

**Modelos de mistura finita usando a
classe de distribuições α potência**

Gualberto Segundo Agamez Montalvo

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Área de concentração: Estatística
Orientador: Prof. Dra. Márcia D'Elia Branco

Durante o desenvolvimento deste trabalho o autor recebeu auxílio
financeiro da CAPES e do CNPq

São Paulo, fevereiro de 2017

Modelos de mistura finita usando a classe de distribuições α potência

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 22/02/2017. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof^a. Dr^a. Márcia D'Elia Branco (orientadora) - IME-USP
- Prof. Dr. Filidor Edilfonso Vilca Labra - UNICAMP
- Prof^a. Dr^a. Viviana Giampaoli - IME-USP
- Prof. Dr. Rafael Bráz Azevedo Farias - UFC
- Prof. Dr. Marcos Oliveira Prates - UFMG

*Dedico esta conquista com muito amor,
a meus pais, Gualberto e Cecilia,
a minha irmã, Carmen (Mi Poli),
e a minha avó, Matilde, (in memoriam).*

Agradecimentos

Em primeiro lugar, agradeço a Deus, por cada dia de vida e por me dar oportunidade de estar aqui. Por me presentear com sabedoria, coragem e força para nunca dar um passo para trás e tentar viver segundo seus princípios.

A meus pais, Gualberto e Cecília, por me educarem e me fornecerem proteção, amor e muita dedicação. Seus esforços deram frutos e hoje podem ter certeza de que sou Doutor graças a vocês. Agradeço sobretudo a minha Mainha (como falariam no Nordeste do Brasil), pois graças a ela eu cresci inspirado a ser sempre um bom homem, gostaria de deixar aqui imortalizado esse fato.

A minha irmã Carmen, por sempre estar por perto de alguma forma e me dar forças, mesmo à distância. Obrigado minha querida irmãzinha por sempre falar que estava orgulhosa de mim, essas palavras marcaram meus momentos de solidão e preencheram eles com alegria e esperança.

A minha avó Matilde (*in memoriam*), por me ver como a luz de seus olhos e sempre confiar em mim. Ela que por tantas vezes me falou que não sabia porque eu estudava tanto, que não entendia nada do que eu fazia, mas, que tinha certeza que tudo estava correto e me beijava com carinho a testa a cada despedida.

A minha orientadora Profa. Dra. Márcia D'Elia Branco, por sua orientação, confiança e principalmente por sua amizade. Ela que é um ser humano impar, sempre atenta e muito justa. Muito obrigado professora por me apoiar em todos os momentos e por ser parte de minha família aqui no Brasil.

A meus tios e minhas tias, especialmente, a tio Robinson e a tia Yazmin, vocês dois são como meus segundos pais e sempre estão em meu coração. Obrigado porque sempre estiveram ao meu lado, nas minhas idas e vindas.

A meu avô Robinson e a minha avó Carmen, por sempre estar em suas orações e todo seu amor.

A meus amigos na Colômbia, Ignacio, Fernando, Angela e Luz Dary, por confiarem em mim. Por acreditarem de coração que tudo ia dar certo, mesmo antes do desenvolvimento deste projeto. A meu amigo Mario Pacheco, por suas palavras de ânimo nos momentos difíceis. A minha querida amiga, Jeny (Minha feinha), pelas tantas vezes que teve que me ouvir falar da tese e de repente uma ideia surgia em minha cabeça, por cuidar de mim tantas vezes e por seu carinho incondicional.

A meus amigos aqui no Brasil, Ana, Eneilson, Francisco Moises, Wando, Ju (Meu time), pela amizade e carinho de todos vocês, que são muito importantes para mim. E com esquecer do grande Sair Máximo? Meu grande amigo, com quem compartilhei muitas alegrias e tristezas também, mais que um amigo ele é um irmão para mim.

A minha namorada Dra. Talita Camila, por todo seu amor. Por chegar a minha vida no momento certo. Por deixar que eu fosse sua *kryptonita* e me presentear com seu maravilhoso coração. Por cada segundo que esteve a meu lado e confiou cegamente em mim. Muito obrigado meu anjo e a toda sua família.

Resumo

Agamez Montalvo, G. S. **Modelos de mistura finita usando a classe de distribuições α potência**. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

As misturas finitas de distribuições de probabilidades permitem a modelagem de dados que podem exibir assimetria, caudas pesadas, dados atípicos ou combinações destes ao mesmo tempo. Neste trabalho, uma classe flexível de modelos usando misturas finitas de distribuições α potência é proposta, e conseqüentemente, a família de misturas finitas de K componentes de mistura de distribuições α potência normal para o caso de localização-escala (MF-PN) é estudada. As abordagens clássica e bayesiana são empregadas para a inferência dos parâmetros do modelo proposto. Para a abordagem clássica, um algoritmo EM é desenvolvido com a finalidade de obter as estimativas de máxima verossimilhança. Para a abordagem bayesiana, diferentes distribuições *a priori* são exploradas e um algoritmo MCMC é descrito. As metodologias propostas são ilustradas usando um conjunto de dados reais sobre o índice de massa corporal de homens nos Estados Unidos da América (USA por suas siglas em inglês). Adicionalmente, para ambas abordagens são realizados estudos de simulação com a finalidade de analisar o comportamento das estimativas dos parâmetros e a habilidade dos modelos de MF-PN de agrupar observações. A taxa de localização (ou classificação) correta é avaliada, dado que uma das principais utilidades das misturas finitas de distribuições é poder identificar grupos ou *clusters* de observações com características similares. Sob um enfoque clássico, o modelo de MF-PN é comparado com alguns modelos pertencentes às misturas finitas de distribuições de mistura de escala da normal assimétrica, concluindo que o modelo proposto é eficiente para classificar dados provenientes dessas mis-

turas. Além disso, a identificabilidade da classe de distribuições α potência (AP) é provada. Finalmente, a metodologia proposta para a classe AP é ilustrada com um conjunto de dados da circunferência da cintura para adolescentes de sexo masculino com idade entre 10 e 20 anos de escolas públicas de São Paulo.

Palavras-chave: Misturas finitas; α potência; algoritmo EM; algoritmo MCMC; classificação; identificabilidade.

Abstract

Agamez Montalvo, G. S. **Finite mixture models using the α power distributions class**. Thesis (Doctoral) - Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2017.

Finite mixtures of probability distributions allow modelling data that may exhibit skewness, heavy tails, outliers or combinations of these at the same time. In this work, a flexible class of models using finite mixtures of α power distribution is proposed, and consequently, the finite mixtures family of K mixture components of power normal distributions for the location-scale case (FM-PN) is studied. The classical and Bayesian approaches are used to the parameters inference of the proposed model. An EM algorithm was developed for the classical approach with the purpose of obtaining the maximum likelihood estimates. In the Bayesian approach, different *prior* distributions are studied and a MCMC algorithm is described. The proposed methodologies are illustrated using a real data set on the body mass index of men in the USA. Additionally, for both approaches simulation studies are performed with the purpose of analyzing the behavior of the estimates parameters and the ability of the FM-PN models of clustering observation. The correct localization (or classification) rate is evaluated, since one of the main utilities of the finite mixture of distributions is identify clusters of observations with similar characteristics. Under a classical approach, the FM-PN model is compared with some models belonging to the finite mixture of scale mixtures of skew-normal distributions, concluding that the proposed model is efficient for classify data from these mixtures. In addition, the identifiability of the α power distributions class (AP) is proved. Finally, the proposed methodology for the AP class is illustrated with a data set of waist circumference for male teenagers aged between ten and twenty years of public schools

in São Paulo.

Keywords: Finite mixtures; α power; EM algorithm; MCMC algorithm; classification; identifiability.

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução	1
2 A classe de distribuições α potência	5
2.1 Introdução	5
2.2 Distribuição α potência normal	7
2.3 Inferência via máxima verossimilhança	9
2.3.1 Família de distribuições PN	9
2.3.2 Caso localização - escala	10
2.4 Estudo de simulação para a família PN	11
2.5 Identificabilidade para a classe de distribuições AP	15
2.6 Considerações finais do capítulo	16
3 Modelos de mistura finita usando a classe de distribuições AP	17
3.1 Introdução	18
3.2 O algoritmo EM para misturas finitas.	19
3.3 Modelos de mistura finita de α potência	22
3.3.1 O modelo	22
3.3.2 Estimação dos parâmetros via algoritmo EM	23
3.3.3 Notas sobre a implementação	24
3.3.4 A matriz de informação observada	25
3.3.5 Critério de classificação	27

3.3.6	Critérios de seleção de modelos	27
3.4	Estudo de simulação	28
3.4.1	Estudo do viés e do erro quadrático médio	28
3.4.2	Agrupamento - Clustering.	33
3.5	Ilustração	36
3.6	Considerações finais do capítulo	39
3.7	Detalhes técnicos do capítulo	41
4	Inferência bayesiana para a classe de distribuições AP	43
4.1	Introdução	44
4.2	Inferência bayesiana para o parâmetro de potência da classe de distribuições AP	45
4.2.1	Conjugação Bayesiana e preditiva	46
4.2.2	Estudo de simulação para a família PN	48
4.3	Inferência bayesiana para o modelo de localização-escala	51
4.3.1	Caso geral do modelo de localização-escala	51
4.3.2	Abordagem bayesiana para a inferência dos parâmetros da família de distribuições $PN(\xi, \eta, \alpha)$	52
4.4	Critérios bayesianos de seleção de modelos	55
4.5	Estudo de simulação para a família $PN(\xi, \eta, \alpha)$	56
4.6	Aplicação: Dados da circunferência da cintura para adolescentes de escolas públicas de São Paulo	59
4.7	Considerações finais do capítulo	62
4.8	Detalhes técnicos do capítulo	63
5	Inferência bayesiana para os modelos de MF-PN	69
5.1	Introdução	70
5.2	Identificabilidade do modelo de MF-PN	71
5.3	Inferência bayesiana para os parâmetros da família de MF-PN	72
5.3.1	Distribuição a priori para os parâmetros do modelo de MF-PN	72
5.3.2	Algoritmo MCMC para o modelo de MF-PN	75

5.3.3	Valores iniciais para o algoritmo MCMC	75
5.3.4	Critério de classificação	76
5.4	Estudo de simulação	76
5.5	Ilustração	79
5.6	Considerações finais do capítulo	81
6	Considerações finais da tese e pesquisas futuras	83
6.1	Considerações finais	83
6.2	Sugestões para pesquisas futuras	85
	Referências Bibliográficas	87

Lista de Figuras

2.1	Função densidade da α potência normal	8
2.2	Duas funções de densidades similares.	14
2.3	Funções de densidades similares.	14
3.1	Viés para o caso $K = 2$	30
3.2	EAMR para o caso $K = 2$	31
3.3	Dados simulados para MF-SN, MF-ST, MF-SSL e MF-SNC ($n = 1000$).	34
3.4	Dados simulados para uma MF-PN com dois componentes.	35
3.5	Ajuste das distribuições MF-SN, MF-N e MF-PN com a resposta do IMC.	37
3.6	Ajuste das distribuições MF-ST, MF-T e MF-PN com a resposta do IMC.	37
4.1	Viés relativo para α	49
4.2	Erro absoluto relativo médio para α	50
4.3	Box plot dos dados de circunferência da cintura.	59
4.4	Histograma dos dados de circunferência da cintura e o ajuste dos modelos Normal e PN.	61
5.1	Histograma de um modelo de MF-PN com $K = 3$ e $n = 1000$	77
5.2	Ajuste dos modelos de MF-N e MF-PN com a resposta do IMC.	80

Lista de Tabelas

2.1	O viés, o VR, o EQM e o EAMR para os parâmetros do modelo PN usando M1.	13
2.2	O viés, o VR, o EQM e o EAMR para os parâmetros do modelo PN usando M2.	13
2.3	<i>Deviance</i> médio (desvio padrão).	15
3.1	Viés e VR.	30
3.2	EQM e EAMR para o caso $K = 2$	31
3.3	VR, EQM e EAMR para o caso $K = 3$	32
3.4	EQM e EAMR para o caso $K = 3$	32
3.5	Taxas médias de localizações para cada MF.	34
3.6	Taxas médias de localizações o modelo de MF-PN.	36
3.7	Estimativas dos dois primeiros momentos e das probabilidades de mistura para os dados de IMC.	38
3.8	Critérios de seleção de modelos para o conjunto de dados do IMC.	38
3.9	Taxas de classificação correta.	39
4.1	Amplitude média e probabilidade de cobertura para o parâmetro da família PN.	49
4.2	VR e EAMR das estimativas Bayesianas para o parâmetro da família PN.	57
4.3	<i>Deviance</i> médio e seus desvio padrão.	58
4.4	Estatísticas descritivas para a variável: circunferência da cintura.	59
4.5	Critérios bayesianos de seleção de modelos para os modelos PN e N.	60

4.6	Estatísticas resumo para a mediana, o primeiro momento e a variância do modelo PN (Jeffreys Truncada) e os parâmetros do modelo Normal.	61
5.1	VR, EAMR e estatísticas resumo.	78
5.2	Taxas de classificação correta.	79
5.3	Critérios bayesianos de seleção de modelos para os modelos de MF-N e MF-PN.	80
5.4	Estatísticas resumo para os três primeiros quantis e as probabilidades de mistura para os dados de IMC.	81

Capítulo 1

Introdução

Os modelos de mistura finita de distribuições de probabilidades são uma ferramenta muito útil para modelar dados que podem exibir assimetria, excesso de curtose, caudas pesadas e observações discrepantes ao mesmo tempo. Modelos de misturas finitas são definidos como uma combinação linear de duas ou mais funções de densidade de probabilidade, permitindo uma grande flexibilidade na modelagem de dados heterogêneos presentes em diferentes áreas do conhecimento como bioestatística, finanças, análise de imagens e ciências médicas, entre outras. Em [McLachlan e Basford \(1988\)](#), [McLachlan e Peel \(2000\)](#), [Frühwirth-Schnatter \(2006\)](#) e [Mengersen *et al.* \(2011\)](#) encontram-se vários exemplos de aplicações das misturas finitas de distribuições.

Uma das principais utilidades das misturas finitas de distribuições é poder identificar grupos de observações com características similares, ou seja, identificar *clusters*, quando não há outras informações além dos valores observados. Neste contexto, cada uma das componentes de mistura pode ser associada a um grupo ou *cluster* diferente. Adicionalmente, no mesmo contexto, as misturas finitas de distribuições podem ser aplicadas na análise discriminante. Para mais detalhes ver [McLachlan e Peel \(2000\)](#).

As misturas finitas de distribuições simétricas foram as primeiras a serem estudadas. No entanto, nos últimos anos os estudos sobre modelos de misturas finitas estão focados principalmente a misturas cujos componentes são distribuições mais flexíveis, como por exemplo, a distribuição normal assimétrica, t-Student assimétrica e slash assimétrica. No caso de variáveis aleatórias univariadas, ver por exemplo, os trabalhos de [Basso *et al.* \(2010\)](#) e [Lin *et al.*](#)

(2007a,b).

Sob um enfoque não paramétrico vários algoritmos de agrupamento foram propostos, ver por exemplo, [Jarvis e Patrick \(1973\)](#), [Comaniciu e Meer \(2002\)](#) e [KumarPatidar *et al.* \(2012\)](#). Porém, a maioria dessas abordagens nem sempre é bem-sucedida em encontrar *clusters* em conjuntos de dados de alta dimensão, dado que é difícil definir a vizinhança de um dado em um espaço de alta dimensão quando o tamanho da amostra é pequeno. Portanto, quase todos os algoritmos não paramétricos baseados nestes métodos têm sido aplicados apenas a problemas de cluster de baixa dimensão. Adicionalmente, com estes métodos não é possível especificar *a priori* o número desejado de *clusters*. Cabe ressaltar que nesta tese, a abordagem não paramétrica não será utilizada. No entanto, achamos relevantes mencionar que existem vários trabalhos desenvolvidos sobre um enfoque não paramétrico para o problema de agrupamento ou classificação.

Esta tese visa apresentar uma classe flexível de modelos estatísticos paramétricos usando misturas finitas de distribuições α potência sobre os enfoques clássico e bayesiano, e consequentemente, utilizar a mistura finita de distribuições α potência normal para o caso de localização-escala em problemas de classificação.

Este trabalho está organizado como segue. Primeiramente, no Capítulo 2, fazemos uma breve revisão bibliográfica da classe de distribuições α potência (AP) e da família de distribuições α potência normal (PN) para o caso de localização-escala ou modelo de três parâmetros. Adicionalmente, neste capítulo são apresentados um estudo de simulação para estudar o comportamento dos parâmetros desta família de distribuições e uma prova formal da identificabilidade da classe de distribuições AP.

No Capítulo 3 propomos o modelo de mistura finita usando a classe de distribuições AP, com ênfase nas misturas finitas usando a distribuição PN para o caso localização-escala. Um algoritmo EM é desenvolvido para obter as estimativas do modelo proposto via máxima verossimilhança e a matriz de informação observada é derivada. Estudos de simulação são apresentados para avaliar as estimativas obtidas e a taxa de classificação correta do modelo. Finalmente, um conjunto dados reais é usado para ilustrar a metodologia proposta.

No Capítulo 4, propomos uma abordagem bayesiana para a inferência do parâmetro de interesse da classe de distribuições AP. Diferentes distribuições *a priori* são usadas e

as distribuições preditivas *a posteriori* são obtidas. Estudos de simulação são apresentados com a finalidade de avaliar o comportamento das estimativas bayesianas para a família PN. Posteriormente, o caso de localização-escala também é abordado. Estudos de simulação para esta família são realizados para estudar o comportamento de alguns quantis obtidos com as estimativas bayesianas dos parâmetros. Finalmente, a metodologia é aplicada a um conjunto de dados reais.

No Capítulo 5 propomos um enfoque bayesiano para o modelo de mistura finita de K componentes de mistura de α potência normal para o caso de localização-escala proposto no Capítulo 3. Um algoritmo de Monte Carlo via cadeias de Markov é proposto para obter as estimativas dos parâmetros do modelo de MF-PN. Os problemas de escolha de valores iniciais e de identificabilidade das etiquetas ou rótulos dos *clusters* são discutidos. Estudos de simulação são realizados para avaliar as estimativas obtidas e a taxa de classificação correta do modelo de MF-PN. Finalizamos esse capítulo apresentando uma aplicação a dados reais.

No Capítulo 6 são apresentadas as considerações finais da tese, assim como os possíveis trabalhos futuros.

Esse trabalho de tese foi escrito de tal forma que cada capítulo pudesse ser lido independentemente dos demais. Porém, todos os capítulos dependem do Capítulo 2, que contém uma breve revisão bibliográfica do modelo α potência.

Os algoritmos utilizados foram implementados no programa estatístico [R Core Team \(2015\)](#). Adicionalmente, cabe ressaltar que foi construído o pacote `mixpn`, que permite:

- Estimar os parâmetros do modelo;
- Classificar as observações;
- Desenhar o histograma e a densidade estimada;
- Calcular os valores iniciais do algoritmo EM, critérios de seleção e a matriz de informação observada;

dado um conjunto de dados e um número fixo de componentes. Adicionalmente, o pacote permite gerar observações de uma FM-PN e ajustar dados da distribuição PN. O instalador de pacote pode ser obtido entrando em contato com o autor.

Capítulo 2

A classe de distribuições α potência

2.1 Introdução

Lehmann (1953) propôs uma classe de distribuições de probabilidades com função de distribuição acumulativa (fda) F_G dada por

$$F_G(z; \alpha) = \{G(z)\}^\alpha, \quad z \in \mathbb{R}, \quad (2.1)$$

sendo G uma fda contínua e diferenciável e α um número inteiro positivo. Em outras palavras, F_G é a potência inteira de uma função de distribuição G . A função de distribuição G é conhecida como função de distribuição base (*'baseline distribution'* em inglês). Na literatura esta classe de distribuições de probabilidade também é conhecida como as *alternativas de Lehmann*.

Se o parâmetro de potência α da classe de distribuições de probabilidade proposta por Lehmann é igual a um, é obtida a mesma distribuição base ($F_G(z; \alpha) = G(z)$). Porém, se este parâmetro é um número inteiro positivo, esta classe de distribuições pode ser vista como a função de distribuição do máximo de uma amostra de tamanho α extraída de uma população cuja fda é G .

A definição de Lehmann foi estendida por Durrans (1992) num contexto hidrológico para $\alpha \in \mathbb{R}^+$, derivando esta classe de distribuições como uma distribuição da estatística de ordem fracionária. No entanto, Durrans não referencia o trabalho de Lehmann (1953). A função densidade de probabilidade (fdp) associada à função de distribuição acumulativa $F_G(z; \alpha)$ é

dada por

$$f_G(z; \alpha) = \alpha g(z) \{G(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \quad \alpha \in \mathbb{R}^+, \quad (2.2)$$

sendo G uma fda (absolutamente contínua) com fdp $g = dG$.

Esta classe de distribuições têm sido amplamente estudada na literatura, principalmente analisando conjuntos de dados de tempos de vida assimétricos e problemas de confiabilidade, usando como função de distribuição base as distribuições Gama, Log-normal, Weibull e Exponencial. Ver por exemplo, Gupta *et al.* (1998), Gupta e Kundu (1999, 2001a,b, 2003, 2004, 2007).

Posteriormente, em Flórez (2011) usa-se a denominação α potência (AP). Enquanto, paralelamente, em Asgharzadeh e Fallah (2011) ela é denominada de classe exponencializada de distribuições. Neste trabalho será adotada a denominação dada em Flórez (2011) e a notação $Z \sim AP(\alpha)$.

Pewsey *et al.* (2012) considera a inferência baseada na função de verossimilhança para a classe de distribuições α potência e apresenta alguns resultados que podem ser utilizados para conduzir a inferência dos três parâmetros da extensão da localização-escala da classe de distribuições AP.

O caso de localização-escala ou modelo de três parâmetros para a classe de distribuições AP é obtido considerando $X = \xi + \eta Z$, onde $Z \sim AP(\alpha)$, $\xi \in \mathbb{R}$ e $\eta \in \mathbb{R}^+$. A função densidade de probabilidade de X é dada por

$$f_G(x; \xi, \eta, \alpha) = \frac{\alpha}{\eta} g\left(\frac{x - \xi}{\eta}\right) \left\{G\left(\frac{x - \xi}{\eta}\right)\right\}^{\alpha-1}, \quad x \in \mathbb{R}, \quad (2.3)$$

sendo ξ o parâmetro de localização e η o parâmetro de escala.

A fda de X é dada por

$$F_G(x; \xi, \eta, \alpha) = \left\{G\left(\frac{x - \xi}{\eta}\right)\right\}^{\alpha}, \quad x \in \mathbb{R}. \quad (2.4)$$

Para o modelo de três parâmetros para a classe de distribuições AP é usada a notação $X \sim AP(\xi, \eta, \alpha)$.

2.2 Distribuição α potência normal

Durrans (1992) estudou o caso em que a função de distribuição base G é a fda da normal padrão. Flórez (2011) chamou este caso de família de distribuições α potência normal assimétrica, cuja função densidade é dada por

$$f_{\phi}(z; \alpha) = \alpha \phi(z) \{\Phi(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \quad (2.5)$$

sendo ϕ e Φ a função densidade e distribuição da normal padrão, respectivamente, e α é um parâmetro de forma. Denotada por $Z \sim PN(\alpha)$. A diferença de Flórez (2011), esta família de distribuições será chamada simplesmente de α potência normal. Dado que, poderia se interpretar que a função de distribuição base é a distribuição normal assimétrica.

A fda de Z é dada por

$$F_{\Phi}(z; \alpha) = \{\Phi(z)\}^{\alpha}. \quad (2.6)$$

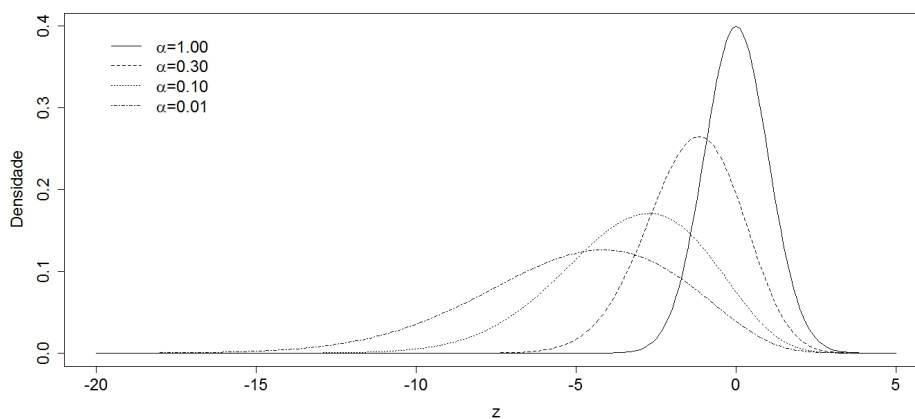
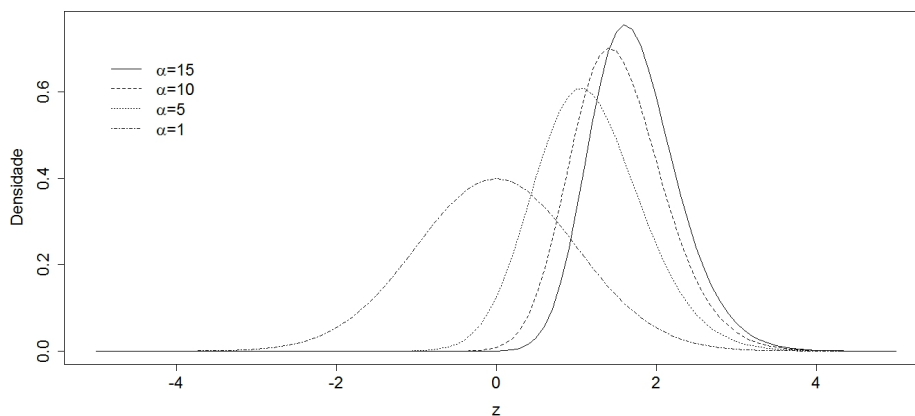
A Figura 2.1 apresenta as funções densidades da distribuição α potência normal para diferentes valores do parâmetro α . Pode-se observar que o parâmetro de potência α tem influência na localização, dispersão, curtose e assimetria da distribuição.

Gupta e Gupta (2008) apresentam uma discussão sobre os momentos da distribuição PN. Os mesmo não têm forma fechada, porém, se $Z \sim PN(\alpha)$ o momento de ordem n pode ser calculado numericamente como

$$E(Z^n) = \alpha \int_0^1 \{\Phi^{-1}(z)\}^n z^{\alpha-1} dz.$$

Também foi considerada a família de distribuição α potência normal para o caso localização-escala, obtida quando a função de distribuição base é Φ na equação (2.3). Denotada por $X \sim PN(\xi, \eta, \alpha)$, com fdp dada por

$$f_{\phi_{LE}}(x; \xi, \eta, \alpha) = \frac{\alpha}{\eta} \phi\left(\frac{x - \xi}{\eta}\right) \left\{ \Phi\left(\frac{x - \xi}{\eta}\right) \right\}^{\alpha-1}, \quad x \in \mathbb{R}. \quad (2.7)$$

(a) α potência normal para $\alpha \leq 1$.(b) α potência normal para $\alpha \geq 1$.Figura 2.1: Função densidade da α potência normal

Em [Pewsey *et al.* \(2012\)](#) foram apresentados os coeficientes de assimetria e curtose para a família de distribuições PN. Esses autores mostraram que para valores de α entre 0,0005 e 50.000 esses coeficientes estão nos intervalos $[-0,6115; 0,9007]$ e $[1,7170; 4,3556]$, respectivamente.

[Azzalini \(1985\)](#) desenvolveu uma classe geral de distribuições assimétricas (univariadas) com fdp dada por

$$\varphi(z; \lambda) = 2f(z)G(\lambda z), \quad z, \lambda \in \mathbb{R}, \quad (2.8)$$

sendo f uma função densidade de probabilidade simétrica ao redor de zero, G uma função de distribuição acumulada absolutamente contínua e simétrica ao redor de zero, e λ o parâmetro de forma que controla a assimetria.

Particularmente, uma variável aleatória Y segue uma distribuição normal assimétrica (SN) com parâmetro de assimetria λ , se sua função densidade é dada por

$$f(y; \xi, \eta, \lambda) = 2\phi(y) \Phi(\lambda y), \quad y \in \mathbb{R}, \quad (2.9)$$

sendo ϕ e Φ a função de densidade e distribuição da normal padrão, respectivamente. Usamos a notação $Y \sim SN(\lambda)$. Esta densidade foi estudada por [Azzalini \(1985, 1986, 2005\)](#), [Henze \(1986\)](#), [Chiogna \(1998\)](#), [Pewsey \(2000, 2006\)](#) e outros. Destes estudos cabe ressaltar que para a distribuição SN, o coeficiente de assimetria toma valores entre $(-0,9953; 0,9953)$ e o coeficiente de curtose entre $[3; 3,8692)$.

Portanto, enquanto a distribuição normal assimétrica é capaz de modelar uma gama mais ampla de assimetria, a distribuição α potência normal modela uma gama maior de curtose.

2.3 Inferência via máxima verossimilhança

Nesta seção são apresentados alguns resultados encontrados por [Gupta e Gupta \(2008\)](#) sobre inferência via máxima verossimilhança para a família de distribuições PN e sua extensão ao modelo de três parâmetros.

2.3.1 Família de distribuições PN

Sejam Z_1, \dots, Z_n variáveis aleatórias i.i.d. seguindo o modelo PN. O estimador de máxima verossimilhança (EMV) de α é dada por

$$\hat{\alpha} = -\frac{n}{\sum_{i=1}^n \ln \Phi(z_i)}.$$

Podemos facilmente observar que

$$E(\hat{\alpha}) = \frac{n}{n-1}\alpha, \quad \text{Var}(\hat{\alpha}) = \frac{n^2}{(n-1)^2(n-2)}\alpha^2$$

e

$$\text{EQM}(\hat{\alpha}) = \frac{n+2}{(n-1)(n-2)}\alpha^2, \quad n > 2.$$

Modificando o EMV para obter o estimador não viesado de α temos

$$\hat{\alpha}_u = \frac{n-1}{n}\hat{\alpha}.$$

Assim $\hat{\alpha}_u$ é o estimador não viesado com a variância $\alpha^2/(n-2)$, $n > 2$. Note-se que a $\text{Var}(\hat{\alpha}_u)$ está mais perto do limite inferior de Cramer-Rao (α^2/n) em relação a $\text{Var}(\hat{\alpha})$. Além disso, em [Gupta e Gupta \(2008\)](#) é apresentado um intervalo de confiança de $(1-\delta) \times 100\%$ para α dado por

$$\left(\frac{X_{2n,\delta/2}^2}{-2 \sum_{i=1}^n \ln \Phi(z_i)}, \frac{X_{2n,1-\delta/2}^2}{-2 \sum_{i=1}^n \ln \Phi(z_i)} \right),$$

devido a que $-\ln \Phi(z_i)$, com $i = 1, \dots, n$ é uma amostra aleatória i.i.d. de uma distribuição exponencial de parâmetro α , implicando que $-2\alpha \ln \Phi(z_i)$ seja uma quantidade pivotal.

2.3.2 Caso localização - escala

Sejam Z_1, \dots, Z_n variáveis aleatórias i.i.d. seguindo o modelo de três parâmetros da família PN. Então, a função de log-verossimilhança é dada por

$$l(\xi, \eta, \alpha) = n \ln \alpha - \frac{1}{2} \sum_{i=1}^n \left(\frac{z_i - \xi}{\eta} \right)^2 + (\alpha - 1) \sum_{i=1}^n \ln \Phi \left(\frac{z_i - \xi}{\eta} \right). \quad (2.10)$$

As equações de estimação usuais são:

$$\frac{\partial l}{\partial \xi} = (\alpha - 1) \sum_{i=1}^n R_I^{-1} \left(\frac{z_i - \xi}{\eta} \right) \left(-\frac{1}{\eta} \right) + \sum_{i=1}^n \left(\frac{z_i - \xi}{\eta^2} \right) = 0; \quad (2.11)$$

$$\frac{\partial l}{\partial \eta} = -\frac{n}{\eta} + (\alpha - 1) \sum_{i=1}^n R_I^{-1} \left(\frac{z_i - \xi}{\eta} \right) \left(-\frac{z_i - \xi}{\eta^2} \right) + \sum_{i=1}^n \frac{(z_i - \xi)^2}{\eta^3} = 0; \quad (2.12)$$

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \ln \left(\frac{z_i - \xi}{\eta} \right) = 0, \quad (2.13)$$

em que

$$R_I(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}, \quad (2.14)$$

e

$$R_I^{-1}(\cdot) = \frac{1}{R_I(\cdot)}.$$

Da equação 2.13 temos que o EMV de α é

$$\hat{\alpha} = -\frac{n}{\sum_{i=1}^n \ln \Phi \left(\frac{z_i - \xi}{\eta} \right)}. \quad (2.15)$$

No entanto, as equações (2.11) e (2.12) não permitem obter expressões de forma fechada para os EMV de ξ e η . Assim, uma opção para obter os estimadores de máxima verossimilhança para ξ e η é substituir (2.15) em (2.11) e (2.12), e logo, resolver estas equações usando algum método numérico iterativo.

A função R_I definida na equação (2.14) é conhecida como taxa de risco invertida. Gupta e Gupta (2007) estudaram o caso em que a função de distribuição base de R_I é uma função de distribuição geral.

2.4 Estudo de simulação para a família PN

Gupta e Gupta (2008) não realizaram estudos de simulação para avaliar as estimativas obtidas via máxima verossimilhança para a família de distribuições PN. Portanto, para preencher essa lacuna, nesta seção o comportamento destas estimativas é estudado.

As medidas de avaliação utilizadas foram o viés, o viés relativo (VR), o erro quadrado médio (EQM) e o erro absoluto médio relativo (EAMR). Diversos estudos de simulação foram desenvolvidos, nos quais o comportamento das estimativas para diferentes conjuntos de parâmetros e tamanhos amostrais foi explorado. Nesta seção são apresentados os resultados

obtidos usando os tamanhos amostrais $n = 50, 100$ e 200 , e os valores dos parâmetros fixados em $\xi = 10$, $\eta = 1$ e $\alpha = 100$, dado que os resultados obtidos nos outros estudos de simulação foram similares.

Foram geradas $m = 1.000$ amostras de tamanho n da família de distribuições PN. Em seguida, estimamos os parâmetros do modelo usando duas metodologias. Na primeira ($M1$) foram estimados os três parâmetros livremente usando a função de log-verossimilhança apresentada em (2.10) e na segunda ($M2$) foi utilizada a metodologia descrita na subseção 2.3.2.

Posteriormente, as medidas de avaliação foram obtidos para todas as amostras. Para ξ , eles são definidos como

$$\begin{aligned} \text{viés} &= \frac{1}{m} \sum_{i=1}^m (\widehat{\xi}_i - \xi), & \text{VR} &= \frac{1}{m} \sum_{i=1}^m \frac{(\widehat{\xi}_i - \xi)}{\xi}, \\ \text{EQM} &= \frac{1}{m} \sum_{i=1}^m (\widehat{\xi}_i - \xi)^2 & \text{e} & \quad \text{EAMR} = \frac{1}{m} \sum_{i=1}^m \left| \frac{\widehat{\xi}_i - \xi}{\xi} \right|, \end{aligned}$$

respectivamente, sendo $\widehat{\xi}_i$ a estimativa de ξ quando os dados correspondem à amostra i . As definições para os outros parâmetros são obtidas analogamente.

Cabe ressaltar que todos os procedimentos de otimização utilizados nesta seção para a obtenção das estimativas de máxima verossimilhança dos parâmetros foram realizados usando o método L-BFGS-B da função `optim` do software [R Core Team \(2015\)](#).

Nas Tabelas 2.1 e 2.2 são apresentados o viés, o VR, o EQM e o EAMR para as estimativas dos parâmetros usando $M1$ e $M2$, respectivamente. Pode-se observar que para a primeira metodologia estudada, o EQM e o EAMR dos três parâmetros diminuem quando o tamanho amostral aumenta, porém, este comportamento não é observado para o viés e o VR. Entretanto, com a segunda metodologia, observou-se que a medida que aumenta o tamanho amostral as quatro medidas de avaliação diminuem.

Adicionalmente, ao analisar os resultados obtidos com as quatro medidas propostas e as duas metodologias utilizadas, observou-se que os parâmetros da família de distribuições PN não foram bem recuperados. Este resultado será discutido posteriormente.

Tabela 2.1: O viés, o VR, o EQM e o EAMR para os parâmetros do modelo PN usando M1.

Medida	Parâmetro	n		
		50	100	200
viés	ξ	-0,6808	-0,6824	-0,5258
	η	0,0147	0,0532	0,0502
	α	2,0172e+04	1,2392e+04	7,0973e+03
VR	ξ	-0,0681	-0,0682	-0,0526
	η	0,0147	0,0532	0,0502
	α	2,0173e+02	1,2392e+02	7,0973e+01
EQM	ξ	7,1160	5,2944	3,5368
	η	0,1998	0,1328	0,0869
	α	1,6275e+09	8,1594e+08	4,0142e+08
EAMR	ξ	0,2246	0,1850	0,1407
	η	0,3855	0,3058	0,2354
	α	2,0254e+02	1,2464e+02	7,1618e+01

Tabela 2.2: O viés, o VR, o EQM e o EAMR para os parâmetros do modelo PN usando M2.

Medida	Parâmetro	n		
		50	100	200
viés	ξ	0,4920	0,3310	0,2554
	η	-0,1576	-0,0989	-0,0697
	α	1,6004e+04	2,4361e+03	6,8892e+02
VR	ξ	0,0492	0,0331	0,0255
	η	-0,1576	-0,0989	-0,0697
	α	1,6004e+02	2,4362e+01	6,8892
EQM	ξ	3,0802	2,0682	1,3215
	η	0,1347	0,0783	0,0476
	α	1,4470e+11	1,0391e+09	4,5522e+07
EAMR	ξ	0,1480	0,1173	0,0928
	η	0,3025	0,2289	0,1775
	α	1,6108e+02	2,5298e+01	7,7563

Outros métodos de estimação clássica foram explorados para tentar recuperar de forma adequada os parâmetros da família de distribuições PN, no entanto, como os resultados obtidos foram similares aos expostos anteriormente decidimos não apresentá-los aqui.

Considerando o problema apontado pelo estudo de simulação na recuperação dos parâmetros, decidiu-se explorar a forma da fdp da família PN. Foi observado que dois conjuntos de parâmetros diferentes podem gerar duas funções de densidade com uma forma muito similar, entretanto, não iguais.

Na Figura 2.2 é apresentada a função densidade obtida com os parâmetros utilizados no estudo de simulação, junto com um outro conjunto de parâmetros que fornece uma densidade muito similar. Para ressaltar este problema (quase identificabilidade), na Figura 2.3 são apresentados três conjuntos de parâmetros com funções de densidade semelhantes. Pode-se observar que enquanto os parâmetros de localização e escala são semelhantes, os parâmetros de potência podem ser muito diferentes.

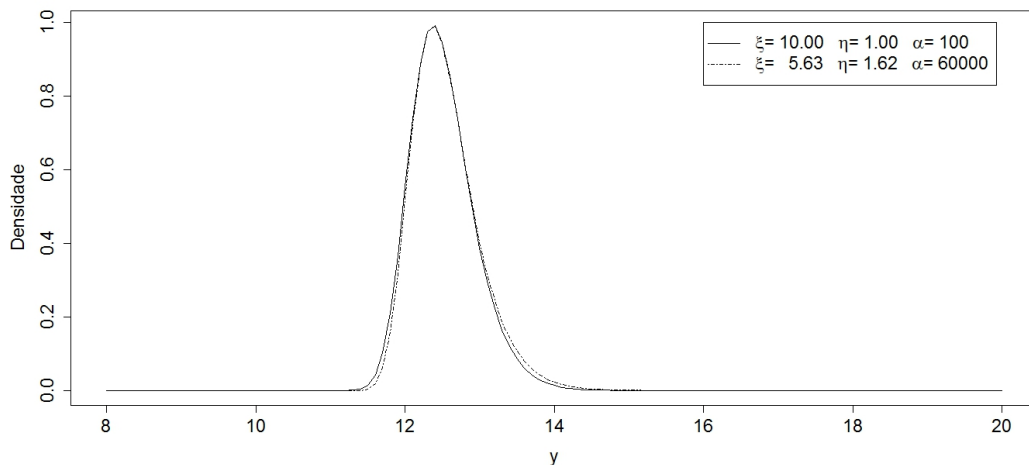


Figura 2.2: Duas funções de densidades similares.

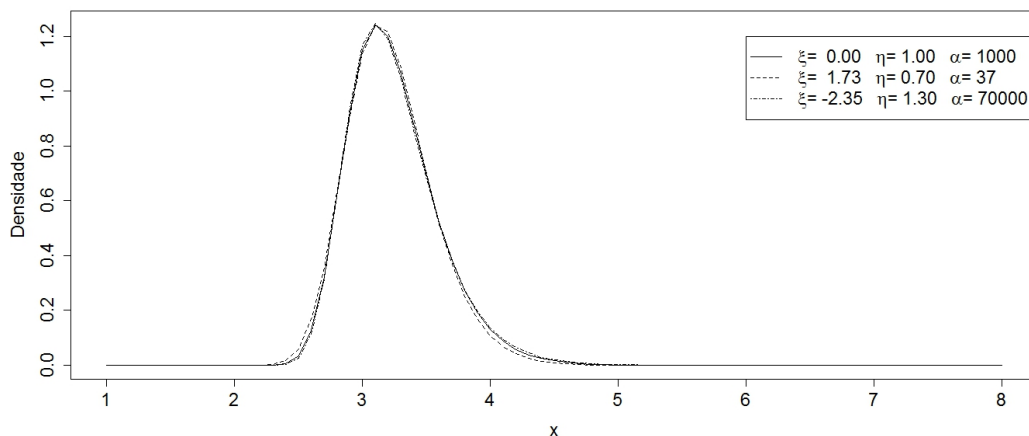


Figura 2.3: Funções de densidades similares.

Finalmente, a função *Deviance* foi utilizada para avaliar o ajuste dos dados. Na Tabela 2.3 são apresentados os *Deviances* médios e seus correspondentes desvios padrões para as duas metodologias utilizadas e os valores reais dos parâmetros. Pode-se observar que os valores dos *Deviances* médios obtidos com as estimativas dos parâmetros do modelo PN foram menores aos calculados com os valores reais. Portanto, as densidades estimadas modelam de forma adequada os dados das amostras.

Tabela 2.3: *Deviance* médio (desvio padrão).

n	Método		
	M1	M2	Valores Reais
50	50,3580 (11,2327)	50,4799 (11,2375)	53,4520 (10,8725)
100	104,9951 (14,7678)	105,1357 (14,7774)	107,8965 (14,4547)
200	212,6529 (21,3019)	212,8146 (21,3229)	215,5923 (21,0546)

2.5 Identificabilidade para a classe de distribuições AP

Nesta seção é apresentada uma prova formal sobre a identificabilidade da classe de distribuições α potência, e conseqüentemente, uma implicação direta para a família de distribuições α potência normal.

Teorema 2.1. *Seja $X \sim AP(\xi, \eta, \alpha)$ com função de distribuição acumulada denotada por $F_G(x, \xi, \eta, \alpha) = \left\{ G\left(\frac{x-\xi}{\eta}\right) \right\}^\alpha$. Se a função de distribuição base G é identificável, então a classe de distribuições α potência é identificável.*

Demonstração (por contradição): Vamos supor que

$$F_G(x, \xi_1, \eta_1, \alpha_1) = F_G(x, \xi_2, \eta_2, \alpha_2) \quad \forall x,$$

em que $\Theta_1 = (\xi_1, \eta_1, \alpha_1) \neq \Theta_2 = (\xi_2, \eta_2, \alpha_2)$.

Então, pela suposição anterior temos que

$$\left\{ G\left(\frac{x-\xi_1}{\eta_1}\right) \right\}^{\alpha_1} = \left\{ G\left(\frac{x-\xi_2}{\eta_2}\right) \right\}^{\alpha_2} \Leftrightarrow \frac{\alpha_1}{\alpha_2} = \frac{\ln G\left(\frac{x-\xi_2}{\eta_2}\right)}{\ln G\left(\frac{x-\xi_1}{\eta_1}\right)} \quad \forall x.$$

Caso 1: Se $(\xi_1, \eta_1) = (\xi_2, \eta_2)$, temos que

$$\frac{\alpha_1}{\alpha_2} \neq 1 = \frac{\ln G\left(\frac{x-\xi_2}{\eta_2}\right)}{\ln G\left(\frac{x-\xi_1}{\eta_1}\right)} \quad \forall x \perp (\text{contradição}),$$

dado que $\alpha_1 \neq \alpha_2$, pois $\Theta_1 \neq \Theta_2$.

Caso 2: Se $(\xi_1, \eta_1) \neq (\xi_2, \eta_2)$, a função $\ln G\left(\frac{x-\xi_2}{\eta_2}\right) / \ln G\left(\frac{x-\xi_1}{\eta_1}\right)$ não é constante para todo x (dado que a função de distribuição G é identificável), ao contrário de α_1/α_2 que é constante para todo x . Obtendo então, novamente, uma contradição.

Portanto, a classe de distribuições α potência é identificável, se sua função de distribuição base for identificável.

Corolário 2.1. *A família de distribuições PN para o caso de localização-escala ou modelo de três parâmetros é identificável.*

Demonstração: Implicação direta do teorema anterior.

2.6 Considerações finais do capítulo

Neste capítulo, apresentamos uma breve revisão bibliográfica da classe de distribuições α potência, e estudos de simulação para estudar o comportamento as estimativas clássicas dos parâmetros da família de distribuições α potência normal para o caso de localização-escala obtidas via máxima verossimilhança.

Nossos estudos de simulação sugerem que os métodos de estimação clássicos permitem obter densidades estimadas que modelam de forma adequada dados provenientes da família de distribuições PN para o modelo de três parâmetros, porém, evidenciou-se um problema de quase identificabilidade intrínseco ao modelo. Finalmente, uma prova formal da identificabilidade da classe de distribuições α potência é apresentada.

Capítulo 3

Modelos de mistura finita usando a classe de distribuições AP

Resumo

Misturas finitas de distribuições fornecem uma ferramenta muito útil para a modelagem e agrupamento de dados heterogêneos; principalmente quando os componentes de mistura são distribuições assimétricas. Neste capítulo, propomos uma classe flexível de modelos usando misturas finitas de distribuições α potência (MF-AP), que permite analisar dados na presença de assimetria, excesso de curtose ou observações atípicas. Um algoritmo EM é usado para calcular as estimativas de máxima verossimilhança do modelo. Além disso, a matriz de informação observada é obtida para as misturas finitas de α potência normal para o caso de localização-escala (MF-PN). Estudos de simulação são apresentados com o objetivo de comparar a taxa de classificação da MF-PN com alguns modelos pertencentes às misturas finitas de distribuições de mistura de escala da normal assimétrica (FM-SMSN por sua sigla em inglês). Adicionalmente, o comportamento do viés e do erro quadrático médio das probabilidades de mistura e dos três primeiros momentos estimados para cada componente de mistura são estudados para diferentes tamanhos amostrais. Finalmente, a metodologia proposta é aplicada a um conjunto de dados reais sobre o índice de massa corporal de homens em Estados Unidos da América.

Palavras-chave: Misturas finitas; dados heterogêneos; α potência; algoritmo EM; modelos baseados em classificação.

3.1 Introdução

A distribuição normal é uma das distribuições contínuas mais usada na literatura, no entanto, não é adequada para modelar dados em presença de assimetria, excesso de curtose ou observações atípicas. A fim de minimizar problemas nas inferências obtidas ao supor normalidade, novas propostas de modelagem têm sido discutidas extensamente. Uma das alternativas para o afastamento da distribuição normal é a classe das distribuições elípticas. Esta classe de distribuições de probabilidades permite reduzir a influência de valores atípicos ou extremos nas inferências, mas requerem a suposição de simetria. Outra alternativa é a distribuição normal assimétrica (skew normal) que permite modelar assimetria e curtose. Esta família foi introduzida formalmente por [Azzalini \(1985\)](#), porém, pode ser vista antes, por exemplo, nos trabalhos de [Roberts \(1966\)](#) e [O'Hagan e Leonard \(1976\)](#).

Na última década, trabalhos realizados por [Gupta e Gupta \(2008\)](#), [Pewsey *et al.* \(2012\)](#), [Flórez \(2011\)](#) e [Martínez-Flórez *et al.* \(2013\)](#) têm retomado os estudos da família α potência e suas propriedades. Proposta por [Lehmann \(1953\)](#), o principal motivo para haver sido tomado em consideração novamente esta família de distribuições é que permite analisar dados na presença de assimetria e observações atípicas.

Não obstante os modelos de misturas finitas (MF) de distribuições são amplamente utilizados em diversas áreas do conhecimento para a modelagem de dados que não apresentam um comportamento usual, isto é, dados que podem exibir assimetria, caudas pesadas e até observações discrepantes ao mesmo tempo; exemplos destas aplicações podem ser encontrados em [McLachlan e Basford \(1988\)](#), [McLachlan e Peel \(2000\)](#), [Frühwirth-Schnatter \(2006\)](#), e [Mengersen *et al.* \(2011\)](#).

As primeiras misturas finitas de distribuições estudadas foram as misturas de distribuições normais. No entanto, nos últimos anos os estudos sobre modelos de misturas finitas estão focados principalmente em misturas cujos componentes são distribuições flexíveis como a distribuição t-Student, skew normal, skew t, skew slash e outras mais (no caso univariado,

ver por exemplo, [Basso *et al.* \(2010\)](#) e [Lin *et al.* \(2007a,b\)](#)).

A mistura finita de distribuições t-Student surge como uma alternativa na modelagem quando um ou vários componentes da mistura têm observações atípicas, mas esta apresenta problemas em modelar dados assimétricos. Uma tentativa de modelar adequadamente dados decorrentes de componentes com observações assimétricas foi a MF de normais assimétricas, no entanto, os parâmetros de assimetria podem ser excessivamente afetados por dados atípicos. Alternativamente, foi desenvolvida a mistura finita da distribuição t-assimétrica para modelar dados cujos componentes de mistura são assimétricos e possuem observações atípicas.

Neste capítulo, será considerada uma outra classe flexível de modelos estatísticos usando misturas finitas de distribuições α potência. Um algoritmo EM será desenvolvido para obter estimativas de máxima verossimilhança de maneira iterativa. Além disso, a matriz de informação observada será obtida explicitamente para as misturas finitas de α potência normal para o caso de localização-escala. O problema de agrupamento (clustering) vai ser abordado, contrastando algumas misturas finitas de distribuições robustas com a MF-PN. Finalmente, a metodologia proposta é ilustrada usando o conjunto de dados reais sobre o índice de massa corporal de homens em Estados Unidos da América estudados anteriormente em [Lin *et al.* \(2007a\)](#) e [Basso *et al.* \(2010\)](#).

3.2 O algoritmo EM para misturas finitas.

O algoritmo EM é um método de calcular a estimativa de máxima verossimilhança quando a maximização direta da função de verossimilhança não é facilmente obtida ([Dempster *et al.*, 1977](#)). O algoritmo EM consiste em dois passos: E (esperança) e M (maximização). No contexto de mistura de densidades o algoritmo EM tem sido usado com bastante sucesso.

Considere uma amostra aleatória z_1, \dots, z_n i.i.d. de tamanho n do modelo de mistura

$$f(z_i; \Theta) = \sum_{j=1}^K \pi_j h_j(z_i; \theta_j) \quad (3.1)$$

sendo $h_j(\cdot)$ a densidade da j -ésima componente da mistura, $\pi_j \geq 0$, $j = 1, \dots, K$, tal que $\sum_{j=1}^K \pi_j = 1$, $\Theta = (\theta_1^t, \dots, \theta_K^t, \pi_1, \dots, \pi_K)^t$ é o vetor com todos os parâmetros do modelo

(3.1) e $\mathbf{z} = (z_1, \dots, z_n)^t$.

Portanto, a função de verossimilhança associada a esta amostra é dada por

$$L_{\text{dados}}(\Theta) = \prod_{i=1}^n \sum_{j=1}^K \pi_j h_j(z_i; \theta_j). \quad (3.2)$$

A maximização direta de (3.2) é um problema de solução numérica difícil e instável. Assim, para desenvolver o algoritmo EM vamos definir uma variável aleatória discreta não observável W tal que

$$P(W_i = j) = \pi_j, \quad \text{com} \quad z_i | W_i = j \sim h_j(\cdot; \theta_j), \quad (3.3)$$

e o vetor aleatório $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iK})^t$, de tal modo que

$$Y_{ij} = 1 \quad \text{se e só se} \quad W_i = j. \quad (3.4)$$

Se $Y_{ij} = 1$, então $Y_{ik} = 0$ para $k \neq j$. A distribuição do vetor \mathbf{Y}_i é multinomial com probabilidades π_1, \dots, π_K . Então, a função de verossimilhança dos dados completos é a função de verossimilhança obtida como se $(\mathbf{z}^t, \mathbf{y}^t)$ foram observáveis, isto é,

$$L(\Theta) = \prod_{i=1}^n \prod_{j=1}^K \pi_j^{y_{ij}} h_j(z_i; \theta_j)^{y_{ij}}. \quad (3.5)$$

Para mais detalhes sobre o algoritmo e suas aplicações em modelos de mistura ver [McLachlan e Krishnan \(2008\)](#).

De forma geral temos que o algoritmo EM esta compreendido pelos passos seguintes:

- Passo E: Calcular a esperança da função de log – verossimilhança em relação à variável latente e condicionada à informação dos dados e a $\Theta^{(t)}$:

$$Q(\Theta | \hat{\Theta}^{(t)}) = E \left[\ln L(\Theta) | \mathbf{z}, \hat{\Theta}^{(t)} \right];$$

- Passo M: Maximize Q com respeito aos parâmetros:

$$\Theta^{(t+1)} = \arg \max_{\Theta} Q \left(\Theta | \widehat{\Theta}^{(t)} \right);$$

- Repita E e M de forma iterativa até alcançar a convergência,

sendo $\widehat{\Theta}^{(t)}$ a estimativa de Θ na etapa t do algoritmo.

Agora, observa-se que

$$E \left[\ln L(\Theta) | \mathbf{z}, \widehat{\Theta}^{(t)} \right] = \sum_{i=1}^n \sum_{j=1}^K E \left(Y_{ij} | \mathbf{z}, \widehat{\Theta}^{(t)} \right) \{ \log \pi_j + \log h(z_i; \theta_j) \}. \quad (3.6)$$

É fácil de calcular $E \left(Y_{ij} | \mathbf{z}, \widehat{\Theta}^{(t)} \right)$, pois a distribuição de Y_{ij} é Bernoulli com probabilidade de sucesso π_j . Então,

$$\hat{y}_{ij}^{(t+1)} = E \left(Y_{ij} | \mathbf{z}, \widehat{\Theta}^{(t)} \right) = P \left(Y_{ij} = 1 | \mathbf{z}, \widehat{\Theta}^{(t)} \right) \propto \hat{\pi}_j^{(t)} h \left(z_i; \hat{\theta}_j^{(t)} \right).$$

Portanto

$$\hat{y}_{ij}^{(t+1)} = \frac{\hat{\pi}_j^{(t)} h \left(z_i; \hat{\theta}_j^{(t)} \right)}{\sum_{m=1}^K \hat{\pi}_m^{(t)} h \left(z_i; \hat{\theta}_m^{(t)} \right)}.$$

Explicitamente precisamos maximizar

$$\frac{\partial Q \left(\Theta | \widehat{\Theta}^{(t)} \right)}{\partial \pi_j} \quad \text{e} \quad \frac{\partial Q \left(\Theta | \widehat{\Theta}^{(t)} \right)}{\partial \theta_j},$$

e de forma geral (usando a restrição $\sum_{j=1}^K \pi_j = 1$) temos que

$$\hat{\pi}_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \hat{y}_{ij}^{(t+1)}. \quad (3.7)$$

3.3 Modelos de mistura finita de α potência

Os modelos de mistura finita de distribuições α potência (MF-AP) são uma ferramenta para modelar e classificar indivíduos em situações onde os dados podem exibir assimetria, caudas pesadas ou/e observações atípicas. Nesta seção são apresentados os modelos de MF-AP e é desenvolvido o algoritmo EM para a classe de mistura finita de α potência normal para o caso localização - escala.

3.3.1 O modelo

Suponha que a população está dividida em K grupos ou componentes e desconhecemos a que componente pertence cada indivíduo. Se a distribuição de cada componente é assumida como uma distribuição α potência, diz-se que um conjunto de n indivíduos independentes $\mathbf{z} = (z_1, \dots, z_n)^t$ seguem uma mistura finita de K componentes de α potência com função de densidade

$$\psi_G(\mathbf{z}_i; \Theta) = \sum_{j=1}^K \pi_j f_G(z_i; \alpha_j, \boldsymbol{\theta}_j), \quad \alpha_j \in \mathbb{R}^+, \quad j = 1, \dots, K, \quad i = 1, \dots, n, \quad (3.8)$$

sendo π_1, \dots, π_K as probabilidades de mistura, que são restritas a ser não-negativas e $\sum_{j=1}^K \pi_j = 1$, $f_G(\cdot; \alpha_j, \boldsymbol{\theta}_j)$ representa a função de densidade da distribuição α potência da componente j , $\boldsymbol{\theta}_j = (\theta_{1j}, \dots, \theta_{qj})^t$ os parâmetros de função de distribuição G , α_j o parâmetro de potência da componente j e Θ todos os parâmetros da mistura.

A função de log-verossimilhança para o modelo geral de mistura finita de α potência é

$$l(\Theta; \mathbf{z}) = \sum_{i=1}^n \ln \left(\sum_{j=1}^K \pi_j f_G(z_i; \alpha_j, \boldsymbol{\theta}_j) \right), \quad (3.9)$$

a qual é uma função usualmente difícil de otimizar, portanto, uma alternativa para obter estimadores dos parâmetros do modelo proposto é o algoritmo EM.

Mistura finita de $PN(\xi, \eta, \alpha)$

A classe de mistura finita de $PN(\xi, \eta, \alpha)$ (MF-PN) é obtida assumindo que cada componente de mistura em (3.8) é a função de distribuição α potência normal para o caso localização-escala, cuja função de densidade é

$$\psi_{\phi_{LE}}(z_i; \Theta) = \sum_{j=1}^K \pi_j \frac{\alpha_j}{\eta_j} \phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \left\{ \Phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \right\}^{\alpha_j - 1}, \quad (3.10)$$

sendo α_j , ξ_j e η_j os parâmetros da distribuição $PN(\xi, \eta, \alpha)$ da j -ésima componente. No caso particular quando $\xi_j = 0$ e $\eta_j = 1$, para todo j , temos a classe de mistura finita de K componentes $PN(0, 1, \alpha)$ com função de densidade dada por

$$\psi_{\phi}(z_i; \Theta) = \sum_{j=1}^K \pi_j \alpha_j \phi(z_i) \{ \Phi(z_i) \}^{\alpha_j - 1}. \quad (3.11)$$

3.3.2 Estimação dos parâmetros via algoritmo EM

Seguindo o procedimento apresentado na Seção 3.2, temos que para a classe de mistura finita de K componentes de α potência normal para o caso localização-escala a função $Q(\Theta | \hat{\Theta}^{(t)})$ está dada por

$$\begin{aligned} Q(\Theta | \hat{\Theta}^{(t)}) &= \sum_{i=1}^n \sum_{j=1}^K \hat{y}_{ij}^{(t)} \left\{ \ln\left(\frac{\alpha_j}{\eta_j}\right) - \ln(\sqrt{2\pi}) - \frac{1}{2} \left(\frac{z_i - \xi_j}{\eta_j}\right)^2 + (\alpha_j - 1) \ln \Phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \right\} \\ &+ \sum_{i=1}^n \sum_{j=1}^K \hat{y}_{ij}^{(t)} \ln \pi_j, \end{aligned} \quad (3.12)$$

sendo

$$\hat{y}_{ij}^{(t)} = \frac{\hat{\pi}_j^{(t)} \frac{\hat{\alpha}_j^{(t)}}{\hat{\eta}_j^{(t)}} \phi\left(\frac{z_i - \hat{\xi}_j^{(t)}}{\hat{\eta}_j^{(t)}}\right) \left\{ \Phi\left(\frac{z_i - \hat{\xi}_j^{(t)}}{\hat{\eta}_j^{(t)}}\right) \right\}^{\hat{\alpha}_j^{(t)} - 1}}{\sum_{m=1}^K \hat{\pi}_m^{(t)} \frac{\hat{\alpha}_m^{(t)}}{\hat{\eta}_m^{(t)}} \phi\left(\frac{z_i - \hat{\xi}_m^{(t)}}{\hat{\eta}_m^{(t)}}\right) \left\{ \Phi\left(\frac{z_i - \hat{\xi}_m^{(t)}}{\hat{\eta}_m^{(t)}}\right) \right\}^{\hat{\alpha}_m^{(t)} - 1}}. \quad (3.13)$$

Assim

$$\hat{\pi}_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \hat{y}_{ij}^{(t)} \quad (3.14)$$

e

$$\hat{\alpha}_j^{(t+1)} = - \frac{\sum_{i=1}^n \hat{y}_{ij}^{(t)}}{\sum_{i=1}^n \hat{y}_{ij}^{(t)} \ln \Phi \left(\frac{z_i - \xi_j^{(t)}}{\hat{\eta}_j^{(t)}} \right)} . \quad (3.15)$$

No entanto, é impossível obter expressões em forma fechada para as estimativas de ξ_j e η_j , isso nos conduz ao uso de métodos numéricos de otimização. No caso da classe de mistura finita de K componentes de $\text{PN}(0, 1, \alpha)$ em que $\xi_j = 0$ e $\eta_j = 1$ para todo $j = 1, \dots, K$, as expressões (3.14) e (3.15) ficam reduzidas a

$$\hat{\pi}_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \hat{y}_{ij}^{(t)}$$

e

$$\hat{\alpha}_j^{(t+1)} = - \frac{\sum_{i=1}^n \hat{y}_{ij}^{(t)}}{\sum_{i=1}^n \hat{y}_{ij}^{(t)} \ln \Phi(z_i)} ,$$

sendo

$$\hat{y}_{ij}^{(t)} = \frac{\hat{\pi}_j^{(t)} \hat{\alpha}_j^{(t)} \phi(z_i) \{\Phi(z_i)\}^{\hat{\alpha}_j^{(t)} - 1}}{\sum_{m=1}^K \hat{\pi}_m^{(t)} \hat{\alpha}_m^{(t)} \phi(z_i) \{\Phi(z_i)\}^{\hat{\alpha}_m^{(t)} - 1}} .$$

As iterações do algoritmo EM são repetidas até que uma regra de convergência adequada é satisfeita.

3.3.3 Notas sobre a implementação

A função de log-verossimilhança no contexto mistura pode ter múltiplas modas e uma das principais críticas nos procedimentos de tipo EM é ficar preso em alguma moda local. Uma maneira conveniente para evitar esse problema é utilizar vários valores iniciais e se existirem várias modas é possível encontrar a moda global, comparando os valores das log-

verossimilhanças.

Neste trabalho foi usada a seguinte metodologia para obter os valores iniciais do modelo de MF-PN. Primeiro, o algoritmo k-means proposto por [Hartigan e Wong \(1979\)](#) e implementado na função `kmeans` do pacote `stats` no software [R Core Team \(2015\)](#) foi empregado para dividir a amostra em K subgrupos ou cluster. Posteriormente, para cada subgrupo foram obtidos os estimadores de máxima verossimilhança da distribuição $PN(\xi, \eta, \alpha)$ ao otimizar a função de log-verossimilhança descrita em (2.10), usando como valores iniciais a média amostral, a variância amostral e 1 para ξ , η e α , respectivamente. Finalmente, os valores iniciais das probabilidades de mistura foram as proporções de indivíduos designados a cada subgrupo com o algoritmo k-means.

Os procedimentos de otimização foram realizados usando o método L-BFGS-B da função `optim` do software [R Core Team \(2015\)](#). Cabe denotar que o algoritmo EM tinha uma taxa de não convergência maior que 0,7 para valores de $\alpha_j < 0,05$, portanto, foram considerados apenas valores de α_j maiores que 0,05.

As regras de convergência para o algoritmo EM mais utilizadas são: $\|\hat{\Theta}^{(t+1)} - \hat{\Theta}^{(t)}\| < \epsilon$, $|l(\hat{\Theta}^{(t+1)}) - l(\hat{\Theta}^{(t)})| < \epsilon$ e $|l(\hat{\Theta}^{(t+1)})/l(\hat{\Theta}^{(t)}) - 1| < \epsilon$, sendo ϵ suficientemente pequeno. No contexto das misturas de distribuições normais univariadas, [Dias e Wedel \(2004\)](#) apresentam uma discussão para a escolha da regra de convergência. Com base na discussão feita por esses autores e em alguns resultados obtidos nos estudos de simulação preliminares, o critério utilizado neste trabalho foi

$$|l(\hat{\Theta}^{(t+1)})/l(\hat{\Theta}^{(t)}) - 1| < 1 \times 10^{-04}.$$

3.3.4 A matriz de informação observada

Sob condições de regularidade, a inversa da matriz de covariância de $\hat{\Theta}$ pode ser aproximada pela inversa da matriz de informação observada I_o . Segundo [McLachlan e Basford \(1988\)](#) temos que

$$I_o(\hat{\Theta}; z) \approx \sum_{i=1}^n \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^t, \quad (3.16)$$

sendo $\hat{\mathbf{u}}_i = \left. \frac{\partial \ln(\psi_G(z_i; \Theta))}{\partial \Theta} \right|_{\Theta = \hat{\Theta}}$.

Para uma mistura finita de K componentes de α potência o vetor $\hat{\mathbf{u}}_i$ pode ser escrito como

$$\hat{\mathbf{u}}_i = \left(\hat{u}_{i,\pi_1}, \dots, \hat{u}_{i,\pi_{K-1}}, \hat{u}_{i,\alpha_1}, \dots, \hat{u}_{i,\alpha_K}, \hat{u}_{i,\theta_{11}}, \dots, \hat{u}_{i,\theta_{1K}}, \dots, \hat{u}_{i,\theta_{q1}}, \dots, \hat{u}_{i,\theta_{qK}} \right)^t,$$

sendo

$$\hat{u}_{i,\pi_r} = \frac{\hat{y}_{ir}}{\hat{\pi}_r} - \frac{\hat{y}_{iK}}{\hat{\pi}_K}, \quad \hat{u}_{i,\alpha_r} = \hat{y}_{ir} \times \left(\frac{1}{\hat{\alpha}_r} + \ln G(z_i; \hat{\boldsymbol{\theta}}_r) \right)$$

e

$$\hat{u}_{i,\theta_{l,r}} = \frac{\pi_r D_{\theta_{l,r}} \left(f_G(z_i; \hat{\alpha}_r, \hat{\boldsymbol{\theta}}_r) \right)}{\psi_G(z_i; \hat{\Theta})},$$

em que

$$D_{\theta_{w,r}} \left(f_G(z_i; \hat{\alpha}_r, \hat{\boldsymbol{\theta}}_r) \right) = \frac{\partial}{\partial \theta_{w,r}} \left(f_G(z_i; \hat{\alpha}_r, \hat{\boldsymbol{\theta}}_r) \right)$$

e

$$\hat{y}_{ir} = \frac{\hat{\pi}_r f_G(z_i; \hat{\alpha}_r, \hat{\boldsymbol{\theta}}_r)}{\psi_G(z_i; \hat{\Theta})},$$

para $r = 1, \dots, K$ e $w = 1, \dots, q$.

Particularizando para o caso de MF-PN temos as seguintes expressões depois de algumas simplificações matemáticas (detalhes das contas podem ser vistos no Seção 3.7).

$$\hat{u}_{i,\pi_r} = \frac{\hat{y}_{ir}}{\hat{\pi}_r} - \frac{\hat{y}_{iK}}{\hat{\pi}_K}, \quad \hat{u}_{i,\alpha_r} = \hat{y}_{ir} \times \left(\frac{1}{\hat{\alpha}_r} + \ln \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right),$$

$$\hat{u}_{i,\xi_r} = \frac{\hat{y}_{ir}}{\hat{\eta}_r} \times \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} + (1 - \hat{\alpha}_r) \times R_I \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right)$$

e

$$\hat{u}_{i,\eta_r} = \frac{\hat{y}_{ir}}{\hat{\eta}_r} \times \left(-1 + \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right)^2 + (1 - \hat{\alpha}_r) \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \times R_I \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right),$$

com \hat{y}_{ir} e $R_I(\cdot)$ são definidos em (3.17) e (2.14), respectivamente.

3.3.5 Critério de classificação

A i -ésima observação será classificada no grupo j se

$$\hat{y}_{ij} = \max \{ \hat{y}_{il} : l = 1, \dots, K \},$$

em que

$$\hat{y}_{ij} = \frac{\hat{\pi}_j \frac{\hat{\alpha}_j}{\hat{\eta}_j} \phi \left(\frac{z_i - \hat{\xi}_j}{\hat{\eta}_j} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_j}{\hat{\eta}_j} \right) \right\}^{\hat{\alpha}_j - 1}}{\sum_{m=1}^K \hat{\pi}_m \frac{\hat{\alpha}_m}{\hat{\eta}_m} \phi \left(\frac{z_i - \hat{\xi}_m}{\hat{\eta}_m} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_m}{\hat{\eta}_m} \right) \right\}^{\hat{\alpha}_m - 1}}, \quad (3.17)$$

$\hat{\xi}_m$, $\hat{\eta}_m$, $\hat{\alpha}_m$ e $\hat{\pi}_m$ denotam as estimativas de máxima verossimilhança obtidas mediante o algoritmo EM para os parâmetros da componente ou grupo m . Note que \hat{y}_{ij} é a estimativa de MV para y_{ij} obtida via algoritmo EM.

3.3.6 Critérios de seleção de modelos

Os três critérios de seleção de modelos que serão utilizados neste trabalho são: o Critério de Informação de Akaike (AIC) (Akaike, 1974), o Critério de Informação Bayesiano (BIC) (Schwarz *et al.*, 1978) e o Critério de Determinação Eficiente (EDC) (Bai *et al.*, 1989). A forma geral destes critérios é dada por

$$-2l(\hat{\Theta}) + \gamma c_n,$$

sendo $l(\cdot)$ a função de log-verossimilhança para o modelo MF-AP definida em (3.9), γ o número de parâmetros do modelo e c_n um termo de penalização que varia de acordo com o critério de informação selecionado.

Para os critérios AIC e BIC, o termo de penalização c_n é igual a 2 e $\log(n)$, respectivamente. No caso do critério EDC, o valor c_n deve ser escolhido tal que as condições $c_n/n \rightarrow 0$ e $c_n/(\log(\log(n))) \rightarrow 0$ quando $n \rightarrow \infty$ sejam satisfeitas. Nós usaremos $c_n = 0.2\sqrt{n}$, assim como foi considerado em Bai *et al.* (1989).

3.4 Estudo de simulação

Nesta seção são apresentados os estudos de simulação realizados para estudar o viés e o erro quadrático médio das probabilidades de mistura e os três primeiros momentos de cada componente de mistura de uma MF-PN, calculados com as estimativas obtidas pelo algoritmo EM para vários tamanhos amostrais. Por outro lado, foi de interesse determinar a capacidade dos modelos de MF-PN para classificar observações provenientes de misturas finitas de distribuições robustas pertencentes às FM-SMSN.

3.4.1 Estudo do viés e do erro quadrático médio

Nesta subseção foi analisado o comportamento das probabilidades de mistura e os momentos de cada componente de mistura (no lugar dos parâmetros) calculados mediante as estimativas de máxima verossimilhança obtidas através do algoritmo EM sobre diferentes tamanhos amostrais, para a classe de mistura finita com 2 e 3 componentes de α potência normal no caso localização - escala.

As medidas de avaliação utilizadas foram o viés (bias), o viés relativo (VR), o erro quadrado médio (EQM) e o erro absoluto médio relativo (EAMR). Os tamanhos das amostras foram $n = 200, 500, 1.000, 2.000, 3.000$ e 5.000 , e os valores dos parâmetros fixados em:

- $K = 2$: $\xi_1 = 2, \xi_2 = 4, \eta_1 = \eta_2 = 1, \alpha_1 = 5, \alpha_2 = 25$ e $\pi_1 = 0, 35$;
- $K = 3$: $\xi_1 = 7, 5, \xi_2 = 10, 0, \xi_3 = 12, 5, \eta_1 = \eta_2 = \eta_3 = 1, \alpha_1 = 2, \alpha_2 = 10, \alpha_3 = 40,$
 $\pi_1 = 0, 25$ e $\pi_2 = 0, 30$;

com o objetivo de explorar um padrão de misturas finitas com dois e três componentes de misturas distinguíveis, respectivamente.

Os valores dos três primeiros momentos para cada componente de mistura correspondentes aos conjuntos de parâmetros apresentados anteriormente são aproximadamente:

- $K = 2$: $E(X)_1 = 3,1631$, $E(X^2)_1 = 10,4526$, $E(X^3)_1 = 35,9832$, $E(X)_2 = 5,9640$, $E(X^2)_2 = 35,8289$, $E(X^3)_2 = 216,8479$;
- $K = 3$: $E(X)_1 = 8,0643$, $E(X^2)_1 = 65,7145$, $E(X^3)_1 = 541,0125$, $E(X)_2 = 11,5395$, $E(X^2)_2 = 133,5033$, $E(X^3)_2 = 1548,5726$, $E(X)_3 = 14,6619$, $E(X^2)_3 = 215,1976$, $E(X^3)_3 = 3161,8937$;

em que $E(X^l)_j$ representa o l -ésimo momento da j -ésimo momento de mistura.

Para cada conjunto de parâmetros e tamanho amostral foram geradas $m = 1.000$ amostras do modelo MF-PN. Em seguida, estimamos os parâmetros do modelo com o algoritmo EM desenvolvido e calculamos os três primeiros momentos de cada componente. Posteriormente, o viés, o VR, o EQM e o EAMR foram obtidos para todas as amostras. Para $E(X)_j$, $j = 1, \dots, K$, eles são definidos como

$$\text{viés} = \frac{1}{m} \sum_{i=1}^m \left(\widehat{E(X)}_j^{(i)} - E(X)_j^{(i)} \right), \quad \text{VR} = \frac{1}{m} \sum_{i=1}^m \frac{\left(\widehat{E(X)}_j^{(i)} - E(X)_j^{(i)} \right)}{E(X)_j^{(i)}},$$

$$\text{EQM} = \frac{1}{m} \sum_{i=1}^m \left(\widehat{E(X)}_j^{(i)} - E(X)_j^{(i)} \right)^2 \quad \text{e} \quad \text{EAMR} = \frac{1}{m} \sum_{i=1}^m \left| \frac{\widehat{E(X)}_j^{(i)} - E(X)_j^{(i)}}{E(X)_j^{(i)}} \right|,$$

respectivamente, sendo $\widehat{E(X)}_j^{(i)}$ a estimativa de $E(X)_j$ quando os dados correspondem à amostra i . As definições para os outros momentos são obtidos por analogia.

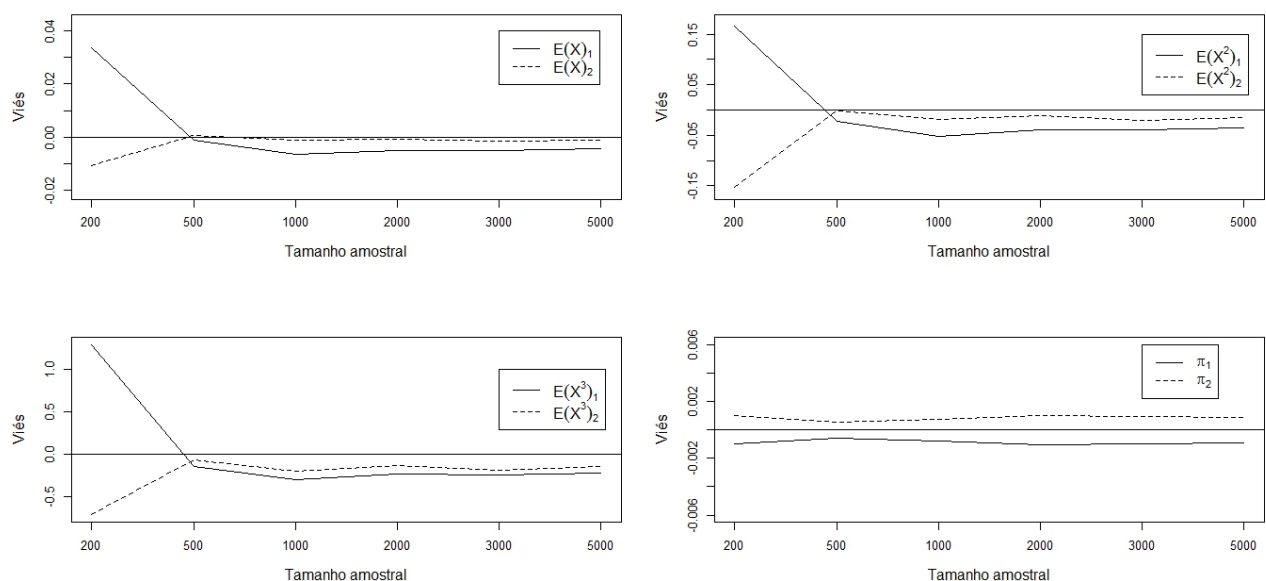
Na Tabela 3.1 são apresentados o viés e o VR para as estimativas dos momentos e das probabilidades de misturas para o caso $K = 2$. Pode ser observado que os vieses e os vieses relativos são pequenos para todos os tamanhos amostrais considerados. Para os momentos de cada componente de mistura, quando n aumenta estas medidas diminuem, no entanto,

para as probabilidades de misturas este comportamento não é observado, embora, os vieses sejam próximos de zero para $n \geq 200$.

Tabela 3.1: Viés e VR.

Medida	Parâmetros	n					
		200	500	1.000	2.000	3.000	5.000
viés	$E(X)_1$	3,3523e-02	-1,3266e-03	-6,5741e-03	-4,9437e-03	-4,9415e-03	-4,2402e-03
	$E(X^2)_1$	1,6523e-01	-2,1403e-02	-5,1625e-02	-3,9133e-02	-3,9742e-02	-3,5327e-02
	$E(X^3)_1$	1,2794e+00	-1,4949e-01	-3,0549e-01	-2,3937e-01	-2,4551e-01	-2,2384e-01
	$E(X)_2$	-1,0849e-02	7,8541e-04	-1,2386e-03	-7,0575e-04	-1,5124e-03	-1,0785e-03
	$E(X^2)_2$	-1,5309e-01	-1,1333e-03	-1,9073e-02	-1,1661e-02	-1,9661e-02	-1,4457e-02
	$E(X^3)_2$	-7,1643e-01	-7,6036e-02	-2,0769e-01	-1,3496e-01	-1,9666e-01	-1,4840e-01
	π_1	-1,0047e-03	-5,8304e-04	-7,5406e-04	-1,0300e-03	-9,5287e-04	-9,0463e-04
VR	$E(X)_1$	1,0600e-02	-4,2752e-04	-2,0817e-03	-1,5647e-03	-1,5635e-03	-1,3416e-03
	$E(X^2)_1$	1,5834e-02	-2,0668e-03	-4,9490e-03	-3,7494e-03	-3,8066e-03	-3,3841e-03
	$E(X^3)_1$	3,5655e-02	-4,1901e-03	-8,5100e-03	-6,6630e-03	-6,8322e-03	-6,2308e-03
	$E(X)_2$	-1,8249e-03	1,3706e-04	-2,0561e-04	-1,1530e-04	-2,4978e-04	-1,7686e-04
	$E(X^2)_2$	-4,2999e-03	-1,8255e-05	-5,2676e-04	-3,1753e-04	-5,3850e-04	-3,9296e-04
	$E(X^3)_2$	-3,3045e-03	-3,2710e-04	-9,4959e-04	-6,1091e-04	-8,9117e-04	-6,6827e-04
	π_1	-2,8705e-03	-1,6658e-03	-2,1544e-03	-2,9428e-03	-2,7225e-03	-2,3893e-03

Adicionalmente, na Figura 3.1 pode-se observar que os estimadores dos momentos de cada componente são viesados. Além disso, para $n \geq 500$, o viés de cada momento é sempre negativo, o que indica que o estimador subestima o verdadeiro valor do momento. No entanto, os vieses relativos para $n \geq 500$ são menores que 0,005.

Figura 3.1: Viés para o caso $K = 2$.

Para o caso $K = 2$ o EQM e o EAMR são apresentados na Tabela 3.2. Pode-se observar

que para $n \geq 500$ os EAMR são menores que 0,05. De forma geral, observa-se que estas duas medidas tendem a zero quando o tamanho amostral aumenta. Além disso, o comportamento da medida EAMR é ilustrado na Figura 3.2.

Tabela 3.2: EQM e EAMR para o caso $K = 2$.

Medida	Parâmetros	n					
		200	500	1.000	2.000	3.000	5.000
EQM	$E(X)_1$	1,3176e-01	6,4370e-03	1,7870e-03	8,6867e-04	6,0646e-04	3,6382e-04
	$E(X^2)_1$	2,5231e+00	2,6138e-01	8,1743e-02	4,0862e-02	2,8972e-02	1,7568e-02
	$E(X^3)_1$	1,2334e+02	7,4640e+00	2,4468e+00	1,2461e+00	8,9415e-01	5,4948e-01
	$E(X)_2$	2,5384e-01	1,5144e-02	7,0678e-04	3,5248e-04	2,6697e-04	1,6101e-04
	$E(X^2)_2$	9,6900e+00	6,1783e-01	8,7667e-02	4,5168e-02	3,3175e-02	1,9868e-02
	$E(X^3)_2$	3,8990e+02	3,3691e+01	6,6689e+00	3,4923e+00	2,5010e+00	1,4918e+00
	π_1	1,2839e-03	4,9146e-04	2,4691e-04	1,1947e-04	7,4387e-05	4,8864e-05
EAMR	$E(X)_1$	4,6749e-02	1,7101e-02	1,0607e-02	7,4385e-03	6,2063e-03	4,8425e-03
	$E(X^2)_1$	7,4232e-02	3,3547e-02	2,1865e-02	1,5530e-02	1,3053e-02	1,0163e-02
	$E(X^3)_1$	1,2624e-01	5,2384e-02	3,4732e-02	2,4979e-02	2,1128e-02	1,6559e-02
	$E(X)_2$	2,1407e-02	6,0779e-03	3,3920e-03	2,3993e-03	2,0653e-03	1,6383e-03
	$E(X^2)_2$	2,7689e-02	1,0216e-02	6,4132e-03	4,5687e-03	3,8778e-03	3,0472e-03
	$E(X^3)_2$	3,4753e-02	1,4461e-02	9,3434e-03	6,6752e-03	5,6131e-03	4,3822e-03
	π_1	8,0160e-02	5,0144e-02	3,5204e-02	2,4414e-02	1,9446e-02	1,6074e-02

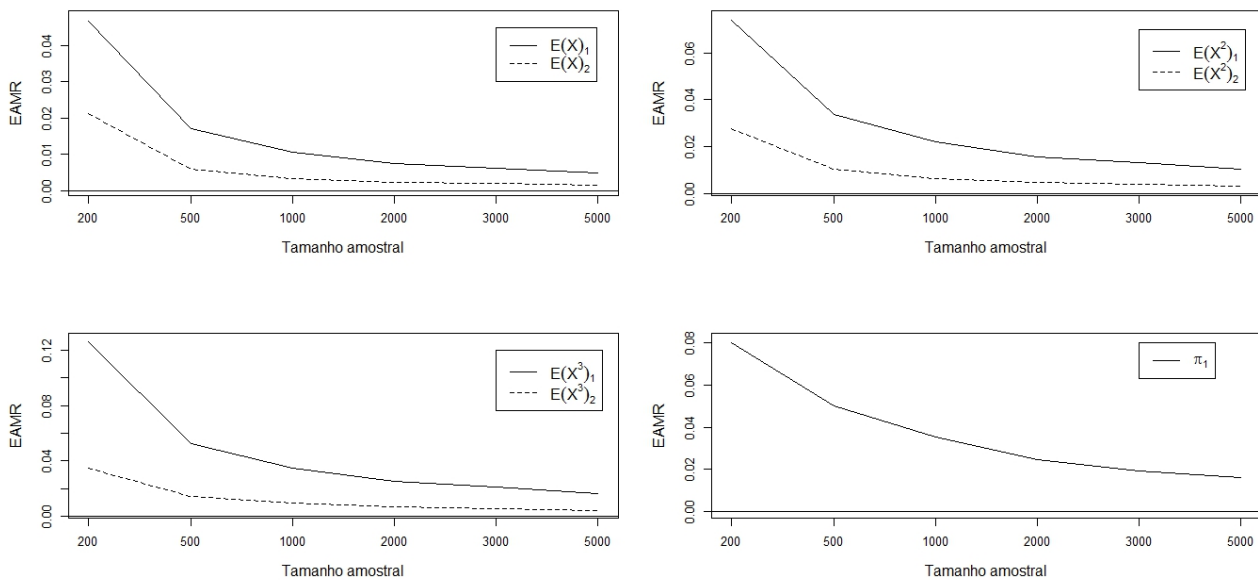


Figura 3.2: EAMR para o caso $K = 2$.

Para o estudo de simulação da classe de mistura finita de três componentes de potência normal no caso localização-escala, os resultados obtidos do viés, o VR, o EQM e o EAMR

são apresentados nas Tabelas 3.3 e 3.4. Pode-se notar que o comportamento destas medidas é similar ao caso $K = 2$.

Tabela 3.3: VR, EQM e EAMR para o caso $K = 3$.

Medida	Parâmetros	n						
		200	500	1.000	2.000	3.000	5.000	
viés	$E(X)_1$	1,2119e-01	3,1629e-02	3,1254e-03	-9,7028e-04	-3,0115e-03	-2,8755e-03	
	$E(X)_2$	3,0225e-03	-8,2716e-03	-4,2282e-03	-3,4812e-03	-4,1786e-03	-3,1166e-03	
	$E(X)_3$	4,6816e-02	8,5944e-03	-1,8859e-03	-9,9424e-04	-6,6866e-04	-1,0975e-04	
	$E(X^2)_1$	1,6880e+00	4,8037e-01	3,7892e-02	-2,1527e-02	-5,5824e-02	-5,3056e-02	
	$E(X^2)_2$	-7,6785e-02	-1,7136e-01	-1,0035e-01	-8,3820e-02	-9,9274e-02	-7,5202e-02	
	$E(X^2)_3$	6,9780e-01	1,6082e-01	-5,7482e-02	-3,1459e-02	-2,4072e-02	-7,2952e-03	
	$E(X^3)_1$	2,1371e+01	5,3490e+00	3,2573e-01	-3,4141e-01	-7,8094e-01	-7,3802e-01	
	$E(X^3)_2$	1,9979e+00	-2,6082e+00	-1,7991e+00	-1,5253e+00	-1,7844e+00	-1,3718e+00	
	$E(X^3)_3$	7,4765e+00	2,2444e+00	-1,3316e+00	-7,5022e-01	-6,3058e-01	-2,5282e-01	
	π_1	-5,1552e-03	-1,1549e-03	-7,2719e-04	-4,6546e-04	-3,6644e-04	-3,2495e-04	
	π_2	6,4627e-03	1,6113e-03	7,8584e-04	7,5028e-05	-9,1642e-05	-1,7241e-04	
	VR	$E(X)_1$	1,5024e-02	3,9231e-03	3,8717e-04	-1,2066e-04	-3,7365e-04	-3,5670e-04
		$E(X)_2$	2,5954e-04	-7,1557e-04	-3,6631e-04	-3,0163e-04	-3,6196e-04	-2,6996e-04
		$E(X)_3$	3,1947e-03	5,8380e-04	-1,2727e-04	-6,6377e-05	-4,3851e-05	-5,7333e-06
$E(X^2)_1$		2,5690e-02	7,3152e-03	5,7507e-04	-3,2897e-04	-8,5032e-04	-8,0788e-04	
$E(X^2)_2$		-5,8139e-04	-1,2798e-03	-7,5121e-04	-6,2769e-04	-7,4308e-04	-5,6294e-04	
$E(X^2)_3$		3,2480e-03	7,3952e-04	-2,6234e-04	-1,4128e-04	-1,0585e-04	-2,7930e-05	
$E(X^3)_1$		3,9482e-02	9,9027e-03	5,9859e-04	-6,3420e-04	-1,4454e-03	-1,3653e-03	
$E(X^3)_2$		1,2795e-03	-1,6777e-03	-1,1607e-03	-9,8478e-04	-1,1513e-03	-8,8524e-04	
$E(X^3)_3$		2,3745e-03	6,9792e-04	-4,1170e-04	-2,2782e-04	-1,8779e-04	-6,8472e-05	
π_1		-2,0621e-02	-4,6196e-03	-2,9088e-03	-1,8619e-03	-1,4658e-03	-1,2998e-03	
π_2		2,1542e-02	5,3711e-03	2,6195e-03	2,5009e-04	-3,0547e-04	-5,7470e-04	

Tabela 3.4: EQM e EAMR para o caso $K = 3$.

Medida	Parâmetros	n						
		200	500	1.000	2.000	3.000	5.000	
EQM	$E(X)_1$	1,0203e+00	5,5942e-02	9,3493e-03	2,7616e-03	1,2860e-03	6,6458e-04	
	$E(X)_2$	5,6501e-01	4,2897e-02	1,5229e-03	7,6780e-04	5,1090e-04	3,0664e-04	
	$E(X)_3$	1,5276e+00	2,4401e-01	2,5237e-03	7,7005e-04	4,6878e-04	3,2528e-04	
	$E(X^2)_1$	6,7105e+01	1,8527e+01	2,4757e+00	7,3939e-01	3,4738e-01	1,8277e-01	
	$E(X^2)_2$	1,1079e+02	1,3536e+01	7,9830e-01	4,0894e-01	2,7228e-01	1,6390e-01	
	$E(X^2)_3$	7,3833e+02	1,2690e+02	1,7450e+00	5,8681e-01	3,6572e-01	2,5108e-01	
	$E(X^3)_1$	1,0048e+04	4,0160e+03	3,7813e+02	1,1472e+02	5,4698e+01	2,9346e+01	
	$E(X^3)_2$	2,1917e+04	2,7819e+03	2,3971e+02	1,2413e+02	8,2682e+01	4,9929e+01	
	$E(X^3)_3$	2,2777e+05	4,0408e+04	7,0949e+02	2,5677e+02	1,6308e+02	1,1062e+02	
	π_1	7,4634e-04	3,1203e-04	1,7057e-04	8,7930e-05	5,6837e-05	3,6980e-05	
	π_2	8,0136e-04	3,9484e-04	2,0861e-04	1,0623e-04	7,6220e-05	4,3938e-05	
	EAMR	$E(X)_1$	3,9816e-02	1,3759e-02	6,9680e-03	4,3610e-03	3,4058e-03	2,5260e-03
		$E(X)_2$	1,8460e-02	5,0929e-03	2,6631e-03	1,9261e-03	1,5659e-03	1,2051e-03
		$E(X)_3$	2,2764e-02	6,2847e-03	1,9225e-03	1,3033e-03	1,1296e-03	9,3999e-04
$E(X^2)_1$		6,0680e-02	2,7810e-02	1,3964e-03	8,8574e-03	6,9043e-03	5,1367e-03	
$E(X^2)_2$		2,8784e-02	9,4814e-03	5,2928e-03	3,8486e-03	3,1347e-03	2,4071e-03	
$E(X^2)_3$		3,6435e-02	1,0787e-02	3,6764e-03	2,5056e-03	2,1674e-03	1,7881e-03	
$E(X^3)_1$		8,9894e-02	4,2758e-02	2,1160e-02	1,3572e-02	1,0572e-02	7,8929e-03	
$E(X^3)_2$		3,7726e-02	1,3586e-02	7,9188e-03	5,7902e-03	4,7208e-03	3,6188e-03	
$E(X^3)_3$		4,5743e-02	1,4322e-02	5,3161e-03	3,6379e-03	3,1348e-03	2,5695e-03	
π_1		8,4733e-02	5,6917e-02	4,1660e-02	2,9510e-02	2,4355e-02	1,9454e-02	
π_2		7,3320e-02	5,3269e-02	3,8657e-02	2,7266e-02	2,2899e-02	1,7655e-02	

Os resultados obtidos nesta subseção para as estimativas das probabilidades de mistura e

dos momentos de cada componente para a MF-PN são satisfatórios. Porém, é recomendado um tamanho amostral $n \geq 500$ para obter um padrão razoável de convergência, mesmo que os resultados obtidos para $n = 200$ não sejam ruins.

3.4.2 Agrupamento - Clustering.

Uma das principais características das misturas finitas de distribuições é o agrupamento ou classificação de observações, isto é, a capacidade de juntar observações que são semelhantes. Portanto, primeiro vamos estudar a habilidade dos modelos de MF-PN para agrupar observações. A taxa de localização correta foi a medida utilizada para avaliar o nível de classificação de cada modelo, isto é, a proporção de vezes que este identifica as observações no grupo correspondente.

Primeiro geramos $m = 500$ amostras de tamanho n de cada uma das diferentes misturas finitas robustas de distribuições: normal assimétrica (MF-SN), slash assimétrica (MF-SSL), normal contaminada assimétrica (MF-SNC) e t assimétrica (MF-ST); as quais pertencem às misturas finitas de distribuições de mistura de escala da normal assimétrica. Para gerar as amostras das FM-SMSN foi utilizada a função `rmix()` do pacote `mixsmsn` (Prates *et al.*, 2013).

A classificação verdadeira para cada uma das amostras geradas foi ignorada no momento de ajustar os modelos de MF-N, MF-SN, MF-T, MF-ST e MF-PN. Posteriormente, classificamos as observações usando as probabilidades *a posteriori* estimadas \hat{y}_{ij} (Ver subseção 3.3.5). Para cada amostra l e cada mistura finita, com $l = 1, \dots, m$, calculamos a taxa de localização ou classificação correta. Esta taxa representa a proporção de indivíduos classificados corretamente e será denotada por r_l .

Fixamos os seguintes valores para os parâmetros dos modelos de MF-SN, MF-SSL e MF-ST em $\mu_1 = 7$, $\mu_2 = 12$, $\eta_1 = 6$, $\eta_2 = 4$, $\lambda_1 = 5$, $\lambda_2 = 7$, $\pi_i = 0.8$ e $\nu_1 = \nu_2 = 3$, e para a MF-SNC fixamos $\nu_1 = \nu_2 = (0, 3; 0, 3)$. Os tamanhos das amostras considerados foram $n = 200, 500, 1.000$ e 1.500 . Na Figura 3.3 apresentamos os histogramas de uma amostra tomada de cada população das misturas finitas de distribuições de probabilidade selecionadas para ser estudadas com um tamanho amostral $n = 1.000$.

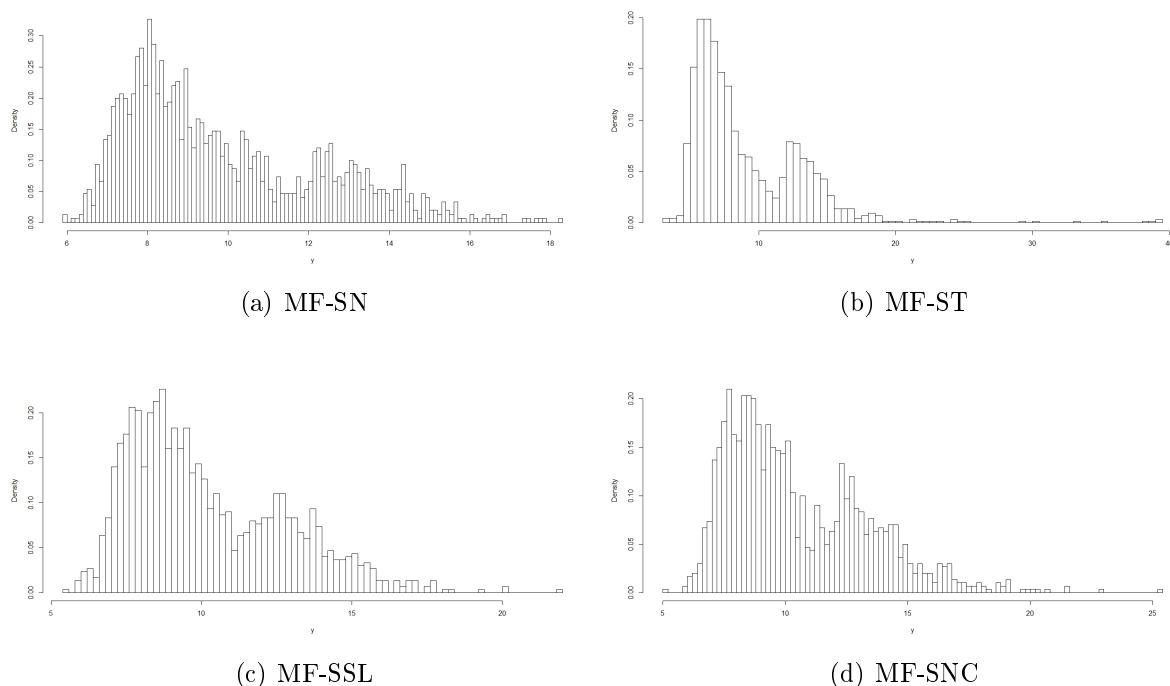


Figura 3.3: Dados simulados para MF-SN, MF-ST, MF-SSL e MF-SNC ($n = 1000$).

Na Tabela 3.5 são apresentados os valores médios das taxas de localização correta, isto é, $\frac{1}{m} \sum_{l=1}^m r_l$. Analisando os resultados obtidos podemos observar que o modelo de MF-PN classifica de forma adequada observações provenientes de outras misturas finitas robustas ou flexíveis.

Tabela 3.5: Taxas médias de localizações para cada MF.

Modelo verdadeiro	n	Modelo ajustado				
		MF-N	MF-SN	MF-T	MF-ST	MF-PN
MF-SN	200	0,8442	0,9104	0,8471	0,9146	0,9138
	500	0,8494	0,9189	0,8517	0,9226	0,9330
	1.000	0,8473	0,9205	0,8496	0,9240	0,9403
	1.500	0,8476	0,9214	0,8509	0,9245	0,9434
MF-SSL	200	0,7692	0,8425	0,7749	0,8523	0,8410
	500	0,7640	0,8436	0,7726	0,8568	0,8630
	1.000	0,7619	0,8458	0,7698	0,8605	0,8719
	1.500	0,7600	0,8462	0,7682	0,8616	0,8742
MF-SNC	200	0,7579	0,8348	0,7660	0,8450	0,8205
	500	0,7546	0,8353	0,7634	0,8530	0,8449
	1.000	0,7525	0,8364	0,7600	0,8566	0,8518
	1.500	0,7514	0,8367	0,7588	0,8587	0,8814
MF-ST	200	0,8223	0,8600	0,8364	0,9263	0,8656
	500	0,8157	0,8460	0,8379	0,9370	0,8717
	1.000	0,8122	0,8296	0,8380	0,9389	0,8686
	1.500	0,8114	0,8179	0,8385	0,9386	0,8670

Outros cenários de simulação foram desenvolvidos com diferentes valores dos parâmetros.

Nestes encontramos comportamentos similares para as taxas de classificação correta, porém, obtivemos casos nos quais as MF-ST tiveram uma taxa de classificação um pouco maior que a MF-PN.

Por outro lado, observou-se que as estimativas das probabilidades de misturas para as MF-N, MF-SN e MF-T diferiam muito dos valores verdadeiros nos diferentes estudos de simulação desenvolvidos, no entanto, as estimativas obtidas com as MF-ST e MF-PN foram próximas dos verdadeiros valores. Assim, como as probabilidades *a posteriori* estimadas (\hat{y}_{ij}) depende das estimativas das probabilidades de misturas ($\hat{\pi}_j$), poderia ser justificado o fato de ter taxas de classificação superiores às obtidas pela distribuição verdadeira (no caso da MF-SN) ao classificar usando as MF-ST e MF-PN.

Dois casos foram explorados para estudar a taxa de classificação correta quando os dados são gerados do modelo de MF-PN. Estes casos dependem da característica de identificação das componentes de mistura, isto é, se as componentes das misturas podem ser distinguíveis e ou não. Os valores dos parâmetros usados para avaliar o nível de classificação nestes dois casos foram:

- *Componentes distinguíveis*: $\mu_1 = 2$, $\mu_2 = 28$, $\eta_1 = 10$, $\eta_2 = 6$, $\alpha_1 = 25$, $\alpha_2 = 5$ e $\pi_i = 0,8$.
- *Componentes não distinguíveis*: $\mu_1 = 5$, $\mu_2 = 12$, $\eta_1 = 6$, $\eta_2 = 4$, $\alpha_1 = 7$, $\alpha_2 = 5$ e $\pi_i = 0,8$.

Na Figura 3.4 são apresentados histogramas de uma das amostras tomada para os casos de componentes distinguíveis e não.

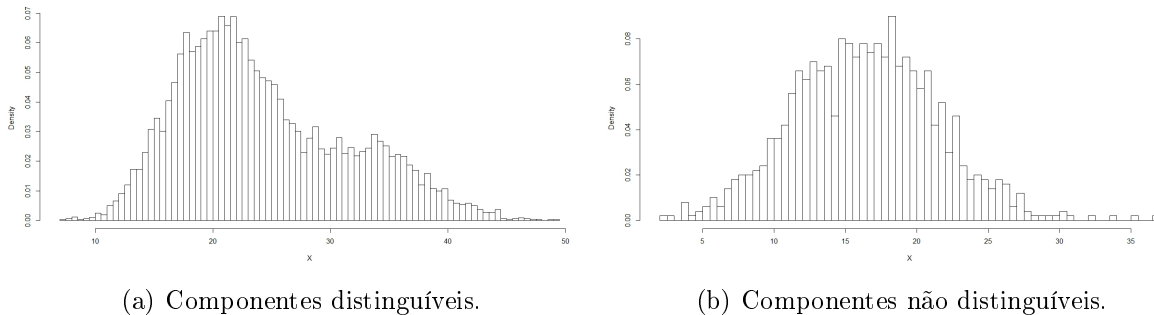


Figura 3.4: Dados simulados para uma MF-PN com dois componentes.

A Tabela 3.6 apresenta os valores médios das taxas de localização correta para os dois casos estudados quando geramos as observações do modelo de MF-PN. Como era esperado, o modelo de MF-PN classifica de forma mais adequada as observações (em ambos casos) em relação às outras misturas finitas robustas utilizadas.

Tabela 3.6: Taxas médias de localizações o modelo de MF-PN.

Modelo verdadeiro	n	Modelo ajustado				
		MF-N	MF-SN	MF-T	MF-ST	MF-PN
MF-PN distinguíveis	500	0,8873	0,8842	0,8875	0,8869	0,9159
	1.000	0,8897	0,8865	0,8901	0,8888	0,9224
	1.500	0,7514	0,8367	0,7585	0,8586	0,8814
MF-PN não distinguíveis	500	0,6415	0,6669	0,6576	0,6672	0,7755
	1.000	0,6501	0,6666	0,6590	0,6679	0,7950
	1.500	0,6518	0,6703	0,6623	0,6702	0,8001

3.5 Ilustração

O índice de massa corporal (IMC) é definido como a razão entre o peso corporal (quilogramas) e o quadrado da altura do corpo (metros quadrados). Nesta seção consideramos o conjunto de dados de IMC da Pesquisa Nacional de Saúde e Nutrição. Esta pesquisa foi feita pelo Centro Nacional para Estatísticas de Saúde (NCHS) do Centro de Controle de Doenças (CDC) realizada nos Estados Unidos e apresentada nos relatórios de 1999-2000 e 2001-2002. Os dados foram utilizados anteriormente por [Lin *et al.* \(2007a\)](#). Eles consideraram os participantes que têm peso dentro (39.50 kg, 70.00 kg) e (95.01 kg, 196.81 kg), permitindo-lhes explorar um padrão de mistura. Dos 4579 participantes originais, foram considerados um total de 2123, onde o primeiro grupo tinha 1069 participantes e o segundo 1054. Para mais detalhes ver [Lin *et al.* \(2007a\)](#) e [Basso *et al.* \(2010\)](#).

Para comparar com o modelo de MF-PN, ajustamos os modelos de misturas finitas de normais, t, skew-normal e skew-t com dois componentes usando a função `smsn.mix()` do pacote `mixsmsn`. Na Figura 3.5 apresentamos o histograma deste conjunto de dados, junto com o ajuste dos modelos MF-SN, MF-N e MF-PN. Na Figura 3.6 são apresentados os

ajustes do modelos MF-SN, MF-N e MF-PN.

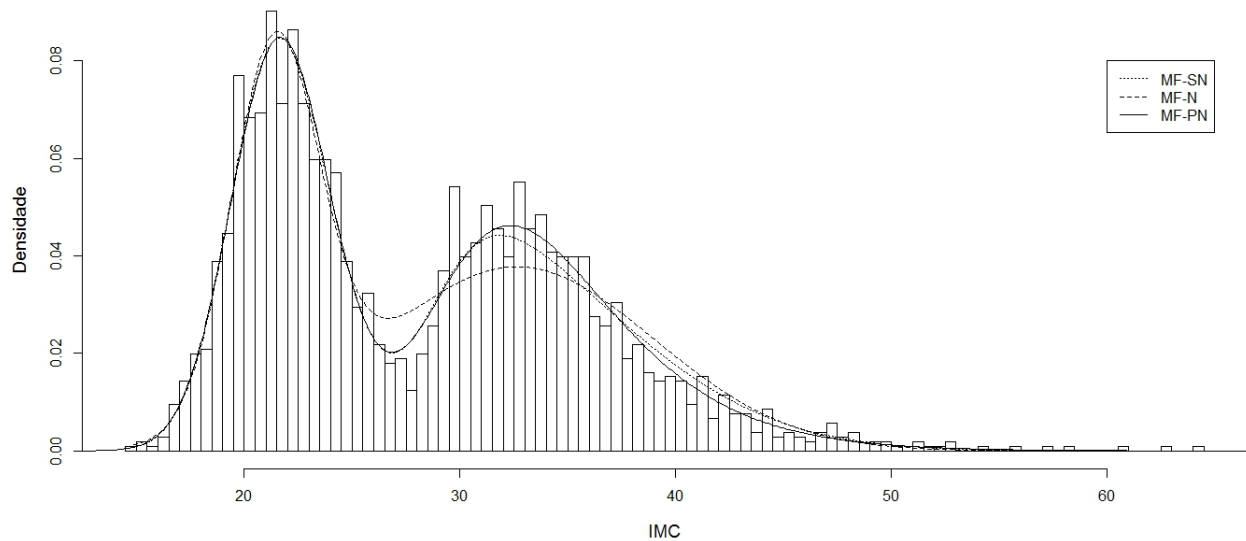


Figura 3.5: Ajuste das distribuições MF-SN, MF-N e MF-PN com a resposta do IMC.

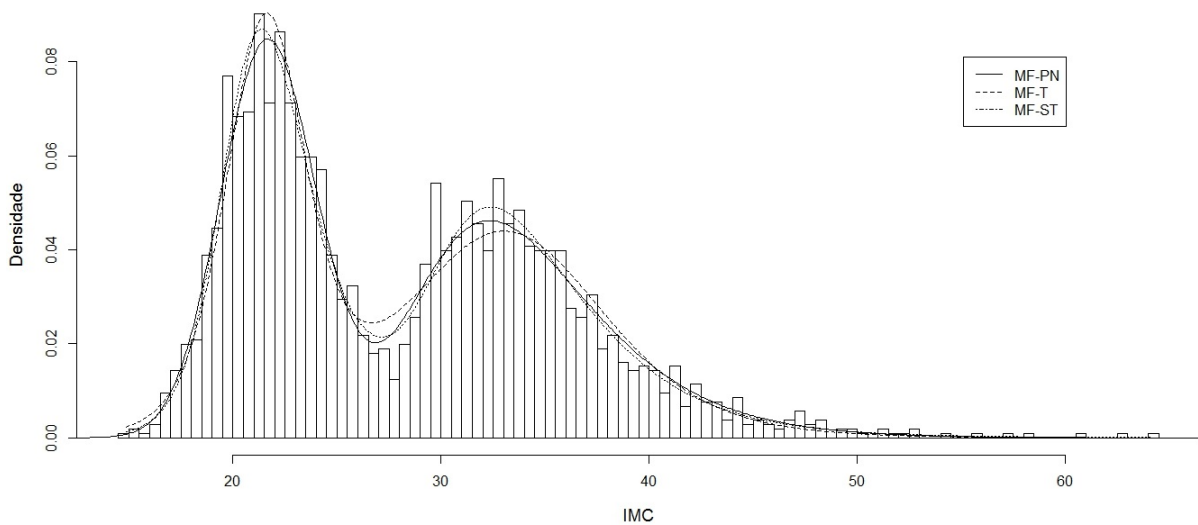


Figura 3.6: Ajuste das distribuições MF-ST, MF-T e MF-PN com a resposta do IMC.

Com base na visualização gráfica obtida mediante as Figuras 3.5 e 3.6 pode ser observado que o modelo MF-PN ajusta melhor os dados do IMC que o modelo MF-N, principalmente os indivíduos pertencentes à segunda componente da mistura. No entanto, ao comparar este modelo com o modelo de MF-SN só é possível perceber levemente um melhor ajuste na segunda componente de mistura. Os modelos MF-T e MF-ST ajustam melhor os dados do IMC pertencentes à primeira componente da mistura que o modelo MF-PN, mas em relação

à segunda componente o modelo de MF-PN ajusta melhor que o modelo MF-T.

Na Tabela 3.7 são apresentadas as estimativas e seus erros padrões (calculados utilizando a técnica de reamostragem Jackknife) correspondentes dos dois primeiros momentos e as probabilidades de mistura para as cinco misturas finitas de distribuições ajustadas para os dados de IMC. Pode-se observar que todos os erros padrões (E_p) são pequenos, então o MLE do parâmetro e as probabilidades de mistura foram obtidos com precisão.

Tabela 3.7: Estimativas dos dois primeiros momentos e das probabilidades de mistura para os dados de IMC.

Momento		MF-N	MF-T	MF-SN	MF-ST	MF-PN
$E(X)_1$	EMV	21,4283	21,5771	21,8277	22,1666	21,8066
	E_p	0,0019	0,0053	0,0102	0,0102	0,0013
$E(X)_2$	EMV	32,6384	33,0509	34,5665	34,5779	34,1573
	E_p	0,0130	0,0128	0,0216	0,0165	0,0613
$E(X^2)_1$	EMV	463,3257	469,4163	482,0053	499,8902	480,9756
	E_p	0,0843	0,2599	0,4878	0,5131	0,5439
$E(X^2)_2$	EMV	1105,8264	1116,6890	1219,9811	1222,0320	1218,9307
	E_p	0,7648	0,7252	1,3770	1,0715	1,4616
π_1	EMV	0,3973	0,4374	0,4954	0,4833	0,4933
	E_p	0,0006	0,0011	0,0012	0,0010	0,0002

Os critérios de AIC, BIC e EDC foram calculados para a comparação dos modelos ajustados ao conjunto de dados do IMC e os valores obtidos são apresentados na Tabela 3.8. Nesta tabela pode-se observar que os três critérios tiveram um comportamento semelhante. O modelo de MF-ST apresenta o melhor ajuste segundo cada critério. Além disso, é importante observar que estes valores são muito próximos aos obtidos com o modelo de MF-PN, afastando-se dos outros modelos ajustados.

Tabela 3.8: Critérios de seleção de modelos para o conjunto de dados do IMC.

Modelo	MF-N	MF-T	MF-SN	MF-ST	MF-PN
AIC	13926,91	13878,73	13858,58	13833,38	13834,50
BIC	13955,21	13907,03	13898,20	13873,01	13874,12
EDC	13962,99	13914,80	13909,09	13883,89	13885,00

Finalmente, na Tabela 3.9 apresentamos as taxas de classificação correta para todos os modelos ajustados. Para este conjunto de dados, os modelos de MF-SN e MF-PN tiveram a

melhor taxa (0,9840).

Tabela 3.9: Taxas de classificação correta.

MF-N	MF-T	MF-SN	MF-ST	MF-PN
0,9463	0,9670	0,9840	0,9816	0,9840

As observações que estão entre as duas componentes de mistura são mais propensas a ser mal classificadas. Nas Figuras 3.5 e 3.6 pode-se observar os modelos de MF-SN e MF-PN ajustaram aparentemente melhor esta região que o modelo de MF-ST, assim as mesmas obtiveram taxas de classificação um pouco maior.

3.6 Considerações finais do capítulo

Neste capítulo foi proposta uma classe flexível de modelos usando misturas finitas de distribuições da classe α potência. O modelo permite analisar dados na presença de multimodalidade, assimetria, excesso de curtose e/ou observações atípicas, sendo assim uma ferramenta muito útil para profissionais de diferentes áreas como biologia, engenharia, medicina, entre outras.

Os estudos de simulação realizados têm mostrado que o algoritmo EM fornece estimativas razoáveis para os momentos do modelo de misturas finitas de K componentes de distribuições α potência normal para o caso localização - escala. Além disso, foi observada que a capacidade do modelo de FM-PN para o agrupamento ou classificação de dados pertencentes a distribuições robustas tais como os modelos MF-SN, MF-SSL, MF-SNC e MF-ST é boa, dado que foram obtidas altas taxas de classificação correta.

Os modelos de MF-N, MF-T e MF-SN podem ser tratados como casos especiais do modelo de MF-ST. Este modelo apresenta vantagens no ajuste de dados pertencentes a uma ou várias componentes de mistura que apresentam simultaneamente assimetria e caudas pesadas em comparação com os casos especiais. A proposta apresentada neste trabalho ajusta de forma adequada dados que seguem o modelo de MF-ST, precisando de um número menor de parâmetros.

O algoritmo EM usado para obter as estimativas de máxima verossimilhança de maneira

iterativa proposto neste trabalho para a família potência normal, pode ser facilmente implementado em qualquer software estatístico. O software [R Core Team \(2015\)](#) foi usado para construir o pacote `mixpn`, que permite:

- Estimar os parâmetros do modelo;
- Classificar as observações;
- Desenhar o histograma e a densidade estimada;
- Calcular os valores iniciais do algoritmo EM, critérios de seleção e a matriz de informação observada;

dado um conjunto de dados e um número fixo de componentes. Adicionalmente, o pacote permite gerar observações de uma FM-PN e ajustar dados da distribuição PN. O instalador de pacote pode ser obtido entrando em contato com o autor.

3.7 Detalhes técnicos do capítulo

Nesta seção calculamos as quantidades \hat{u}_{i,ξ_r} , \hat{u}_{i,η_r} e \hat{u}_{i,α_r} para a classe de mistura finita de K componentes de α potência normal para o caso localização-escala.

$$\begin{aligned}
 \star \quad \hat{u}_{i,\xi_r} &= \frac{1}{\psi_{\phi_{LE}}(z_i; \hat{\Theta})} \times \left[\hat{\pi}_r \frac{\hat{\alpha}_r}{\hat{\eta}_r} \times \left(\phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r^2} \right) \times \right. \\
 &\times \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} + \frac{1 - \hat{\alpha}_r}{\hat{\eta}_r} \left\{ \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^2 \\
 &\times \left. \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} \right] \\
 &= \frac{1}{\psi_{\phi_{LE}}(z_i; \hat{\Theta})} \times \left[\hat{\pi}_r \frac{\hat{\alpha}_r}{\hat{\eta}_r} \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} \right] \times \\
 &\times \left[\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r^2} + \frac{1 - \hat{\alpha}_r}{\hat{\eta}_r} R_I \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right] \\
 &= \frac{\hat{y}_{ir}}{\hat{\eta}_t} \times \left[\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} + (1 - \hat{\alpha}_r) R_I \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right],
 \end{aligned}$$

com \hat{y}_{ir} e $R_I(\cdot)$ são definidos em (3.17) e (2.14), respectivamente.

$$\begin{aligned}
\star \quad \hat{u}_{i,\eta_r} &= \frac{1}{\psi_{\phi_{LE}}(z_i; \hat{\Theta})} \times \left[\hat{\pi}_r \hat{\alpha}_r \times \left(-\frac{1}{\hat{\eta}_r^2} \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} + \right. \\
&+ \frac{1}{\hat{\eta}_r} \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \left(\frac{(z_i - \hat{\xi}_r)^2}{\hat{\eta}_r^3} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} + \\
&\left. + \frac{1 - \hat{\alpha}_r}{\hat{\eta}_r} \left\{ \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^2 \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r^2} \right) \right] \\
&= \frac{1}{\psi_{\phi_{LE}}(z_i; \hat{\Theta})} \times \left[\hat{\pi}_r \frac{\hat{\alpha}_r}{\hat{\eta}_r} \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} \right] \times \\
&\times \left[-\frac{1}{\hat{\eta}_r} + \frac{(z_i - \hat{\xi}_r)^2}{\hat{\eta}_r^3} + (1 - \hat{\alpha}_r) \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r^2} \right) R_I \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right] \\
&= \frac{\hat{y}_{ir}}{\hat{\eta}_t} \times \left[-1 + \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right)^2 + (1 - \hat{\alpha}_r) \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) R_I \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right].
\end{aligned}$$

$$\begin{aligned}
\star \quad \hat{u}_{i,\alpha_r} &= \frac{1}{\psi_{\phi_{LE}}(z_i; \hat{\Theta})} \times \left[\frac{\hat{\pi}_r}{\hat{\eta}_r} \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right] \times \left[\left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} + \right. \\
&\left. + \hat{\alpha}_r \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} \ln \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right] \\
&= \frac{1}{\psi_{\phi_{LE}}(z_i; \hat{\Theta})} \times \left[\frac{\hat{\pi}_r}{\hat{\eta}_r} \phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \left\{ \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right\}^{\hat{\alpha}_r - 1} \right] \times \\
&\times \left[1 + \hat{\alpha}_r \ln \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right] \\
&= \hat{y}_{ir} \times \left(\frac{1}{\hat{\alpha}_r} + \ln \Phi \left(\frac{z_i - \hat{\xi}_r}{\hat{\eta}_r} \right) \right).
\end{aligned}$$

Capítulo 4

Inferência bayesiana para a classe de distribuições AP

Resumo

A classe de distribuições α potência (AP) é usada para a modelagem de dados em presença de assimetria, curtoses ou dados atípicos. Neste capítulo propomos uma abordagem bayesiana para a inferência dos parâmetros na classe de distribuições AP. Três distribuições *a priori* para o parâmetro de potência α são usadas: *a priori* não informativa de Jeffreys, uma uniforme imprópria e uma conjugada. As distribuições *a posteriori* e preditivas são obtidas com cada uma das distribuições *a priori* utilizadas. Estudos de simulação são apresentados com o objetivo de avaliar o comportamento do viés relativo (VR) e o erro absoluto médio relativo (EAMR) das estimativas bayesianas obtidas para a família α potência normal (PN). Em seguida é abordada a extensão da classe AP e da família PN ao caso de localização-escala, e estudos de simulação são realizados para estudar o comportamento do VR e o EAMR de alguns quantis obtidos com as estimativas bayesianas dos parâmetros. Finalmente, a metodologia é aplicada a um conjunto de dados reais sob a circunferência da cintura para adolescentes de sexo masculino com idade entre 10 e 20 anos de escolas públicas de São Paulo.

Palavras-chave: Inferência bayesiana; α potência; α potência normal; *a priori* de Jeffreys;

distribuição preditiva *a posteriori*.

4.1 Introdução

No Capítulo 2 foi comentado brevemente que a classe de distribuições α potência também é conhecida como a família exponencializada de distribuições, e que existem vários trabalhos analisando conjuntos de dados de tempos de vida assimétricos e problemas de confiabilidade. No caso em que a função de distribuição base é a função de distribuição exponencial (conhecida como distribuição exponencial generalizada), alguns autores como, por exemplo, Gupta e Kundu (1999, 2001a,b), Raqab (2002), Raqab e Madi (2005), estudaram o comportamento dos parâmetros sobre um enfoque clássico e bayesiano. Além disso, Mudholkar e Srivastava (1993) e Nassar e Eissa (2005) estudaram o caso em que a função de distribuição base é a distribuição Weibull.

Posteriormente, Dey e Maiti (2011) abordaram a família exponencializada de distribuições, assumindo o parâmetro de potência como desconhecido e que a função de distribuição base é uma função de distribuição geral. Os autores derivaram os estimadores bayesianos considerando a extensão da distribuição *a priori* não-informativos de Jeffreys e uma distribuição *a priori* conjugada sob diferentes funções de perda invariante sobre escala. Porém, esses autores não trataram o caso em que a função base é a distribuição Normal.

O principal objetivo deste capítulo é desenvolver uma abordagem bayesiana para a inferência dos parâmetros de interesse da classe de distribuições α potência e da família de distribuições α potência normal para o caso de localização-escala (PN).

O capítulo está organizado da seguinte forma. Na Seção 4.2, introduzimos uma abordagem bayesiana para a inferência do parâmetro de potência na classe de distribuições AP usando uma distribuição *a priori* não informativa de Jeffreys, uma uniforme imprópria e uma Gama (dado que permite conjugação). A distribuição preditiva *a posteriori* é obtida para cada uma das distribuições *a priori*. Estudos de simulação são apresentados para analisar o comportamento de viés relativo (VR) e o erro absoluto médio relativo (EAMR) das estimativas bayesianas obtidas para a família PN. Na Seção 4.3, estudamos o caso geral do modelo de localização-escala e particularizamos para a família $PN(\xi, \eta, \alpha)$. Na Seção 4.5,

são apresentados estudos de simulação para analisar o comportamento do VR e o EAMR de alguns quantis obtidos com as estimativas bayesianas obtidas para a família $PN(\xi, \eta, \alpha)$. A metodologia proposta é ilustrada na Seção 4.6, considerando um conjunto de dados da circunferência da cintura para adolescentes de sexo masculino com idade entre 10 e 20 anos de escolas públicas de São Paulo. As demonstrações dos resultados teóricos são apresentadas na Seção 4.8.

Considerando que a classe de distribuições AP é adequada para modelar dados com assimetria, a mediana ($Q_{0,5}$) é uma medida mais adequada que a média para resumir esses dados. Portanto, decidiu-se estudar aqui o comportamento dos quantis da família de distribuições PN.

O quantil p , denotado por Q_p , para a família PN é dado por

$$Q_p = \Phi^{-1} \left(p^{\frac{1}{\alpha}} \right).$$

Note que os quantis da família PN podem ser facilmente obtidos em comparação com os momentos. Para a classe geral definida em (2.1), o mesmo resultado pode ser observado, dado que $Q_p = G^{-1} \left(p^{\frac{1}{\alpha}} \right)$.

4.2 Inferência bayesiana para o parâmetro de potência da classe de distribuições AP

Uma abordagem bayesiana para a inferência do parâmetro de potência da classe de distribuições AP é desenvolvida nesta seção. Três propostas de distribuições *a priori* são apresentadas: a primeira é *a priori* não informativa de Jeffreys, a segunda uma uniforme imprópria, e a fim de obter uma *a priori* conjugada, a última proposta é baseada na forma da função de verossimilhança. Adicionalmente, a distribuição preditiva *a posteriori* é obtida, e finalmente, um estudo de simulação apresentado para ilustrar os resultados obtidos.

4.2.1 Conjugação Bayesiana e preditiva

Proposição 1. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim AP(\alpha)$, $i = 1, \dots, n$, com função de densidade de probabilidade definida em (2.2). Então,*

$$T(\mathbf{z}) = \sum_{i=1}^n \log [G(z_i)]^{-1} \quad (4.1)$$

é uma estatística suficiente minimal para α .

Demonstração: Ver Seção 4.8.

Proposição 2. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim AP(\alpha)$, $i = 1, \dots, n$. Se $\alpha \sim \text{Gama}(a_0, b_0)$, isto é, a função de densidade para α é dada por*

$$f(\alpha; a_0, b_0) = \frac{b_0^{a_0} \exp(-b_0\alpha) \alpha^{a_0-1}}{\Gamma(a_0)},$$

sendo, $a_0 > 0$ e $b_0 > 0$ os hiperparâmetros da distribuição a priori. Então a distribuição a posteriori, a média e variância a posteriori de α são, respectivamente, dadas por

- $\alpha|\mathbf{z} \sim \text{Gama}(n + a_0, T(\mathbf{z}) + b_0)$,
- $E(\alpha|\mathbf{z}) = \frac{n+a_0}{T(\mathbf{z})+b_0}$ e $\text{Var}(\alpha|\mathbf{z}) = \frac{n+a_0}{(T(\mathbf{z})+b_0)^2}$,

sendo $T(\mathbf{z})$ a estatística suficiente minimal definida em (4.1).

Demonstração: Ver Seção 4.8.

Corolário 4.1. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim AP(\alpha)$, $i = 1, \dots, n$. Se a distribuição a priori de α é a priori não informativa de Jeffreys ($f(\alpha) \propto \frac{1}{\alpha}$), então a distribuição a posteriori, a média e variância a posteriori de α são, respectivamente, dadas por*

- $\alpha|\mathbf{z} \sim \text{Gama}(n, T(\mathbf{z}))$,

- $E(\alpha|\mathbf{z}) = \frac{n}{T(\mathbf{z})}$ e $\text{Var}(\alpha|\mathbf{z}) = \frac{n}{(T(\mathbf{z}))^2}$,

sendo $T(\mathbf{z})$ a estatística suficiente minimal definida em (4.1).

Demonstração: Ver Seção 4.8.

Corolário 4.2. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim AP(\alpha)$, $i = 1, \dots, n$. Se a distribuição a priori de α é uma distribuição uniforme imprópria no intervalo $(0, \infty)$, isto é, $f(\alpha) \propto \mathbf{1}_{(0, \infty)}(\alpha)$. Então, a distribuição a posteriori, a média e variância a posteriori de α são, respectivamente, dadas por*

- $\alpha|\mathbf{z} \sim \text{Gama}(n + 1, T(\mathbf{z}))$,
- $E(\alpha|\mathbf{z}) = \frac{n+1}{T(\mathbf{z})}$ e $\text{Var}(\alpha|\mathbf{z}) = \frac{n+1}{(T(\mathbf{z}))^2}$,

sendo $T(\mathbf{z})$ a estatística suficiente minimal definida em (4.1).

A demonstração é omitida.

Observação: Note que as distribuições a posteriori obtidas nos corolários 4.1 e 4.2, são casos limite da distribuição a posteriori obtida na proposição 2. A primeira é obtida fazendo a_0 e b_0 tender a 0, enquanto a segunda é obtida fixando $a_0 = 1$ e fazendo b_0 tender a 0.

Proposição 3. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim AP(\alpha)$, $i = 1, \dots, n$, e $z_{\text{nov}} uma observação futura independente de \mathbf{z} . Então, se $\alpha \sim \text{Gama}(a_0, b_0)$, a distribuição preditiva a posteriori de z_{nov} é dada por$*

$$f(z_{\text{nov}}|\mathbf{z}) = \frac{(n + a_0) g(z_{\text{nov}}) [b_0 + T(\mathbf{z})]^{n+a_0}}{G(z_{\text{nov}}) [b_0 + T(\mathbf{z}) + \log [G(z_{\text{nov}})]^{-1}]^{n+a_0+1}},$$

sendo $T(\mathbf{z})$ a estatística suficiente e minimal definida em (4.1).

Demonstração: Ver Seção 4.8.

Observação: Nos casos limites quando a_0 e b_0 tendem a 0 ($f(\alpha) \propto \frac{1}{\alpha}$) e quando $a_0 = 1$ e b_0 tende a 0 [$f(\alpha) \propto \mathbf{1}_{(0,\infty)}(\alpha)$], as distribuições preditivas *a posteriori* são dadas por

$$f(z_{\text{nov}}|\mathbf{z}) = \frac{(n+1)g(z_{\text{nov}})[T(\mathbf{z})]^{n+1}}{G(z_{\text{nov}})[T(\mathbf{z}) + \log[G(z_{\text{nov}})]^{-1}]^{n+2}},$$

e

$$f(z_{\text{nov}}|\mathbf{z}) = \frac{ng(z_{\text{nov}})[T(\mathbf{z})]^n}{G(z_{\text{nov}})[T(\mathbf{z}) + \log[G(z_{\text{nov}})]^{-1}]^{n+1}},$$

respectivamente.

A distribuição preditiva *a posteriori* é útil em seleção de modelos e também como uma ferramenta para a detecção de outliers mediante a estatística CPO_i , mais detalhes ver [Gelfand *et al.* \(1992\)](#).

4.2.2 Estudo de simulação para a família PN

Para analisar o comportamento das estimativas bayesianas obtidas sob a metodologia proposta nesta seção, foram geradas 1000 amostras com tamanhos $n = 30, 50, 100$ e 200 , e diferentes valores do parâmetro de potência da família $PN(\alpha)$. Para cada uma das amostras foram obtidas a média e mediana *a posteriori* usando quatro especificações de distribuições *a priori*. Essas medidas foram avaliadas mediante o viés relativo (VR) e o erro absoluto médio relativo (EAMR). Também foram calculados os intervalos de credibilidade HPD com probabilidade de 0,95. As distribuições *a priori* utilizadas foram

- Uniforme imprópria;
- Jeffreys;
- Gama($\frac{1}{1000}, \frac{1}{1000}$);
- Gama($\frac{1}{100}, \frac{1}{100}$).

Na Tabela 4.1 são apresentadas a amplitude e a cobertura média dos intervalos de credibilidade HPD. Pode-se observar que a medida que o tamanho amostral aumenta, a amplitude

média dos intervalos de credibilidade diminui para todos os casos estudados. As amplitudes médias obtidas com *a priori* $\text{Gama}(\frac{1}{100}, \frac{1}{100})$ foram um pouco menores, enquanto *a priori* Uniforme apresentou as maiores amplitudes. Os resultados com as distribuições *a priori* de Jeffreys e a $\text{Gama}(\frac{1}{1000}, \frac{1}{1000})$ foram semelhantes. Para as coberturas médias foi observado que as três primeiras distribuições *a priori* obtiveram resultados similares, enquanto *a priori* $\text{Gama}(\frac{1}{100}, \frac{1}{100})$ apresentou coberturas um pouco diferentes, porém todas estão próximas de 0,95.

Tabela 4.1: Amplitude média e probabilidade de cobertura para o parâmetro da família PN.

n	α	Uniforme		Jeffreys		$\text{Gama}(\frac{1}{1000}, \frac{1}{1000})$		$\text{Gama}(\frac{1}{100}, \frac{1}{100})$	
		Amplitude	Cobertura	Amplitude	Cobertura	Amplitude	Cobertura	Amplitude	Cobertura
30	0,05	0,0377	0,9410	0,0368	0,9400	0,0368	0,9400	0,0368	0,9400
	25	18,8630	0,9470	18,4288	0,9600	18,4125	0,9600	18,2679	0,9600
	50	37,6462	0,9450	36,7796	0,9500	36,7145	0,9510	36,1394	0,9500
50	0,05	0,0283	0,9510	0,0279	0,9540	0,0279	0,9540	0,0279	0,9530
	25	14,3607	0,9400	14,1558	0,9500	14,1485	0,9500	14,0833	0,9490
	50	28,7087	0,9390	28,2992	0,9480	28,2698	0,9470	28,0080	0,9450
100	0,05	0,0199	0,9480	0,0197	0,9420	0,0197	0,9420	0,0197	0,9420
	25	10,0011	0,9400	9,9310	0,9520	9,9285	0,9520	9,9061	0,9510
	50	19,9934	0,9400	19,8533	0,9530	19,8433	0,9530	19,7529	0,9520
200	0,05	0,0140	0,9520	0,0139	0,9550	0,0139	0,9550	0,0139	0,9570
	25	6,9655	0,9510	6,9456	0,9540	6,9448	0,9540	6,9371	0,9530
	50	13,9915	0,9520	13,9517	0,9500	13,9482	0,9500	13,9168	0,9500

Na Figura 4.1 são apresentados os vieses relativos obtidos para os diferentes estudos de simulação usando a mediana *a posteriori* como estimador bayesiano.

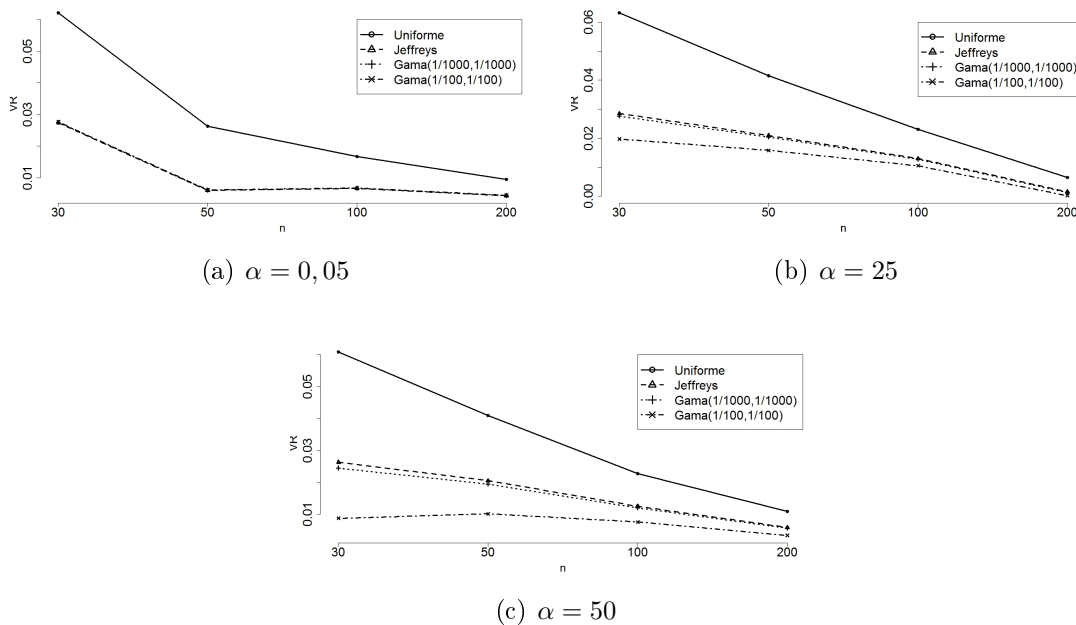


Figura 4.1: Viés relativo para α

Para $\alpha = 0,05$ os vieses relativos obtidos com as distribuições *a priori* de Jeffreys, $\text{Gama}(\frac{1}{1000}, \frac{1}{1000})$ e $\text{Gama}(\frac{1}{100}, \frac{1}{100})$ foram muito similares, enquanto *a priori* Uniforme apresentou valores maiores (outros valores de α menores que 1 foram estudados e em todos os casos o comportamento foi similar ao apresentado na Figura 4.1(a)). Para os valores de $\alpha = 25$ e 50 os melhores resultados foram obtidos usando *a priori* $\text{Gama}(\frac{1}{100}, \frac{1}{100})$, enquanto *a priori* Uniforme apresentou vieses maiores. Adicionalmente, pode se observar que os resultados usando as distribuições *a priori* de Jeffreys e a $\text{Gama}(\frac{1}{1000}, \frac{1}{1000})$ foram semelhantes. Um aspecto importante é que os VR's foram menores que 0,03 para todos os tamanhos amostrais usando as distribuições *a priori* de Jeffreys, $\text{Gama}(\frac{1}{1000}, \frac{1}{1000})$ e $\text{Gama}(\frac{1}{100}, \frac{1}{100})$.

Na Figura 4.2 são apresentados os erros absolutos relativos médios obtidos para os diferentes estudos de simulação usando a mediana *a posteriori* como estimador bayesiano.

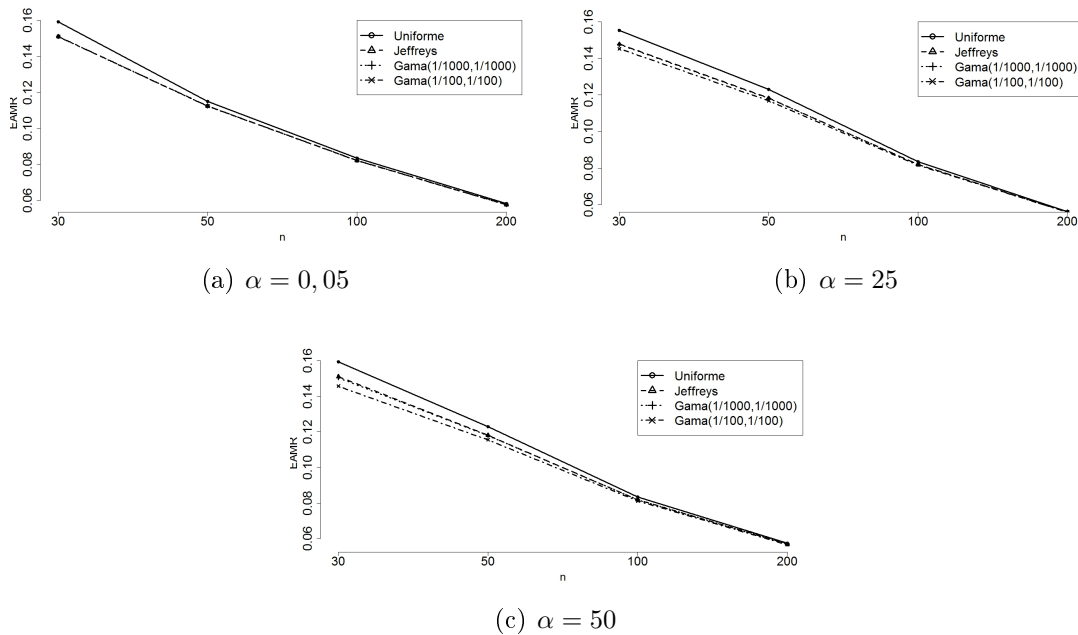


Figura 4.2: Erro absoluto relativo médio para α

Pode-se observar que os EAMR's diminuem quando o tamanho amostral aumenta para as quatro especificações de distribuições *a priori* escolhidas. O comportamento dos EAMR's são semelhantes aos VR's, porém quando o tamanho da amostra é grande ($n = 100$ e 200) os EAMR's começam a ser muito semelhantes para todas as distribuições *a priori* usadas.

De forma geral, pode se observar que os melhores resultados foram obtidos usando a distribuição *a priori* $\text{Gama}(\frac{1}{100}, \frac{1}{100})$, enquanto os piores utilizando *a priori* Uniforme im-

própria. Por outro lado, os resultados usando as distribuições *a priori* não informativa de Jeffreys e a Gama $\left(\frac{1}{1000}, \frac{1}{1000}\right)$ foram semelhantes.

Os resultados obtidos usando a média *a posteriori* como estimador bayesiano foram omitidos, dado que o comportamento dos VR's e os EAMR's foram semelhantes aos obtidos usando a mediana *a posteriori*. Porém, a mediana *a posteriori* apresentou valores menores dessas medidas em comparação com a média *a posteriori* em todos os estudos de simulação.

4.3 Inferência bayesiana para o modelo de localização-escala

Nesta seção uma abordagem bayesiana para a inferência dos parâmetros da classe de distribuições α potência para o caso localização-escala é proposta, enfatizando a família de distribuições $PN(\xi, \eta, \alpha)$. Distribuições *a priori* serão propostas para os parâmetros de interesse e suas correspondentes distribuições *a posteriori* deduzidas. Algumas demonstrações serão omitidas, dado que são similares às apresentadas na seção anterior.

4.3.1 Caso geral do modelo de localização-escala

O parâmetro de interesse para o modelo de localização-escala é o vetor $\theta = (\alpha, \xi, \eta)^T$. Suponha que existe independência *a priori* entre α e do vetor (ξ, η) , isto é,

$$f(\theta) = f(\alpha) f(\xi, \eta).$$

Proposição 4. *Seja $\mathbf{x} = (x_1, \dots, x_n)^t$ uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim AP(\xi, \eta, \alpha)$, $i = 1, \dots, n$. Se a distribuição *a priori* do vetor (ξ, η) é independente da distribuição *a priori* de α e a distribuição *a priori* de α é a priori não informativa de Jeffreys, então*

- $\alpha | \mathbf{x}, \xi, \eta \sim \text{Gama} \left(n, \sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1} \right)$,
- $E(\alpha | \mathbf{x}, \xi, \eta) = \frac{n}{\sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1}}$ e $\text{Var}(\alpha | \mathbf{x}, \xi, \eta) = \frac{n}{\left(\sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1} \right)^2}$.

Demonstração: Ver Seção 4.8.

Proposição 5. *Seja $\mathbf{x} = (x_1, \dots, x_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim AP(\xi, \eta, \alpha)$, $i = 1, \dots, n$. Se a distribuição a priori do vetor (ξ, η) é independente da distribuição a priori de α e $\alpha \sim \text{Gama}(a_0, b_0)$, então*

- $\alpha | \mathbf{x}, \xi, \eta \sim \text{Gama} \left(n + a_0, \sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1} + b_0 \right)$,
- $E(\alpha | \mathbf{x}, \xi, \eta) = \frac{n + a_0}{\sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1} + b_0}$ e $\text{Var}(\alpha | \mathbf{x}, \xi, \eta) = \frac{n + a_0}{\left(\sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1} + b_0 \right)^2}$.

Demonstração: Ver Seção 4.8.

Como $G(\cdot)$ é uma função de distribuição geral, as distribuições *a posteriori* de $[\xi | \mathbf{X}, \eta, \alpha]$ e $[\eta | \mathbf{X}, \xi, \alpha]$ não podem ser especificadas para obter as distribuições condicionais completas *a posteriori*. Na próxima subseção estudaremos a família de distribuições α potência normal para o caso de localização-escala.

4.3.2 Abordagem bayesiana para a inferência dos parâmetros da família de distribuições $PN(\xi, \eta, \alpha)$

- **Verossimilhança e distribuição conjunta a posteriori**

Suponha que $\mathbf{x} = (x_1, \dots, x_n)^T$ é uma amostra aleatória (i.i.d.) com $X_i \sim PN(\xi, \eta, \alpha)$. A função de verossimilhança é dada por

$$\begin{aligned}
 L(\mathbf{x}; \xi, \eta^2, \alpha) &= \left(\frac{\alpha}{\eta} \right)^n \prod_{i=1}^n \left[\phi \left(\frac{x_i - \xi}{\eta} \right) \left\{ \Phi \left(\frac{x_i - \xi}{\eta} \right) \right\}^{\alpha-1} \right] \\
 &= \left(\frac{\alpha}{\eta} \right)^n \prod_{i=1}^n \left[\frac{\phi \left(\frac{x_i - \xi}{\eta} \right)}{\Phi \left(\frac{x_i - \xi}{\eta} \right)} \right] \left\{ \prod_{i=1}^n \left[\Phi \left(\frac{x_i - \xi}{\eta} \right) \right] \right\}^{\alpha} \\
 &= \left(\frac{\alpha}{\eta} \right)^n T_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) \left\{ \Phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) \right\}^{\alpha}, \tag{4.2}
 \end{aligned}$$

sendo: $\mathbf{1}_n = (1, \dots, 1)_{(n \times 1)}^T$, $\phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) = \prod_{i=1}^n \phi \left(\frac{x_i - \xi}{\eta} \right)$, $\Phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) = \prod_{i=1}^n \Phi \left(\frac{x_i - \xi}{\eta} \right)$

e

$$T_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) = \frac{\phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right)}{\Phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right)}.$$

Cabe ressaltar que $\phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right)$ é a densidade de uma variável aleatória normal multivariada com vetor de média $\xi \mathbf{1}_n^T$ e matriz de covariâncias $\eta^{-2} \mathbf{I}_n$, em que \mathbf{I}_n é a matriz identidade de dimensão n , e $\Phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right)$ sua correspondente função de distribuição. Se $f(\alpha, \xi, \eta^2) = f(\alpha) f(\xi, \eta^2)$, a distribuição conjunta *a posteriori* do vetor θ é dada por

$$f(\xi, \eta^2, \alpha | \mathbf{x}) = \left(\frac{\alpha}{\eta} \right)^n T_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) \left\{ \Phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) \right\}^\alpha f(\alpha) f(\xi, \eta^2). \quad (4.3)$$

• Distribuição a priori para os parâmetros

Pensamos em utilizar as mesmas distribuições *a priori* especificadas anteriormente para α . No entanto, para poder provar a existência da distribuição conjunta *a posteriori* foi necessário restringir a domínio de α ao intervalo $[1, \infty)$.

Proposição 6. *Seja $\mathbf{x} = (x_1, \dots, x_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim PN(\alpha, \xi, \eta)$, $i = 1, \dots, n$. Suponha que $f(\alpha, \xi, \eta^2) = f(\alpha) f(\xi, \eta^2)$, $f(\xi, \eta^2) = \eta^{-2}$ e a distribuição a priori de α é Gama truncada no intervalo $[1, \infty)$ ($\alpha \sim TGama(a_0, b_0, 1, \infty)$), isto é,*

$$f(\alpha; a_0, b_0, 1, \infty) = K(a_0, b_0) \exp(-b_0 \alpha) \alpha^{a_0-1} I_{[1, \infty)}(\alpha),$$

sendo:

$$K(a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0) [1 - H(a_0, b_0)]} \quad e \quad H(a_0, b_0) = \int_0^1 \frac{b_0^{a_0} \exp(-b_0 \alpha) \alpha^{a_0-1}}{\Gamma(a_0)}.$$

Então, a distribuição conjunta a posteriori de (α, ξ, η) é própria.

Demonstração: Ver Seção 4.8.

Proposição 7. Seja $\mathbf{x} = (x_1, \dots, x_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim PN(\alpha, \xi, \eta)$, $i = 1, \dots, n$. Suponha que $f(\alpha, \xi, \eta^2) = f(\alpha) f(\xi, \eta^2)$, $f(\xi, \eta^2) = \eta^{-2}$ e $\alpha \sim TGama(a_0, b_0, 1, \infty)$, então

- $\alpha | \xi, \eta^2, \mathbf{x} \sim TGama\left(n + a_0, \log\left[\Phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right)\right]^{-1} + b_0, 1, \infty\right)$,
- $f(\xi, \eta^2 | \alpha, \mathbf{x}) \propto \left(\frac{1}{\eta^{n+2}}\right) \phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right) \left\{\Phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right)\right\}^{\alpha-1}$.

A demonstração é omitida.

Proposição 8. Seja $\mathbf{x} = (x_1, \dots, x_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim PN(\alpha, \xi, \eta)$, $i = 1, \dots, n$. Suponha que $f(\alpha, \xi, \eta^2) = f(\alpha) f(\xi, \eta^2)$, $f(\xi, \eta^2) = \eta^{-2}$ e $f(\alpha) \propto \mathbf{1}_{(1, \infty)}(\alpha)$. Então, temos que

- a distribuição conjunta a posteriori de (α, ξ, η) é própria;
- $\alpha | \xi, \eta^2, \mathbf{x} \sim TGama\left(n + 1, \log\left[\Phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right)\right]^{-1}, 1, \infty\right)$;
- $f(\xi, \eta^2 | \alpha, \mathbf{x}) \propto \left(\frac{1}{\eta^{n+2}}\right) \phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right) \left\{\Phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right)\right\}^{\alpha-1}$.

A demonstração é omitida.

Proposição 9. Seja $\mathbf{x} = (x_1, \dots, x_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim PN(\alpha, \xi, \eta)$, $i = 1, \dots, n$. Suponha que $f(\alpha, \xi, \eta^2) = f(\alpha) f(\xi, \eta^2)$, $f(\xi, \eta^2) = \eta^{-2}$ e $f(\alpha) \propto \frac{1}{\alpha} \times \mathbf{1}_{(1, \infty)}(\alpha)$. Então, temos que

- a distribuição conjunta a posteriori de (α, ξ, η) é própria;
- $\alpha | \xi, \eta^2, \mathbf{x} \sim TGama\left(n, \log\left[\Phi_n\left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}\right)\right]^{-1}, 1, \infty\right)$;

$$\bullet f(\xi, \eta^2 | \alpha, \mathbf{x}) \propto \left(\frac{1}{\eta^{n+2}} \right) \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) \left\{ \Phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) \right\}^{\alpha-1}.$$

A demonstração é omitida.

Das proposições anteriores pode-se observar que não é fácil gerar de ξ e η , portanto, se propõe implementar um algoritmo de Metropolis Hastings dentro do amostrador de Gibbs. Neste capítulo foi utilizada a função `Metro_Hastings` do pacote `MHadaptive` disponível no software [R Core Team \(2015\)](#) para o algoritmo de Metropolis Hastings. Esta função considera como proposta uma distribuição normal univariada ou multivariada, dependendo do caso. Se o usuário não fornece uma matriz de variâncias e covariâncias para esta distribuição normal, o pacote usa uma estimativa baseada na informação de Fisher calculada numericamente. Nos estudos de simulação foram consideradas as opções por default da função.

4.4 Critérios bayesianos de seleção de modelos

Os três critérios bayesianos de seleção de modelos que serão utilizados neste trabalho são: o Critério de Informação da Deviance (DIC) ([Spiegelhalter et al., 2002](#)), o valor esperado do Critério de Informação de Akaike (EAIC) ([Brooks, 2002](#)) e o valor esperado do Critério de Informação Bayesiano (EBIC) ([Brooks, 2002](#)). A forma geral destes critérios é dada por

$$E(D(\Theta)) + p_d,$$

sendo $E(D(\Theta))$ a média posterior da *Deviance* e p_d um termo de penalização que varia de acordo com o critério de informação selecionado.

Para os critérios EAIC e EBIC, o termo de penalização p_d é igual a $2p$ e $p \log(n)$, respectivamente, sendo p o número de parâmetros do modelo e n o número de observações. No caso do critério DIC, o valor p_d é chamado número efetivo de parâmetros e é definido por

$$p_d = E(D(\Theta)) - D(E(\Theta)).$$

A média posterior da *Deviance* pode se aproximar usando métodos MCMC por

$$D_{bar} = \frac{1}{B} \sum_{b=1}^B D(\Theta_b),$$

sendo que b representa a b -ésima realização de um total de B realizações.

4.5 Estudo de simulação para a família $PN(\xi, \eta, \alpha)$

Neste seção é estudado o comportamento dos quantis 0,05, 0,25, 0,50, 0,75 e 0,95, obtidos com as estimativas bayesianas dos parâmetros do modelo. As medidas de avaliação utilizadas foram o viés relativo e a erro absoluto médio relativo.

Primeiramente, foram geradas 1000 amostras com tamanhos $n = 50, 100, 200$ e 500 , e valores dos parâmetros de família de distribuições $PN(\xi, \eta, \alpha)$ fixados em $\xi = -2, \eta = 35$ e $\alpha = 100$ (estes valores foram escolhidos dado que são semelhantes aos obtidos na aplicação que será apresentada na próxima seção). Em seguida, foram estimados os parâmetros sob a metodologia apresentada na seção 4.3.2. As distribuições *a priori* usadas foram:

- Gama Truncada: $f(\xi, \eta^2) = \eta^{-2}$ e $\alpha \sim \text{TGama}(1/100, 1/100, 1, \infty)$;
- Uniforme(1, ∞): $f(\xi, \eta^2) = \eta^{-2}$ e $f(\alpha) \propto \mathbf{1}_{(1, \infty)}(\alpha)$;
- Jeffreys Truncada: $f(\xi, \eta^2) = \eta^{-2}$ e $f(\alpha) \propto \frac{1}{\alpha} \times \mathbf{1}_{(1, \infty)}(\alpha)$.

Na Tabela 4.2 são apresentados o VR e o EAMR calculados para os quantis 0,05, 0,25, 0,50, 0,75 e 0,95, obtidos mediante as estimativas bayesianas dos parâmetros do modelo para as três especificações das distribuições *a priori*. Pode-se observar que o EAMR diminui quando a tamanho amostral aumenta para os dois estimadores bayesianos usados e em quase todos os casos a mediana *a posteriori* apresentou valores menores que com a média *a posteriori*.

Os VR's obtidos com os dois estimadores bayesianos escolhidos têm magnitudes pequenas. Para a média *a posteriori* foram obtidos valores menores que 0,023, 0,033 e 0,028

para as distribuições *a priori* Gama Truncada, Uniforme(1, ∞) e Jeffreys Truncada, respectivamente. Enquanto para a mediana *a posteriori* foram menores que 0,016, 0,016 e 0,008 para as distribuições *a priori* Gama Truncada, Uniforme(1, ∞) e Jeffreys Truncada, respectivamente. Portanto, as três propostas de distribuições *a priori* permitem obter boas estimativas dos parâmetros do modelo.

Tabela 4.2: VR e EAMR das estimativas Bayesianas para o parâmetro da família PN.

n	Quantil	Estimador	Gama Truncada		Uniforme(1, ∞)		Jeffreys Truncada		
			VR	EAMR	VR	EAMR	VR	EAMR	
50	$Q_{0,05}$	Media	0,0174	0,0378	0,0326	0,0418	0,0278	0,0413	
		Mediana	-0,0157	0,0334	0,0110	0,0320	-0,0022	0,0319	
	$Q_{0,25}$	Media	0,0174	0,0283	0,0189	0,0267	0,0201	0,0285	
		Mediana	-0,0088	0,0242	0,0013	0,0210	-0,0038	0,0223	
	$Q_{0,50}$	Media	0,0172	0,0260	0,0108	0,0218	0,0155	0,0244	
		Mediana	-0,0050	0,0224	-0,0043	0,0200	-0,0048	0,0213	
	$Q_{0,75}$	Media	0,0168	0,0259	0,0038	0,0220	0,0114	0,0241	
		Mediana	-0,0019	0,0237	-0,0093	0,0234	-0,0058	0,0237	
	$Q_{0,95}$	Media	0,0158	0,0288	-0,0046	0,0279	0,0064	0,0280	
		Mediana	0,0014	0,0286	-0,0152	0,0307	-0,0071	0,0295	
	100	$Q_{0,05}$	Media	0,0224	0,0353	0,0173	0,0255	0,0237	0,0336
			Mediana	-0,0103	0,0240	0,0065	0,0212	-0,0029	0,0228
$Q_{0,25}$		Media	0,0188	0,0259	0,0100	0,0163	0,0172	0,0233	
		Mediana	-0,0063	0,0174	0,0012	0,0138	-0,0034	0,0163	
$Q_{0,50}$		Media	0,0162	0,0221	0,0057	0,0140	0,0132	0,0196	
		Mediana	-0,0044	0,0163	-0,0019	0,0134	-0,0040	0,0157	
$Q_{0,75}$		Media	0,0135	0,0204	0,0020	0,0155	0,0094	0,0186	
		Mediana	-0,0033	0,0178	-0,0045	0,0160	-0,0047	0,0177	
$Q_{0,95}$		Media	0,0095	0,0209	-0,0024	0,0201	0,0045	0,0203	
		Mediana	-0,0028	0,0224	-0,0077	0,0211	-0,0059	0,0223	
200		$Q_{0,05}$	Media	0,0220	0,0325	0,0085	0,0175	0,0215	0,0303
			Mediana	-0,0104	0,0204	0,0018	0,0157	-0,0068	0,0198
	$Q_{0,25}$	Media	0,0184	0,0242	0,0053	0,0119	0,0166	0,0220	
		Mediana	-0,0064	0,0149	3,0393e-06	0,0108	-0,0053	0,0145	
	$Q_{0,50}$	Media	0,0156	0,0202	0,0035	0,0106	0,0132	0,0184	
		Mediana	-0,0047	0,0136	-0,0011	0,0100	-0,0047	0,0134	
	$Q_{0,75}$	Media	0,0124	0,0175	0,0019	0,0115	0,0099	0,0164	
		Mediana	-0,0039	0,0143	-0,0020	0,0113	-0,0046	0,0141	
	$Q_{0,95}$	Media	0,0075	0,0160	-0,0001	0,0142	0,0053	0,0157	
		Mediana	-0,0041	0,0175	-0,0031	0,0144	-0,0051	0,0169	
	500	$Q_{0,05}$	Media	0,0226	0,0270	0,0099	0,0154	0,0203	0,0248
			Mediana	-0,0091	0,0147	-0,0017	0,0112	-0,0065	0,0136
$Q_{0,25}$		Media	0,0185	0,0211	0,0071	0,0109	0,0161	0,0189	
		Mediana	-0,0059	0,0113	-0,0021	0,0082	-0,0047	0,0104	
$Q_{0,50}$		Media	0,0151	0,0174	0,0052	0,0092	0,0128	0,0155	
		Mediana	-0,0049	0,0103	-0,0025	0,0077	-0,0043	0,0096	
$Q_{0,75}$		Media	0,0112	0,0139	0,0035	0,0086	0,0093	0,0126	
		Mediana	-0,0050	0,0103	-0,0028	0,0084	-0,0045	0,0098	
$Q_{0,95}$		Media	0,0049	0,0106	0,0011	0,0093	0,0039	0,0102	
		Mediana	-0,0065	0,0128	-0,0035	0,0104	-0,0058	0,0119	

Por outro lado, a função *Deviance* foi utilizada para avaliar o ajuste dos dados. Na Tabela

4.3 são apresentados os *Deviances* médios (\bar{D}) e seus correspondentes desvios padrões (dp_D) para as três especificações de distribuições *a priori* com as 1000 amostras geradas e os valores reais dos parâmetros. Como os parâmetros são variáveis aleatórias, os *Deviances* foram calculados usando a média e a mediana *a posteriori*.

Tabela 4.3: Deviance médio e seus desvio padrão.

n	Estimador	Gama Truncada		Uniforme($1, \infty$)		Jeffreys Truncada		Valor Real	
		D	dp_D	D	dp_D	D	dp_D	D	dp_D
50	Media	413,7805	16,8303	408,3057	10,3128	409,0168	12,1459	409,5786	10,4261
	Mediana	411,9343	11,8464	407,6464	10,5197	407,7032	10,5316		
100	Media	822,9734	25,9308	817,0728	14,9964	821,8388	26,5899	818,6527	14,9486
	Mediana	817,6884	15,7793	816,7701	15,0585	817,2420	15,2412		
200	Media	1646,4260	31,2631	1637,7280	21,6430	1645,7810	33,5254	1639,3530	21,0796
	Mediana	1639,0210	22,2599	1637,3300	21,0741	1638,7740	21,8550		
500	Media	4106,1620	39,0079	4096,7930	36,1332	4104,4710	39,4621	4094,9970	33,4807
	Mediana	4096,2060	34,9303	4093,8550	33,6588	4095,0310	33,7937		

Pode-se observar que em todos os casos os *Deviances* foram menores quando a mediana *a posteriori* é usada como estimador bayesiano. Os maiores valores foram obtidos com a priori Gama Truncada, enquanto a priori Uniforme($1, \infty$) foram os menores. Também foi observado que os resultados usando as distribuições *a priori* Uniforme($1, \infty$) e Jeffreys Truncada foram semelhantes. Portanto, os melhores resultados são obtidos ao utilizar estas distribuições *a priori*.

Um aspecto interessante é que alguns valores dos *Deviances* médios obtidos com as estimativas bayesianas dos parâmetros do modelo PN(ξ, η, α) foram menores aos calculados com os valores reais, porém pode ser observado que são muito similares. Isto pode ter sido causado porque as estimativas bayesianas, ajustam melhor os dados das amostras nesses casos.

De forma geral, nesta seção foi observado que os melhores resultados foram obtidos usando como estimador a mediana *a posteriori* e sob as distribuições *a priori* Uniforme($1, \infty$) e Jeffreys Truncada. Em outras palavras, em relação à distribuição *a priori*, os piores resultados foram obtidos usando uma distribuição conjugada para o parâmetro de potência e $f(\xi, \eta) \propto \eta^{-2}$.

4.6 Aplicação: Dados da circunferência da cintura para adolescentes de escolas públicas de São Paulo

Os dados sobre a circunferência da cintura (cc) para adolescentes de sexo masculino com idade entre 10 e 20 anos de escolas públicas de São Paulo coletados entre os anos de 2008 e 2011 são utilizados nesta subseção para ilustrar a metodologia proposta. Estes dados estão disponíveis no Relatório de Análise Estatística (RAE) número 35 de 2012 do Centro de Estatística Aplicada (CEA) do Instituto de Matemáticas e Estatística da Universidade de São Paulo (IME-USP). Nesse relatório, foi observado que a circunferência da cintura está correlacionada a fatores de risco associados a doenças cardiovasculares. Para mais detalhes ver Branco *et al.* (2012).

Na Figura 4.3 é apresentado um diagrama de caixa dos dados de circunferência da cintura, enquanto na Tabela 4.4 são apresentadas algumas estatísticas descritivas: mínimo (\min), mediana (med), média (\bar{x}), máximo (\max), desvio padrão (s), coeficiente de assimetria ($\sqrt{b_1}$) e coeficiente de curtose (b_2). Pode-se observar que este conjunto de dados apresentam assimetria e dados atípicos.

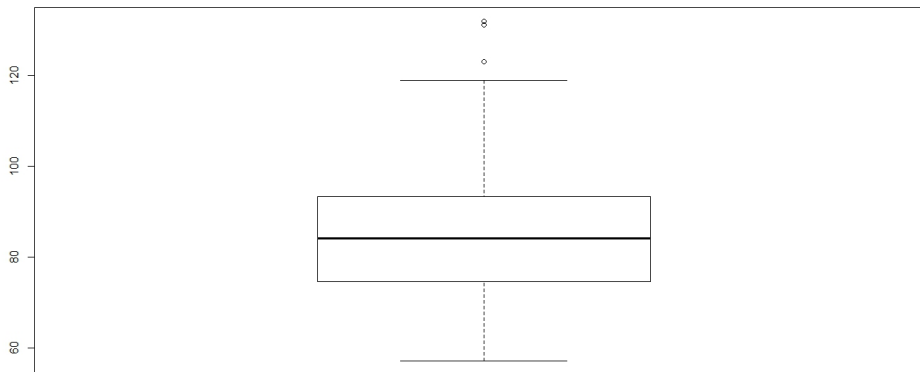


Figura 4.3: Box plot dos dados de circunferência da cintura.

Tabela 4.4: Estatísticas descritivas para a variável: circunferência da cintura.

n	min	med	\bar{x}	max	s	$\sqrt{b_1}$	b_2
94	57,10	84,05	85,91	132,00	15,51	0,84	0,44

Os modelos α potência normal para o caso de localização - escala e Normal (com média μ e variância σ^2) foram ajustados aos dados de circunferência da cintura sob um enfoque bayesiano. Para o modelo $PN(\xi, \eta, \alpha)$ foram usadas cadeias com 30.000 iterações, um período de *burn in* de 10.000 iterações e um espaçamento de tamanho 2 entre iterações, e as distribuições *a priori* descritas na subseção 4.3.2, isto é, usando $f(\xi, \eta^2) \propto \eta^{-2}$ e para o parâmetro de potência as distribuições *a priori*

- Gama Truncada: $\alpha \sim \text{TGama}(1/100, 1/100, 1, \infty)$;
- Jeffreys Truncada: $f(\alpha) \propto \frac{1}{\alpha} \times \mathbf{1}_{(1, \infty)}(\alpha)$;
- Uniforme(1, ∞): $f(\alpha) \propto \mathbf{1}_{(1, \infty)}(\alpha)$.

Enquanto que, para o modelo Normal foi assumida como distribuição *a priori* $f(\mu, \sigma^2) \propto \sigma^{-2}$.

Na Tabela 4.5 são apresentados os critérios bayesianos de seleção de modelos *DIC*, *EAIC* e *EBIC* para as quatro especificações dos modelos descritos anteriormente. Pode-se observar que as três especificações para o modelo PN tiveram valores menores para cada critério em comparação com o modelo Normal. Das três especificações para o modelo PN, a segunda apresentou valores um pouco menores, isto é, o modelo ajusta melhor quando a distribuição *a priori* não informativa de Jeffreys restrita no intervalo $(1, \infty)$ para o parâmetro de potência é usada.

Tabela 4.5: Critérios bayesianos de seleção de modelos para os modelos PN e N.

Modelo	Método		
	DIC	EAIC	EBIC
PN (Gama Truncada)	772,8731	777,0662	784,6961
PN (Jeffreys Truncada)	772,7152	776,9991	784,6290
PN (Uniforme(1, ∞))	772,9200	777,2535	784,8834
Normal	785,1289	787,1574	792,2440

Os parâmetros dos modelos PN e Normal não podem ser comparados diretamente. Portanto, com as cadeias dos parâmetros do modelo PN usando *a priori* não informativa de Jeffreys Truncada foram obtidas as cadeias para a mediana e os dois primeiros momentos da família $PN(\xi, \eta, \alpha)$.

Na Tabela 4.6 são apresentados algumas estatísticas resumos para a mediana, o primeiro momento e a variância para a variável circunferência da cintura obtidos com o modelo PN, e os parâmetros do modelo Normal, assim como, intervalos de credibilidade com probabilidade de 0,95. Pode-se observar que a média das cadeias da mediana e do primeiro momento para o modelo $PN(\xi, \eta, \alpha)$ são semelhantes à mediana e média amostras dos dados (apresentadas na Tabela 4.4), respectivamente. Adicionalmente, estes valores diferem um pouco dos obtidos para o modelo Normal, no qual, a média e mediana são iguais. Em relação à variância, pode-se observar que a variabilidade do modelo $PN(\xi, \eta, \alpha)$ é menor que a obtida pela modelo Normal.

Tabela 4.6: Estatísticas resumo para a mediana, o primeiro momento e a variância do modelo PN (Jeffreys Truncada) e os parâmetros do modelo Normal.

Parâmetro	Estatística								
	Min	$Q_{0,25}$	$Q_{0,50}$	Mean	$Q_{0,75}$	Max	Sd	$Q_{0,025}$	$Q_{0,975}$
Mediana(X)	78,6204	83,0982	84,0422	84,0811	85,0373	89,3471	1,4533	81,299	87,0350
$E(X)$	80,0385	84,6443	85,6203	85,6542	86,6504	91,2238	1,5059	82,7711	88,7185
$Var(X)$	66,8994	190,0370	214,1380	215,6130	239,2570	414,9810	37,1691	146,6167	293,2326
μ	79,7333	84,8474	85,9170	85,9272	87,0026	93,0811	1,6159	82,7609	89,0966
σ^2	147,4432	219,1712	241,8433	245,4545	267,6251	440,3916	36,8690	184,5366	328,5034

Finalmente, na Figura 4.4 apresentamos o histograma da variável circunferência da cintura, e as densidades estimadas para os modelos Normal e $PN(\xi, \eta, \alpha)$ usando como estimador bayesiano dos parâmetros a mediana *a posteriori*. Pode-se observar que o modelo $PN(\xi, \eta, \alpha)$ ajusta aparentemente melhor os dados, permitindo modelar um maior grau de assimetria e curtose que o modelo Normal.

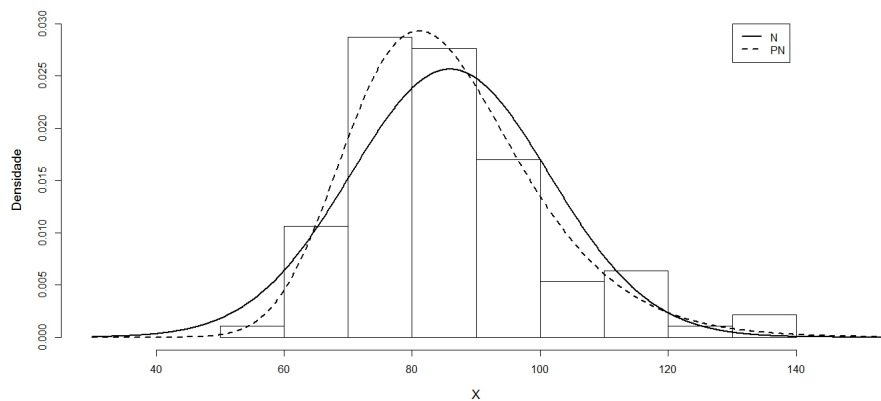


Figura 4.4: Histograma dos dados de circunferência da cintura e o ajuste dos modelos Normal e PN.

4.7 Considerações finais do capítulo

Neste capítulo foi proposta uma abordagem bayesiana para a inferência do parâmetro de interesse da classe de distribuições α potência. Uma estatística suficiente minimal que facilita a inferência clássica e bayesiana para o parâmetro de potência da classe de distribuições AP foi encontrada. Diferentes distribuições *a priori* foram estudadas, incluindo *à priori* não informativa de Jeffreys, e conseqüentemente, as correspondentes distribuições *a posteriori* e preditivas foram deduzidas. Cabe ressaltar que as distribuições preditivas *a posteriori* podem ser usadas para comparar modelos e detectar outliers.

Os estudos de simulação realizados para a família de distribuições α potência normal, mostraram que a média e a mediana *a posteriori* como estimadores bayesianos para o parâmetro de potência desta família, têm melhores desempenhos usando uma distribuição *a priori* Gama ‘pouco informativa’ em comparação com a distribuição *a priori* não informativa de Jeffreys. Adicionalmente, observou-se que de forma geral a mediana *a posteriori* apresentou melhor resultados que a média *a posteriori*.

Neste capítulo, também foi estudada a extensão localização-escala para o modelo α potência. Nesse caso, o domínio da distribuição *a priori* não informativa de Jeffreys para o parâmetro de potência da família de distribuições α potência normal, teve que ser restrito ao intervalo $[1, \infty)$ para garantir que a distribuição conjunta *a posteriori* dos parâmetros fosse própria quando $f(\xi, \eta) \propto \eta^{-2}$. Os estudos de simulação mostraram que para o modelo de localização-escala com $f(\xi, \eta) \propto \eta^{-2}$, é melhor usar as distribuições *a priori* não informativas estudadas neste capítulo para o parâmetro de potência, em vez de uma *a priori* conjugada ‘pouco informativa’.

Para o modelo de localização-escala não foi possível obter distribuição *a posteriori* com formas conhecidas para estes dois parâmetros, portanto, foi necessária utilizar Métodos de Monte Carlo via Cadeias de Markov (MCMC). Propusemos utilizar um algoritmo de Metropolis-Hastings dentro de um algoritmo de Gibbs Sampling.

Se um conjunto de dados \mathbf{X} apresenta assimetria à esquerda, o valor do parâmetro de potência para a família PN(ξ, η, α) será menor que um. Nestes casos se recomenda usar os dados transformados $\mathbf{Y} = -\mathbf{X}$ para utilizar a abordagem bayesiana apresentada neste

trabalho, pois os dados transformados terão assimetria à direita e o valor do parâmetro de potência será maior que um. Outra solução, seria utilizar distribuições próprias para todos os parâmetros.

4.8 Detalhes técnicos do capítulo

Nesta seção apresentamos as demonstrações correspondentes aos resultados obtidos no presente capítulo.

Demonstração da proposição 1

Suficiente: Seja $\mathbf{z} = (z_1, \dots, z_n)^t$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim AP(\alpha)$, $i = 1, \dots, n$. Então, a função de verossimilhança de \mathbf{Z} pode ser escrita como

$$\begin{aligned}
 L(\mathbf{z}; \alpha) &= \prod_{i=1}^n \alpha g(z_i) \{G(z_i)\}^{\alpha-1} \\
 &= \alpha^n \left\{ \prod_{i=1}^n g(z_i) \right\} \left\{ \prod_{i=1}^n G(z_i) \right\}^{\alpha-1} \\
 &= \alpha^n \left\{ \prod_{i=1}^n g(z_i) \right\} \exp \left((\alpha - 1) \sum_{i=1}^n \log G(z_i) \right) \\
 &= \alpha^n c(\mathbf{z}) \exp \left(-\alpha \left\{ \sum_{i=1}^n \log [G(z_i)]^{-1} \right\} \right), \tag{4.4}
 \end{aligned}$$

sendo:

$$c(\mathbf{z}) = \left\{ \prod_{i=1}^n g(z_i) \right\} \exp \left(\sum_{i=1}^n \log [G(z_i)]^{-1} \right).$$

Portanto, fazendo

$$D(\alpha, T(\mathbf{z})) = \alpha^n \exp(-\alpha T(\mathbf{z})) \quad \text{e} \quad T(\mathbf{z}) = \sum_{i=1}^n \log [G(z_i)]^{-1},$$

temos pelo Teorema da Fatoração que $T(\mathbf{z})$ é uma estatística suficiente para α . Então, a distribuição *a posteriori* dependerá apenas de T .

Minimal:

$$\begin{aligned} \frac{L(\mathbf{z}; \alpha)}{L(\mathbf{x}; \alpha)} &= \frac{\alpha^n c(\mathbf{z}) \exp(-\alpha \{\sum_{i=1}^n \log [G(z_i)]^{-1}\})}{\alpha^n c(\mathbf{x}) \exp(-\alpha \{\sum_{i=1}^n \log [G(x_i)]^{-1}\})} \\ &= \frac{c(\mathbf{z})}{c(\mathbf{x})} \exp\left(-\alpha \left\{ \sum_{i=1}^n \log [G(z_i)]^{-1} - \sum_{i=1}^n \log [G(x_i)]^{-1} \right\}\right) \\ &= \frac{c(\mathbf{z})}{c(\mathbf{x})} \exp(-\alpha \{T(\mathbf{z}) - T(\mathbf{x})\}). \end{aligned} \quad (4.5)$$

Claramente (4.5) não depende de α , se e somente se $T(\mathbf{z}) = T(\mathbf{x})$. Portanto, $T(\mathbf{z}) = \sum_{i=1}^n \log [G(z_i)]^{-1}$ é minimal para α .

Demonstração da proposição 2

Suponha que o parâmetro de potência tem distribuição *a priori* Gama, isto é, $\alpha \sim \text{Gamma}(a_0, b_0)$.

Assim temos que

$$\begin{aligned} f(\alpha; \mathbf{Z}) &\propto \alpha^n \exp\left(-\alpha \left\{ \sum_{i=1}^n \log [G(z_i)]^{-1} \right\}\right) \exp(-b_0 \alpha) \alpha^{a_0-1} \\ &\propto \exp\left(-\alpha \left\{ b_0 + \sum_{i=1}^n \log [G(z_i)]^{-1} \right\}\right) \alpha^{a_0+n-1} \end{aligned}$$

Portanto,

$$\alpha | \mathbf{Z} \sim \text{Gama}\left(a_0 + n, b_0 + \sum_{i=1}^n \log [G(z_i)]^{-1}\right). \quad (4.6)$$

As expressões da média e variância *a posteriori* de α são implicações da distribuição *a posteriori* de α .

Demonstração do corolário 4.1

A distribuição *a priori* não informativa de Jeffreys é proporcional à raiz quadrada da informação esperada de Fisher de α ($I(\alpha)$). Da equação 4.4, temos que

$$\frac{\partial \log L(\mathbf{Z}; \alpha)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log G(z_i) \quad \text{e} \quad \frac{\partial^2 \log L(\mathbf{Z}; \alpha)}{\partial \alpha^2} = -\frac{n}{\alpha^2}.$$

Portanto, é fácil ver que

$$I(\alpha) = \frac{n}{\alpha^2} \quad \text{e} \quad f(\alpha) \propto \frac{1}{\alpha}.$$

A distribuição *a posteriori* de α dada *a priori* não informativa de Jeffreys é

$$\begin{aligned} f(\alpha; \mathbf{Z}) &\propto \alpha^n \exp\left(-\alpha \left\{ \sum_{i=1}^n \log [G(z_i)]^{-1} \right\}\right) \frac{1}{\alpha} \\ &\propto \exp\left(-\alpha \left\{ \sum_{i=1}^n \log [G(z_i)]^{-1} \right\}\right) \alpha^{n-1} \end{aligned}$$

Assim,

$$\alpha | \mathbf{Z} \sim \text{Gama} \left(n, \sum_{i=1}^n \log [G(z_i)]^{-1} \right). \quad (4.7)$$

As expressões da média e variância *a posteriori* de α são implicações da distribuição *a posteriori* de α .

Demonstração da proposição 3

$$\begin{aligned}
f(z_{\text{nov}}; \mathbf{z}) &= \int_0^\infty f(z_{\text{nov}}, \alpha | \mathbf{z}) d\alpha = \int_0^\infty f(z_{\text{nov}} | \alpha) f(\alpha | \mathbf{z}) d\alpha \\
&= \int_0^\infty \alpha g(z_{\text{nov}}) \{G(z_{\text{nov}})\}^{\alpha-1} \frac{[b_0 + T(\mathbf{z})]^{n+a_0}}{\Gamma(n+a_0)} \exp(-(b_0 + T(\mathbf{z}))\alpha) \alpha^{n+a_0-1} d\alpha \\
&= \frac{g(z_{\text{nov}})}{G(z_{\text{nov}})} \frac{[b_0 + T(\mathbf{z})]^{n+a_0}}{\Gamma(n+a_0)} \int_0^\infty \alpha^{n+a_0+1-1} \exp(-(b_0 + T(\mathbf{z}) + \log[G(z_{\text{nov}})]^{-1})\alpha) d\alpha \\
&= \frac{g(z_{\text{nov}})}{G(z_{\text{nov}})} \frac{[b_0 + T(\mathbf{z})]^{n+a_0}}{\Gamma(n+a_0)} \frac{\Gamma(n+a_0+1)}{[b_0 + T(\mathbf{z}) + \log[G(z_{\text{nov}})]^{-1}]^{n+a_0+1}} \\
&= \frac{(n+a_0) g(z_{\text{nov}}) [b_0 + T(\mathbf{z})]^{n+a_0}}{G(z_{\text{nov}}) [b_0 + T(\mathbf{z}) + \log[G(z_{\text{nov}})]^{-1}]^{n+a_0+1}}
\end{aligned}$$

Demonstrações das proposições 4 e 5

Suponha que $\mathbf{x} = (x_1, \dots, x_n)$ é uma amostra aleatória (i.i.d.) de tamanho n com $X_i \sim AP(\xi, \eta, \alpha)$, $i = 1, \dots, n$. A função de verossimilhança de \mathbf{X} é dada por

$$\begin{aligned}
L(\mathbf{X}; \xi, \eta, \alpha) &= \prod_{i=1}^n \frac{\alpha}{\eta} g\left(\frac{x_i - \xi}{\eta}\right) \left\{G\left(\frac{x_i - \xi}{\eta}\right)\right\}^{\alpha-1} \\
&= \left(\frac{\alpha}{\eta}\right)^n \left\{ \prod_{i=1}^n g\left(\frac{x_i - \xi}{\eta}\right) \right\} \exp\left((\alpha - 1) \sum_{i=1}^n \log G\left(\frac{x_i - \xi}{\eta}\right)\right). \quad (4.8)
\end{aligned}$$

Portanto,

$$\frac{\partial \log L(\mathbf{X}; \alpha, \xi, \eta)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log G\left(\frac{x_i - \xi}{\eta}\right), \quad \frac{\partial^2 \log L(\mathbf{X}; \alpha, \xi, \eta)}{\partial \alpha^2} = -\frac{n}{\alpha^2}.$$

$$I(\alpha) = \frac{n}{\alpha^2} \quad \text{e} \quad f(\alpha) \propto \frac{1}{\alpha}.$$

Assim, é fácil ver que

$$\alpha | \mathbf{X}, \xi, \eta \sim \text{Gama}\left(n, + \sum_{i=1}^n \log \left[G\left(\frac{x_i - \xi}{\eta}\right)\right]^{-1}\right). \quad (4.9)$$

Similarmente, de (4.8) e supondo que $\alpha \sim \text{Gama}(a_0, b_0)$ temos que

$$\alpha | \mathbf{X}, \xi, \eta \sim \text{Gama} \left(a_0 + n, b_0 + \sum_{i=1}^n \log \left[G \left(\frac{x_i - \xi}{\eta} \right) \right]^{-1} \right). \quad (4.10)$$

As expressões da média e variância *a posteriori* de α são implicações das distribuições *a posteriori* de α definidas em (4.9) e (4.10).

Demonstração da proposição 6

É suficiente provar que

$$m(\mathbf{x}) = \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} f(\alpha, \xi, \eta | \mathbf{x}) d\alpha d\eta^2 d\xi < \infty.$$

Da equação (4.2) tem-se que

$$\begin{aligned} m(\mathbf{x}) &= \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \left(\frac{\alpha}{\eta} \right)^n \phi_n(\mathbf{z}) \{\Phi_n(\mathbf{z})\}^{\alpha-1} f(\alpha) f(\xi, \eta^2) d\alpha d\eta^2 d\xi \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} \int_1^{\infty} \alpha^n \frac{1}{\eta^{n+2}} \phi_n(\mathbf{z}) \{\Phi_n(\mathbf{z})\}^{\alpha-1} K(a_0, b_0) \exp(-b_0\alpha) \alpha^{a_0-1} d\alpha d\eta^2 d\xi, \end{aligned}$$

sendo $\mathbf{z} = \frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta}$. Como $0 < \Phi(\cdot) < 1$ e $\alpha \geq 1$, então

$$\begin{aligned} m(\mathbf{x}) &\leq \int_{-\infty}^{\infty} \int_0^{\infty} \int_1^{\infty} \alpha^n \frac{1}{\eta^{n+2}} \phi_n(\mathbf{z}) K(a_0, b_0) \exp(-b_0\alpha) \alpha^{n+a_0-1} d\alpha d\eta^2 d\xi \\ &= \left[\int_1^{\infty} K(a_0, b_0) \exp(-b_0\alpha) \alpha^{n+a_0-1} d\alpha \right] \left[\int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\eta^{n+2}} \phi_n(\mathbf{z}) d\eta^2 d\xi \right]. \end{aligned}$$

Note-se que

$$\int_1^{\infty} K(a_0, b_0) \exp(-b_0\alpha) \alpha^{n+a_0-1} d\alpha = \frac{K(a_0, b_0)}{K(n+a_0, b_0)} < \infty,$$

e adicionalmente,

$$\int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\eta^{n+2}} \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\eta} \right) d\eta^2 d\xi < \infty,$$

pois a integral resultante é equivalente ao caso de uma variável aleatória $X \sim N(\mu, \sigma^2)$, com ambos parâmetros desconhecidos e $f(\mu, \sigma^2) \propto \sigma^{-2}$. Portanto, obtemos que $m(\mathbf{x}) < \infty$.

Capítulo 5

Inferência bayesiana para os modelos de MF-PN

Resumo

Modelos de misturas finitas de K componentes de distribuições α potência normal para o caso de localização-escala (MF-PN) são usadas para a modelagem de dados em presença de assimetria, excesso de curtose, multimodalidade, dados atípicos ou combinações destas situações. Neste capítulo propomos uma abordagem bayesiana para a inferência dos parâmetros do modelo de MF-PN. Diferentes distribuições *a priori* para os parâmetros de interesse são usadas e suas correspondentes distribuições *a posteriori* são obtidas. Um algoritmo de Monte Carlo via Cadeias de Markov é proposto para obter as estimativas dos parâmetros do modelo de MF-PN. Os problemas de escolha de valores iniciais e de identificabilidade das etiquetas dos *clusters* são discutidos. Estudos de simulação são apresentados com o objetivo de estudar a taxa de classificação correta do modelo de MF-PN e analisar o comportamento do viés e do erro absoluto médio relativos das probabilidades de mistura e alguns quantis estimados para cada componente de mistura. Finalmente, a metodologia é aplicada a um conjunto de dados reais sobre o índice de massa corporal de homens em Estados Unidos da América.

Palavras-chave: Inferência bayesiana; misturas finitas; α potência normal; algoritmo MCMC;

modelos baseados em classificação; identificabilidade.

5.1 Introdução

Os modelos de misturas finitas de distribuições de probabilidade possibilitam a modelagem de uma grande gama de conjuntos de dados que podem apresentar assimetria, excesso de curtose, multimodalidade, observações atípicas ou combinações destas ao mesmo tempo, em outras palavras, permitem ajustar dados heterogêneos. Portanto, os modelos de misturas finitas são utilizados amplamente em diversas áreas do conhecimento; exemplos de suas aplicações são encontrados em [McLachlan e Basford \(1988\)](#), [McLachlan e Peel \(2000\)](#), [Frühwirth-Schnatter \(2006\)](#) e [Mengersen *et al.* \(2011\)](#).

Um caso particular dos modelos de misturas finitas são as misturas de distribuições normais (MF-N), as quais têm sido consideravelmente estudadas na literatura. O modelo de MF-N permite modelar muitas das distribuições de probabilidade existentes ao aumentar o número de componentes de misturas, porém, pode levar a uma inferência errada do número de *clusters* dos dados ajustados. Para mais detalhes ver [McLachlan e Peel \(2000\)](#) e [Marin *et al.* \(2005\)](#).

Nos últimos anos os estudos de misturas finitas de distribuições têm sido orientados principalmente a modelos cujos componentes de misturas são distribuições mais flexíveis. Por exemplo, para o caso de variáveis aleatórias univariadas e sobre uma abordagem bayesiana podem ser vistos os trabalhos de [Shawky e Bakoban \(2009\)](#), [AL-Hussaini e Hussein \(2012\)](#) e [Mohamed *et al.* \(2014\)](#), que estudam misturas finitas de dois componentes de distribuições gamas exponencializadas, exponenciais exponencializadas e exponenciais generalizadas (esta última com enfoque na predição bayesiana), respectivamente.

Este capítulo tem como objetivo principal desenvolver uma abordagem bayesiana para a inferência dos parâmetros de interesse do modelo de mistura finita de K componentes de distribuições α potência normal para o caso de localização-escala (ou modelo de três parâmetros) apresentado no Capítulo 3.

O capítulo está organizado da seguinte forma. Na Seção 5.2, a identificabilidade do modelo de mistura finita de distribuições α potência normal para o caso de localização-escala

é estudada. Uma abordagem bayesiana para a inferência dos parâmetros de interesse do modelo de MF-PN usando diferentes distribuições *a priori* é introduzida na Seção 5.3, assim como, um algoritmo de Monte Carlo via Cadeias de Markov (MCMC por sua sigla em inglês). O problema de valores iniciais para o algoritmo também é discutido nesta seção. Na Seção 5.4 são apresentados estudos de simulação nos quais a taxa de classificação correta do modelo de MF-PN é estudada, adicionalmente, o comportamento do viés e do erro absoluto médio relativos das probabilidades de mistura e alguns quantis estimados para cada componente de mistura são analisados. A metodologia proposta é ilustrada na Seção 5.5, considerando o conjunto de dados sobre o índice de massa corporal de homens em Estados Unidos da América (USA por sua sigla em inglês) já estudado anteriormente no Capítulo 3 sob uma abordagem clássica.

5.2 Identificabilidade do modelo de MF-PN

Nos modelos de misturas finitas de distribuições de probabilidade cada observação ou indivíduo pertence a um único componente de mistura. No entanto, esta suposição conduz ao problema de não identificabilidade dos componentes de mistura devido à permutação de rótulos ou etiquetas dos indivíduos. A permutação nas etiquetas dos indivíduos induz a uma permutação nas etiquetas dos parâmetros entre os componentes de mistura. Este problema tem sido discutido por diferentes autores, por exemplo, ver [McLachlan e Peel \(2000\)](#), [Marin *et al.* \(2005\)](#) e [Frühwirth-Schnatter \(2006\)](#).

[Marin *et al.* \(2005\)](#) afirmam que uma característica básica das misturas finitas de distribuições é precisamente ser invariante sob permutação das etiquetas dos componentes, implicando que os parâmetros dos componentes de mistura não são identificáveis marginalmente. Portanto, esta característica de não identificabilidade é um grande problema para a inferência bayesiana, principalmente no momento da utilização de métodos de Monte Carlo via cadeias de Markov.

No caso das misturas finitas de distribuições normais, diferentes autores incorporam restrições sob os parâmetros para resolver o problema de não identificabilidade devido às etiquetas dos componentes. As principais restrições e as mais simples, estão baseadas no

reordenamento dos parâmetros em relação à ordem das médias, das variâncias ou dos pesos das componentes de mistura.

Em [McLachlan e Peel \(2000\)](#) pode ser encontrada uma discussão na qual os autores estabelecem que se o problema de não identificabilidade induzido pela permutação de etiquetas é resolvido, impondo uma restrição de identificabilidade nos componentes de mistura, a maioria dos modelos de mistura são identificáveis.

Portanto, para o modelo de misturas finitas de K componentes de distribuições α potência normal de três parâmetros, propomos restringir a ordem das etiquetas dos parâmetros realizando um reordenando segundo as medianas dos componentes de mistura. Em outras palavras, se Θ_k e $Q_{k;0,5}$ representam o vetor de parâmetros e a correspondente mediana da k -ésima componente de mistura, respectivamente, então

$$Q_{1;0,5} < Q_{2;0,5} < \cdots < Q_{K;0,5}.$$

5.3 Inferência bayesiana para os parâmetros da família de MF-PN

Nesta seção, uma abordagem bayesiana para a inferência dos parâmetros do modelo de misturas finitas de distribuições α potência normal para o caso localização-escala é proposta. Para algumas distribuições *a priori* consideradas para os parâmetros de interesse, as correspondentes distribuições *a posteriori* condicionais são deduzidas e um algoritmo MCMC é apresentado.

5.3.1 Distribuição a priori para os parâmetros do modelo de MF-PN

Considere uma amostra aleatória (i.i.d.) de n indivíduos $\mathbf{z} = (z_1, \dots, z_n)^t$ que seguem uma distribuição de mistura finita de K componentes de α potência normal para o caso localização-escala¹. Suponha que a amostra está dividida em K grupos e que não se conhece

¹Este modelo foi apresentado na Seção 3.3

a qual destes grupos pertence cada indivíduo.

A função de verossimilhança associada a esta amostra é dado por

$$L_{\text{dados}}(\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\pi} | \mathbf{z}) = \prod_{i=1}^n \sum_{j=1}^K \pi_j \frac{\alpha_j}{\eta_j} \phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \left\{ \Phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \right\}^{\alpha_j - 1}, \quad (5.1)$$

sendo α_j , ξ_j e η_j os parâmetros da distribuição $PN(\xi, \eta, \alpha)$ da j -ésima componente de mistura, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)^t$, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)^t$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^t$ e $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)^t$. No entanto, esta expressão da função de verossimilhança tem um grande custo computacional para tamanhos amostrais muito grande ao momento de sua maximização.

Para simplificar a função de verossimilhança apresentada na equação (5.1), introduzimos variáveis latentes Y_i tais que

$$Z_i | Y_i = j \sim f_{\phi_{LE}}(z_i | \xi_j, \eta_j, \alpha_j) \quad \text{e} \quad P(Y_i = j) = \pi_j.$$

Portanto, é obtido um conjunto de dados ausentes ou não observados $\mathbf{y} = (y_1, \dots, y_n)^t$, associado ao vetor de dados observados \mathbf{z} . Este vetor não observado \mathbf{y} fornece as etiquetas indicando a que componentes de mistura ou grupo pertencem cada indivíduo.

Então, a verossimilhança dos dados completos é a função de verossimilhança obtida como se $(\mathbf{z}^t, \mathbf{y}^t)$ foram observáveis, isto é,

$$L(\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\pi} | \mathbf{z}, \mathbf{y}) = \prod_{j=1}^K \pi_j^{n_j} \prod_{i: y_i=j} \left[\frac{\alpha_j}{\eta_j} \phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \left\{ \Phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \right\}^{\alpha_j - 1} \right], \quad (5.2)$$

sendo $n_j = \#\{y_i = j\}$ e $n = \sum_{j=1}^K n_j$.

De forma geral, a probabilidade *a posteriori* que a observação z_i pertença à j -ésima componente é dada por

$$P(y_i = j | z_i, \boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\pi}) = \frac{\pi_j \frac{\alpha_j}{\eta_j} \phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \left\{ \Phi\left(\frac{z_i - \xi_j}{\eta_j}\right) \right\}^{\alpha_j - 1}}{\sum_{m=1}^K \pi_m \frac{\alpha_m}{\eta_m} \phi\left(\frac{z_i - \xi_m}{\eta_m}\right) \left\{ \Phi\left(\frac{z_i - \xi_m}{\eta_m}\right) \right\}^{\alpha_m - 1}}.$$

As demonstrações das seguintes proposições são omitidas, dado que são similares às apresentadas no capítulo anterior.

Proposição 10. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim MF - PN$, $i = 1, \dots, n$. Suponha que $\boldsymbol{\pi} \sim D(\pi_{1;0}, \dots, \pi_{K;0})$ e que para cada componente de mistura $f(\alpha_j, \xi_j, \eta_j^2) = f(\alpha_j) f(\xi_j, \eta_j^2)$, $f(\xi_j, \eta_j^2) = \eta_j^{-2}$ e $\alpha_j \sim TGama(a_0, b_0, 1, \infty)$.*

Então

- $\boldsymbol{\pi} | \mathbf{z}, \mathbf{y} \sim D(\pi_{1;0} + n_1, \dots, \pi_{K;0} + n_K)$,
- $\alpha_j | \boldsymbol{\xi}, \boldsymbol{\eta}, \mathbf{z}, \mathbf{y} \sim TGama\left(n_j + a_0, \log\left[\Phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right)\right]^{-1} + b_0, 1, \infty\right)$,
- $f(\xi_j, \eta_j^2 | \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{z}, \mathbf{y}) \propto \left(\frac{\pi_j^{n_j}}{\eta_j^{n_j+2}}\right) \phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right) \left\{\Phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right)\right\}^{\alpha_j - 1}$,

em que $D(\cdot)$ representa a distribuição Dirichlet, K é o número de componentes de mistura, $\mathbf{1}_{n_j} = (1, \dots, 1)_{(n_j \times 1)}^T$, $\phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right) = \prod_{i=1}^{n_j} \phi\left(\frac{z_i - \xi_j}{\eta_j}\right)$ e $\Phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right) = \prod_{i=1}^{n_j} \Phi\left(\frac{z_i - \xi_j}{\eta_j}\right)$.

Proposição 11. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim MF - PN$, $i = 1, \dots, n$. Suponha que $\boldsymbol{\pi} \sim D(\pi_{1;0}, \dots, \pi_{K;0})$ e que para cada componente de mistura $f(\alpha_j, \xi_j, \eta_j^2) = f(\alpha_j) f(\xi_j, \eta_j^2)$, $f(\xi_j, \eta_j^2) = \eta_j^{-2}$ e $f(\alpha_j) \propto \mathbf{1}_{(1, \infty)}(\alpha_j)$.*

Então,

- $\boldsymbol{\pi} | \mathbf{z}, \mathbf{y} \sim D(\pi_{1;0} + n_1, \dots, \pi_{K;0} + n_K)$,
- $\alpha_j | \boldsymbol{\xi}, \boldsymbol{\eta}, \mathbf{z}, \mathbf{y} \sim TGama\left(n_j + 1, \log\left[\Phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right)\right]^{-1}, 1, \infty\right)$;
- $f(\xi_j, \eta_j^2 | \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{z}, \mathbf{y}) \propto \left(\frac{\pi_j^{n_j}}{\eta_j^{n_j+2}}\right) \phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right) \left\{\Phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right)\right\}^{\alpha_j - 1}$.

Proposição 12. *Seja $\mathbf{z} = (z_1, \dots, z_n)^T$ uma amostra aleatória (i.i.d.) de tamanho n com $Z_i \sim MF - PN$, $i = 1, \dots, n$. Suponha que $\boldsymbol{\pi} \sim D(\pi_{1;0}, \dots, \pi_{K;0})$ e que para componente de mistura $f(\alpha_j, \xi_j, \eta_j^2) = f(\alpha_j) f(\xi_j, \eta_j^2)$, $f(\xi_j, \eta_j^2) = \eta_j^{-2}$ e $f(\alpha_j) \propto \frac{1}{\alpha_j} \times \mathbf{1}_{(1, \infty)}(\alpha_j)$.*

Então

- $\boldsymbol{\pi} | \mathbf{z}, \mathbf{y} \sim D(\pi_{1;0} + n_1, \dots, \pi_{K;0} + n_K)$,
- $\alpha_j | \boldsymbol{\xi}, \boldsymbol{\eta}, \mathbf{z}, \mathbf{y} \sim TGama\left(n_j, \log\left[\Phi_{n_j}\left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j}\right)\right]^{-1}, 1, \infty\right)$;

$$\bullet f(\xi_j, \eta_j^2 | \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{z}, \mathbf{y}) \propto \left(\frac{\pi_j^{n_j}}{\eta_j^{n_j+2}} \right) \phi_{n_j} \left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j} \right) \left\{ \Phi_{n_j} \left(\frac{\mathbf{z} - \xi_j \mathbf{1}_{n_j}}{\eta_j} \right) \right\}^{\alpha_j - 1}.$$

Neste capítulo foi utilizada a função `Metro_Hastings` do pacote `MHadaptive` disponível no software [R Core Team \(2015\)](#) para o algoritmo de Metropolis Hastings. Esta função considera como proposta uma distribuição normal univariada ou multivariada, dependendo do caso. Se o usuário não fornece uma matriz de variâncias e covariâncias para esta distribuição normal, o pacote usa uma estimativa baseada na informação de Fisher calculada numericamente. Nos estudos de simulação foram consideradas as opções por default da função.

5.3.2 Algoritmo MCMC para o modelo de MF-PN

Nesta subseção um algoritmo de MCMC para gerar amostras das distribuições *a posteriori* dos parâmetros do modelo de MF-PN é apresentado. Cabe ressaltar que o número de componentes de mistura (K) é considerado um valor fixo.

1. Escolher valores iniciais para $\pi^{(0)}$, $\xi^{(0)}$, $\eta^{(0)}$ e $\alpha^{(0)}$. Fazer $l = 0$.
2. Gerar $\mathbf{y}^{(l+1)}$ amostrando de $\mathbf{y} | \mathbf{z}, \pi^{(l)}, \xi^{(l)}, \eta^{(l)}, \alpha^{(l)}$.
3. Gerar $\boldsymbol{\pi}^{(l+1)}$ amostrando de $\boldsymbol{\pi} | \mathbf{z}, \mathbf{y}^{(l+1)}$.
4. Gerar $\alpha^{(l+1)}$ amostrando de $\alpha | \mathbf{z}, \mathbf{y}^{(l+1)}, \pi^{(l+1)}, \xi^{(l)}, \eta^{(l)}$.
5. Gerar $(\xi^{(l+1)}, \eta^{(l+1)})^T$ amostrando de $(\xi, \eta)^T | \mathbf{z}, \mathbf{y}^{(l+1)}, \pi^{(l+1)}, \alpha^{(l+1)}$.
6. Ordenar $\pi^{(l+1)}$, $\xi^{(l+1)}$, $\eta^{(l+1)}$ e $\alpha^{(l+1)}$, segundo a ordem das medianas das K componentes de mistura.
7. Fazer $l = l + 1$ e voltar a 2.

5.3.3 Valores iniciais para o algoritmo MCMC

Neste trabalho foi usada a seguinte metodologia para obter os valores iniciais do modelo de MF-PN necessários para a utilização do algoritmo MCMC descrito na subseção imediatamente anterior.

Primeiro, foi empregado o algoritmo k-means proposto por [Hartigan e Wong \(1979\)](#) e implementado na função `kmeans` do pacote `stats` no software [R Core Team \(2015\)](#), para dividir a amostra de n indivíduos em K subgrupos ou *cluster*. Posteriormente, para cada subgrupo foram geradas valores do vetor $(\xi_j, \eta_j, \alpha_j)^t$ conjuntamente utilizando um algoritmo de Metropolis Hastings (M-H). Para o algoritmo de M-H foi usado como valores iniciais a média amostral e o desvio padrão amostral de cada componente de mistura para ξ_j e η_j , respectivamente. Os valores iniciais para os parâmetros de potência foram fixados em 1. Finalmente, os valores iniciais das probabilidades de mistura foram as proporções de indivíduos designados a cada subgrupo com o algoritmo k-means.

5.3.4 Critério de classificação

A i -ésima observação será classificada no grupo j se

$$P(y_i = j | z_i, \hat{\xi}, \hat{\eta}, \hat{\alpha}, \hat{\pi}) = \max_{l=1, \dots, K} \left\{ P(y_i = l | z_i, \hat{\xi}, \hat{\eta}, \hat{\alpha}, \hat{\pi}) \right\},$$

em que

$$P(y_i = j | z_i, \hat{\xi}, \hat{\eta}, \hat{\alpha}, \hat{\pi}) = \frac{\hat{\pi}_j \frac{\hat{\alpha}_j}{\hat{\eta}_j} \phi\left(\frac{z_i - \hat{\xi}_j}{\hat{\eta}_j}\right) \left\{ \Phi\left(\frac{z_i - \hat{\xi}_j}{\hat{\eta}_j}\right) \right\}^{\hat{\alpha}_j - 1}}{\sum_{m=1}^K \hat{\pi}_m \frac{\hat{\alpha}_m}{\hat{\eta}_m} \phi\left(\frac{z_i - \hat{\xi}_m}{\hat{\eta}_m}\right) \left\{ \Phi\left(\frac{z_i - \hat{\xi}_m}{\hat{\eta}_m}\right) \right\}^{\hat{\alpha}_m - 1}}, \quad (5.3)$$

$\hat{\xi}_k$, $\hat{\eta}_k$, $\hat{\alpha}_k$ e $\hat{\pi}_k$ denotam as estimativas bayesianas obtidas mediante o algoritmo MCMC para os parâmetros do k -ésimo componente de mistura.

5.4 Estudo de simulação

Nesta seção são apresentados os estudos de simulação realizados para analisar o comportamento do viés e do erro absoluto médio relativos das probabilidades de mistura e dos quantis 0, 25, 0, 50 e 0, 75 de cada componente de mistura do modelo de MF-PN para vários tamanhos amostrais. Estes quantis foram calculados usando as estimativas bayesianas obtidas pelo algoritmo MCMC apresentado na Seção 5.3.2. Além disso, a capacidade do modelo de MF-PN para classificar corretamente as observações em seus correspondentes *clusters*

também foi estudada.

No Capítulo 4 foram estudadas diferentes distribuições *a priori* para os parâmetros da família α potência normal de três parâmetros. Nesse capítulo foi observado que é melhor usar como estimador bayesiano a mediana *a posteriori* em vez da média *a posteriori*. Além disso, foram obtidos bons resultados utilizando as distribuições *a priori* $f(\xi, \eta) \propto \eta^{-2}$ para os parâmetros de localização-escala e para o parâmetro de potência uma uniforme impropria ou a não informativa de Jeffreys truncada no intervalo $[1, \infty)$, especificidades nas proposições 8 e 9, respectivamente.

Portanto, neste capítulo a mediana *a posteriori* é usada como estimador bayesiano para os parâmetros de modelo de MF-PN. Adicionalmente, as distribuições *a priori* apresentadas na Proposição 12 são utilizadas, dado que nesta proposição são empregadas as distribuições *a priori* da Proposição 9.

Os parâmetros do modelo de MF-PN com $K = 3$ componentes de mistura foram fixados em: $\xi_1 = 100$, $\xi_2 = 130$, $\xi_3 = 155$, $\eta_1 = 10$, $\eta_2 = 8$, $\eta_3 = 6$, $\alpha_1 = 200$, $\alpha_2 = 50$, $\alpha_3 = 10$ e $\pi_1 = \pi_2 = 0,35$. Os tamanhos das amostras utilizados foram $n = 500, 1000, 1500$ e 2000 . Para cada conjunto de parâmetros e tamanho amostral foram geradas $m = 500$ amostras do modelo MF-PN. Na Figura 5.1 é apresentado o histograma de uma amostra simulada com $n = 1000$.

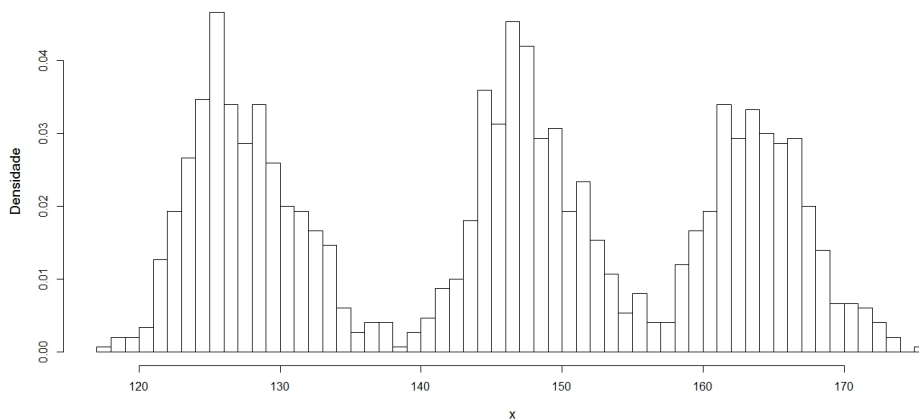


Figura 5.1: Histograma de um modelo de MF-PN com $K = 3$ e $n = 1000$.

Na Tabela 5.1 são apresentados a média, o desvio padrão (DP), os intervalos de credibilidade médios com probabilidade de 0,95, o VR e o EAMR para as estimativas dos quantis

0, 25, 0, 50 e 0, 75, e das probabilidades de mistura para cada componente ou *cluster* obtidas com as 500 réplicas geradas para cada combinação de parâmetros e tamanhos amostrais.

Tabela 5.1: VR, EAMR e estatísticas resumo.

n	Parâmetros	Valor verdadeiro	Média	DP	$Q_{0,025}$	$Q_{0,975}$	VR	EAMR
500	$Q_{1;0,25}$	124,6204	124,6486	0,3335	123,9807	125,3392	2,2609e-04	2,1416e-03
	$Q_{2;0,25}$	145,3706	145,3110	0,3166	144,6638	145,8679	-4,1029e-04	1,7511e-03
	$Q_{3;0,25}$	161,7740	161,6816	0,4328	160,8283	162,4893	-5,7133e-04	2,0984e-03
	$Q_{1;0,50}$	127,0070	126,9986	0,3621	126,2737	127,7249	-6,5609e-05	2,2710e-03
	$Q_{2;0,50}$	147,6308	147,5580	0,3417	146,8519	148,2182	-4,9350e-04	1,8234e-03
	$Q_{3;0,50}$	163,9926	163,9615	0,3649	163,2648	164,5967	-1,8991e-04	1,7450e-03
	$Q_{1;0,75}$	129,8082	129,6524	0,4373	128,7980	130,4832	-1,2003e-03	2,8733e-03
	$Q_{2;0,75}$	150,2233	150,0734	0,4636	149,1999	151,1030	-9,9765e-04	2,5405e-03
	$Q_{3;0,75}$	166,4329	166,4348	0,3858	165,6671	167,1271	1,1174e-05	1,8297e-03
π_1	0,3500	0,3469	0,0223	0,3010	0,3912	-8,8670e-03	5,1212e-02	
π_2	0,3500	0,3478	0,0232	0,3057	0,3937	-6,1775e-03	5,2900e-02	
1000	$Q_{1;0,25}$	124,6204	124,6507	0,2228	124,1745	125,0423	2,4317e-04	1,4508e-03
	$Q_{2;0,25}$	145,3706	145,3111	0,2176	144,9063	145,7517	-4,0978e-04	1,2236e-03
	$Q_{3;0,25}$	161,7740	161,6846	0,2900	161,0848	162,1843	-5,5280e-04	1,4710e-03
	$Q_{1;0,50}$	127,0070	127,0069	0,2519	126,4932	127,4927	-7,6087e-07	1,5907e-03
	$Q_{2;0,50}$	147,6308	147,5441	0,2363	147,1316	148,0434	-5,8778e-04	1,3590e-03
	$Q_{3;0,50}$	163,9926	163,9701	0,2360	163,5146	164,4058	-1,3700e-04	1,1555e-03
	$Q_{1;0,75}$	129,8082	129,6877	0,3135	129,0835	130,3046	-9,2817e-04	2,0787e-03
	$Q_{2;0,75}$	150,2233	150,0340	0,3339	149,4236	150,7317	-1,2604e-03	2,1239e-03
	$Q_{3;0,75}$	166,4329	166,4380	0,2546	165,9511	166,9563	3,0298e-05	1,2270e-03
π_1	0,3500	0,3486	0,0146	0,3177	0,3754	-4,0838e-03	3,3043e-02	
π_2	0,3500	0,3479	0,0161	0,3187	0,3811	-6,1169e-03	3,6848e-02	
1500	$Q_{1;0,25}$	124,6204	124,6473	0,1838	124,2869	125,0065	2,1632e-04	1,1849e-03
	$Q_{2;0,25}$	145,3706	145,3101	0,1724	144,9504	145,6475	-4,1662e-04	9,8724e-04
	$Q_{3;0,25}$	161,7740	161,6750	0,2412	161,2222	162,0911	-6,1163e-04	1,28440e-03
	$Q_{1;0,50}$	127,0070	127,0082	0,2006	126,5990	127,4117	9,6048e-06	1,23640e-03
	$Q_{2;0,50}$	147,6308	147,5533	0,1838	147,1901	147,9489	-5,2515e-04	1,0548e-03
	$Q_{3;0,50}$	163,9926	163,9640	0,1959	163,5927	164,3411	-1,7433e-04	9,6127e-04
	$Q_{1;0,75}$	129,8082	129,7027	0,2390	129,2081	130,1343	-8,1332e-04	1,59840e-03
	$Q_{2;0,75}$	150,2233	150,0547	0,2592	149,5688	150,5934	-1,1225e-03	1,6787e-03
	$Q_{3;0,75}$	166,4329	166,4328	0,2079	166,0252	166,8340	-6,6313e-07	9,9387e-04
π_1	0,3500	0,3475	0,0125	0,3235	0,3713	-7,1543e-03	2,9456e-02	
π_2	0,3500	0,3482	0,0131	0,3225	0,3755	-5,0218e-03	2,9774e-02	
2000	$Q_{1;0,25}$	124,6204	124,6639	0,1594	124,3429	124,9644	3,4903e-04	1,0861e-03
	$Q_{2;0,25}$	145,3706	145,3355	0,1564	144,9928	145,6316	-2,4191e-04	8,5682e-04
	$Q_{3;0,25}$	161,7740	161,7086	0,2014	161,3069	162,0691	-4,0420e-04	1,0348e-03
	$Q_{1;0,50}$	127,0070	127,0295	0,1682	126,7016	127,3532	1,7784e-04	1,08710e-03
	$Q_{2;0,50}$	147,6308	147,5885	0,1653	147,2497	147,8785	-2,8702e-04	9,0902e-04
	$Q_{3;0,50}$	163,9926	163,9759	0,1703	163,6266	164,3067	-1,0206e-04	8,0371e-04
	$Q_{1;0,75}$	129,8082	129,7343	0,2013	129,3243	130,1135	-5,6925e-04	1,3321e-03
	$Q_{2;0,75}$	150,2233	150,0958	0,2246	149,6408	150,5379	-8,4908e-04	1,37280e-03
	$Q_{3;0,75}$	166,4329	166,4309	0,1904	166,0938	166,7973	-1,2199e-05	9,0773e-04
π_1	0,3500	0,3474	0,0111	0,3272	0,3704	-7,3068e-03	2,6340e-02	
π_2	0,3500	0,3489	0,0116	0,3264	0,3722	-3,1699e-03	2,6439e-02	

Pode-se observar que os valores das médias das estimativas obtidas para os quantis e as probabilidades de misturas com as 500 réplicas são próximos dos valores reais para todos os tamanhos amostrais considerados, e que seus correspondentes desvios padrões, vieses relativos e erros absolutos médios relativos são pequenos (estes dois últimos próximos de zero). Analisando os intervalos de credibilidade, pode-se perceber que as amplitudes médias destes diminuem ao aumentar os tamanhos amostrais. Um comportamento similar observou-se para os desvios padrões e os erros absolutos médios relativos.

Finalmente, na Tabela 5.2 são apresentados os resultados obtidos para as taxas de localização ou classificação correta. Pode-se observar que para todos os casos estudados, as taxas médias foram acima de 0,985 e que seus respectivos desvios padrões pequenos. Adicionalmente, as amplitudes médias dos intervalos de credibilidade de probabilidade 0,95 diminuem ao aumentar o tamanho amostral.

Tabela 5.2: Taxas de classificação correta.

n	Média	DP	LI	LS
500	0,9858	6,6902e-03	0,9730	0,9960
1000	0,9866	3,7835e-03	0,9790	0,9940
1500	0,9866	3,3945e-03	0,9793	0,9927
2000	0,9867	2,6295e-03	0,9810	0,9910

De forma geral, pode-se afirmar que os resultados obtidos nesta seção para as estimativas das probabilidades de mistura e dos quantis de cada componente para o modelo de MF-PN sobre uma abordagem bayesiana são muito bons, assim como, para a taxa de classificação correta.

5.5 Ilustração

Nesta seção, consideramos o conjunto de dados do índice de massa corporal da Pesquisa Nacional de Saúde e Nutrição nos Estados Unidos analisado anteriormente sobre uma abordagem clássica no Capítulo 3 para ilustrar a metodologia proposta.

Os modelos de misturas finitas de distribuições α potência normal para o caso de localização - escala e Normal (com média e variância desconhecidas) foram ajustados aos dados de IMC sob um enfoque bayesiano. Para ambos modelos foram usadas cadeias com 20.000 iterações e um período de *burn in* de 15.000 iterações. As distribuições *a priori* utilizadas para o modelo MF-PN foram descritas na proposição 12, enquanto que, para o modelo MF-N foi usada como distribuição *a priori* para a i -ésima componente de mistura $f(\mu_i, \sigma_i^2) = \sigma_i^{-2}$.

Na Figura 5.2 apresentamos o histograma do conjunto de dados do IMC, junto com o ajuste dos modelos MF-PN e MF-N usando como estimador a mediana *a posteriori*. Pode

ser observado que o modelo MF-PN ajusta melhor os dados do IMC que o modelo MF-N, principalmente os indivíduos pertencentes à segunda componente da mistura.

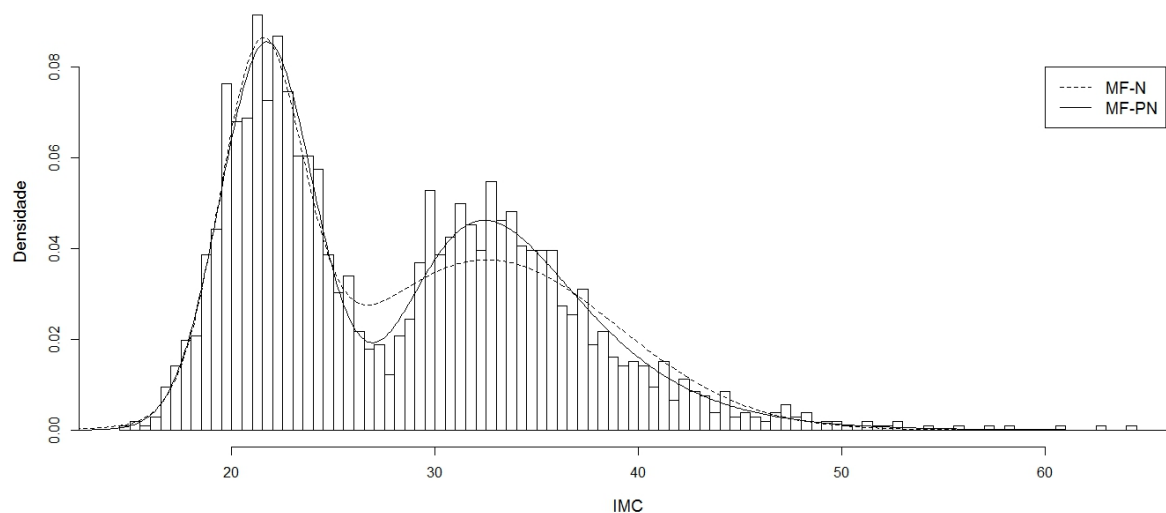


Figura 5.2: Ajuste dos modelos de MF-N e MF-PN com a resposta do IMC.

Os critérios bayesianos de seleção de modelos DIC , $EAIC$ e $EBIC$ para os dois modelos de misturas finitas de distribuições ajustados ao conjunto de dados do IMC são apresentado na Tabela 5.3. Pode-se observar que as três medidas são menores para o modelo de MF-PN, indicando que este modelo ajusta melhor os dados.

Tabela 5.3: Critérios bayesianos de seleção de modelos para os modelos de MF-N e MF-PN.

Modelo	Método		
	DIC	EAIC	EBIC
MF-N	13970,70	13980,70	13989,68
MF-PN	13699,67	13835,98	13844,96

Na Tabela 5.4 são apresentados algumas estatísticas resumos para os três primeiros quantis e as probabilidades de misturas para cada componente dos modelos de MF-N e MF-PN, assim como, intervalos de credibilidade de probabilidade 0,95. Pode-se observar que os desvios *a posteriori* para todas as medidas avaliadas são maiores para o modelo de MF-N que os obtidos com o modelo de MF-PN. Adicionalmente, o mesmo comportamento pode ser observado ao analisar as amplitudes dos intervalos de credibilidade.

Tabela 5.4: Estatísticas resumo para os três primeiros quantis e as probabilidades de mistura para os dados de IMC.

Modelo	Parâmetro	Estatística								
		Min	$Q_{0,25}$	$Q_{0,50}$	Mean	$Q_{0,75}$	Max	Sd	$Q_{0,025}$	$Q_{0,975}$
MF-N	$Q_{1;0,25}$	19,8169	20,0843	20,1418	20,1406	20,1984	20,4598	0,0834	19,9754	20,3017
	$Q_{2;0,25}$	28,5225	29,3335	29,5144	29,5103	29,6897	30,5770	0,2630	28,9843	30,0168
	$Q_{1;0,50}$	21,3129	21,5688	21,6255	21,6254	21,6819	21,9325	0,0844	21,4595	21,7923
	$Q_{2;0,50}$	32,6760	33,3451	33,4911	33,4889	33,6315	34,3671	0,2130	33,0685	33,9055
	$Q_{1;0,75}$	22,7001	23,0392	23,1099	23,1102	23,1802	23,5857	0,1057	22,9029	23,3213
	$Q_{2;0,75}$	36,6538	37,3302	37,4667	37,4675	37,6033	38,2326	0,2049	37,0671	37,8791
	π_1	0,3225	0,3834	0,3949	0,3948	0,4061	0,4625	0,0167	0,3621	0,4272
MF-PN	$Q_{1;0,25}$	20,1382	20,2259	20,2430	20,2424	20,2592	20,3363	0,0249	20,1922	20,2903
	$Q_{2;0,25}$	30,1076	30,7540	30,8659	30,8623	30,9767	31,5134	0,1683	30,5155	31,1814
	$Q_{1;0,50}$	21,6153	21,7611	21,7922	21,7919	21,8229	21,9608	0,0459	21,7002	21,8812
	$Q_{2;0,50}$	33,1157	33,5678	33,6605	33,6590	33,7521	34,2546	0,1375	33,3834	33,9214
	$Q_{1;0,75}$	23,1126	23,3252	23,3733	23,3734	23,4214	23,6490	0,0721	23,2333	23,5152
	$Q_{2;0,75}$	36,5063	37,0192	37,1115	37,1120	37,2053	37,6992	0,1398	36,8355	37,3871
	π_1	0,4440	0,4886	0,4967	0,4967	0,5048	0,5444	0,0119	0,4735	0,5200

Finalmente, utilizando a metodologia apresentada na subseção 5.3.4 para o modelo de MF-PN e sua equivalente para o modelo de MF-N, os indivíduos foram classificados em um dos dois grupos pelos dois modelos ajustados e as taxas de classificação correta foram calculadas. Cabe ressaltar que, a mediana *a posteriori* foi empregada como estimador bayesiano para os parâmetros. As taxas obtidas foram 94,16% e 98,40% para os modelos de MF-N e MF-PN, respectivamente. Portanto, o modelo de MF-PN apresentou melhores resultados que o modelo de MF-N.

5.6 Considerações finais do capítulo

Neste capítulo foi proposta uma abordagem bayesiana para a inferência dos parâmetros do modelo de misturas finitas de K componentes de distribuições α potência normal para o caso localização - escala. Mediante esta abordagem é simples obter estimativas intervalares para os parâmetros de modelo de MF-PN ou funções dos parâmetros, tais como os quantis de cada componente de mistura.

Para resolver o problema de identificabilidade das etiquetas dos parâmetros de diferentes modelos de mistura finita de K componentes, têm sido propostas restrições de reordenamento dos parâmetros em relação ao parâmetro de localização ou à média de cada componente. Dado que para a distribuição α potência normal para o caso localização - escala, o parâmetro de potência tem influência na localização e a média não têm uma expressão em forma fechada,

uma restrição de reordenamento dos parâmetros em relação à mediana de cada componente foi imposta para o modelo de MF-PN.

Estudos de simulação realizados têm mostrado que o algoritmo MCMC proposto fornece estimativas razoáveis para os parâmetros do modelo de MF-PN, e que a restrição imposta sobre o reordenamento dos parâmetros é adequada. Cabe ressaltar, que este algoritmo pode ser facilmente implementado em qualquer software estatístico. Neste trabalho foi usado o software [R Core Team \(2015\)](#).

Capítulo 6

Considerações finais da tese e pesquisas futuras

6.1 Considerações finais

Neste trabalho de tese, uma classe flexível de modelos estatísticos usando misturas finitas de distribuições α potência foi apresentada sob as perspectivas de inferência clássica e bayesiana. O caso particular do modelo de mistura finita de distribuições α potência normal para o caso de localização-escala foi utilizado em problemas de classificação. Este modelo permite analisar dados na presença de multimodalidade, assimetria, excesso de curtose, caudas pesadas e/ou observações atípicas, sendo assim uma ferramenta muito útil para profissionais de diferentes áreas do conhecimento.

No Capítulo 2, uma prova formal sobre a identificabilidade da classe de distribuições α potência foi deduzida, e conseqüentemente, para a família de distribuições α potência normal de três parâmetros. Porém, evidenciou-se um problema de quase identificabilidade intrínseco a esta família de distribuições. Este fato, nos leva a dificuldades na estimação dos parâmetros originais do modelo. Portanto, optou-se por estudar o comportamento dos momentos ou dos quantis da família de distribuições PN, dado que estas quantidades permitem representar esta família de distribuições.

No Capítulo 3, o modelo de misturas finitas de distribuições da classe α potência foi proposto. Os estudos de simulação realizados têm mostrado que o algoritmo EM fornece

estimativas razoáveis para os parâmetros do modelo de misturas finitas de K componentes de distribuições α potência normal para o caso localização - escala. Além disso, foi observada que a capacidade do modelo de MF-PN para o agrupamento ou classificação de dados pertencentes a distribuições robustas tais como os modelos MF-SN, MF-SSL, MF-SNC e MF-ST é boa, dado que foram obtidas altas taxas de classificação correta. Nosso modelo é competitivo inclusive com o modelo de MF-ST, precisando de um número menor de parâmetros.

No Capítulo 4, uma abordagem bayesiana para a inferência do parâmetro de interesse da classe de distribuições α potência foi proposta, assim como, para a família de distribuições α potência normal de três parâmetros. Para este último, o domínio da distribuição *a priori* não informativa de Jeffreys para o parâmetro de potência, teve que ser restrito ao intervalo $[1, \infty)$ para garantir que a distribuição conjunta *a posteriori* dos parâmetros fosse própria quando $f(\xi, \eta) \propto \eta^{-2}$. Estas distribuições foram escolhidas considerando não haver informação subjetiva disponível *a priori* a respeito dos parâmetros. Caso essa informação esteja disponível, poderia se trabalhar com distribuições *a priori* próprias que permitam incorporar esta informação. Os estudos de simulação mostraram que para o modelo de localização-escala com $f(\xi, \eta) \propto \eta^{-2}$, é melhor usar as distribuições *a priori* não informativas impróprias para o parâmetro de potência, em vez de uma *a priori* conjugada com grande variância.

No Capítulo 5, um enfoque bayesiano para o modelo de mistura finita de K componentes de mistura de α potência normal para o caso de localização-escala foi proposto. Uma das vantagens desta abordagem é obter estimativas intervalares de maneira simples dos parâmetros de modelo de MF-PN ou funções dos mesmos, tais como os quantis de cada componente de mistura. Os estudos de simulação mostraram que o algoritmo MCMC proposto fornece boas estimativas para os parâmetros do modelo de MF-PN, indicando que a restrição imposta para resolver o problema de identificabilidade das etiquetas dos *clusters* é adequada.

As duas metodologias propostas para o modelo de MF-PN foram ilustradas com um conjunto de dados reais do índice de massa corporal de homens em USA. Pode-se observar que o ajuste dos dados sob ambas abordagens foram semelhantes, assim como, a taxa de classificação correta.

Sob um enfoque clássico é preciso a utilização de métodos de reamostragem como Jackknife ou Bootstrap para obter estimativas intervalares de quantidades de interesse, como por

exemplo, os momentos ou os quantis de cada componente de mistura do modelo de MF-PN. No entanto, sob uma abordagem bayesiana é mais natural e simples obter estas estimativas intervalares empregando as cadeias dos parâmetros.

De forma geral, foi observado que sobre ambas metodologias o modelo de MF-PN apresentou um bom comportamento ao classificar indivíduos em seus correspondentes *clusters*. Os algoritmos propostos para estimar os parâmetros sobre cada abordagem podem ser facilmente implementados em qualquer software estatístico. Cabe ressaltar que, para o enfoque clássico um pacote foi construído no software [R Core Team \(2015\)](#).

Como resultado deste trabalho de tese, um artigo intitulado '*Finite mixture of the α power distributions applied to classification problem*' foi submetido no *Journal of Statistical Computation and Simulation*. Além disso, dois artigos sob um enfoque bayesiano estão sendo elaborados. Um primeiro deles aborda a classe de distribuições AP e a identifiabilidade desta, entretanto, o segundo discute o modelo de MF-PN.

6.2 Sugestões para pesquisas futuras

- Estudar o modelo de mistura finita de distribuições α potência quando a distribuição de probabilidade base da classe AP é a distribuição t -Student.
- Estender o modelo de MF-AP ao caso de K componentes de distribuições α potência multivariadas sobre uma abordagem clássica e bayesiana.
- Desenvolver um modelo de regressão considerando uma mistura finita de distribuições α potência para os erros.
- Explorar o caso em que o número de componentes de mistura do modelo de MF-PN é uma variável aleatória.
- Estudar os modelos de regressão quantílica para a classe de distribuições α potência.

Referências Bibliográficas

- Akaike(1974)** Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723. Citado na pág. 27
- AL-Hussaini e Hussein(2012)** Essam K. AL-Hussaini e Mohamed Hussein. Estimation under a finite mixture of exponentiated exponential components model and balanced square error loss. *Open Journal of Statistics*, 2(1):28–38. ISSN 2161-7198. doi: 10.4236/ojs.2012.21004. URL <http://dx.doi.org/10.4236/ojs.2012.21004>. Citado na pág. 70
- Asgharzadeh e Fallah(2011)** A. Asgharzadeh e A. Fallah. Estimation and prediction for exponentiated family of distributions based on records. *Communications in Statistics - Theory and Methods*, 40(1):68–83. ISSN 0361-0926. doi: 10.1080/03610920903350564. URL <http://dx.doi.org/10.1080/03610920903350564>. Citado na pág. 6
- Azzalini(1985)** Adelchi Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 12(2):171–178. Citado na pág. 8, 9, 18
- Azzalini(1986)** Adelchi Azzalini. Further results on a class of distributions which includes the normal ones. *Statistica*, 46(2):199–208. Citado na pág. 9
- Azzalini(2005)** Adelchi Azzalini. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32(2):159–188. Citado na pág. 9
- Bai et al.(1989)** Zhi-Dong Bai, Paruchuri R Krishnaiah e Lin-Cheng Zhao. On rates of convergence of efficient detection criteria in signal processing with white noise. *IEEE Transactions on Information Theory*, 35:380–388. Citado na pág. 27, 28
- Basso et al.(2010)** Rodrigo M Basso, Víctor H Lachos, Celso Rômulo Barbosa Cabral e Pulak Ghosh. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*, 54(12):2926–2941. Citado na pág. 1, 19, 36
- Branco et al.(2012)** M Branco, F Suzuki e V Verzani. Relatório de análise estatística sobre o projeto: Circunferência da cintura é mais importante do que resistência à insulina na análise do índice de risco cardiometabólico em adolescentes. Relatório Técnico RAE-CEA-12P35, Universidade de São Paulo. Citado na pág. 59
- Brooks(2002)** S.P. Brooks. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *Journal Royal Statistical Society B*, 64(3):616–618. Citado na pág. 55
- Chiogna(1998)** Monica Chiogna. Some results on the scalar skew-normal distribution. *Journal of the Italian Statistical Society*, 7(1):1–13. Citado na pág. 9

- Comaniciu e Meer(2002)** D. Comaniciu e P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619. ISSN 0162-8828. doi: 10.1109/34.1000236. URL <http://dx.doi.org/10.1109/34.1000236>. Citado na pág. 2
- Dempster et al.(1977)** Arthur P Dempster, Nan M Laird e Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (Methodological)*, 39:1–38. Citado na pág. 19
- Dey e Maiti(2011)** Sanku Dey e Sudhansu S. Maiti. Bayesian inference on the shape parameter and future observation of exponentiated family of distributions. *Journal of Probability and Statistics*, 2011:1–17. ISSN 1687-9538. doi: 10.1155/2011/457472. URL <http://dx.doi.org/10.1155/2011/457472>. Citado na pág. 44
- Dias e Wedel(2004)** José G Dias e Michel Wedel. An empirical comparison of em, sem and mcmc performance for problematic gaussian mixture likelihoods. *Statistics and Computing*, 14(4):323–332. Citado na pág. 25
- Durrans(1992)** S Durrans. Distributions of fractional order statistics in hydrology. *Water Resources Research*, 28(6):1649–1655. Citado na pág. 5, 7
- Flórez(2011)** Guillermo Domingo Martinez Flórez. *Extensões do modelo α -potência*. Tese de Doutorado, Universidade de São Paulo. Citado na pág. 6, 7, 18
- Frühwirth-Schnatter(2006)** Sylvia Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media. Citado na pág. 1, 18, 70, 71
- Gelfand et al.(1992)** Alan E Gelfand, Dipak K Dey e Hong Chang. Model determination using predictive distributions with implementation via sampling-based methods. Relatório técnico, Stanford University. Citado na pág. 48
- Gupta e Gupta(2007)** Ramesh C. Gupta e Rameshwar D. Gupta. Proportional reversed hazard rate model and its applications. *Journal of Statistical Planning and Inference*, 137(11):3525–3536. ISSN 0378-3758. doi: 10.1016/j.jspi.2007.03.029. URL <http://dx.doi.org/10.1016/j.jspi.2007.03.029>. Citado na pág. 11
- Gupta et al.(1998)** Ramesh C Gupta, Pushpa L Gupta e Rameshwar D Gupta. Modeling failure time data by lehman alternatives. *Communications in Statistics-Theory and Methods*, 27(4):887–904. Citado na pág. 6
- Gupta e Gupta(2008)** Rameshwar D. Gupta e Ramesh C. Gupta. Analyzing skewed data by power normal model. *TEST*, 17(1):197–210. ISSN 1863-8260. doi: 10.1007/s11749-006-0030-x. URL <http://dx.doi.org/10.1007/s11749-006-0030-x>. Citado na pág. 7, 9, 10, 11, 18
- Gupta e Kundu(1999)** Rameshwar D. Gupta e Debasis Kundu. Generalized exponential distributions. *Australian New Zealand Journal of Statistics*, 41(2):173–188. ISSN 1467-842X. doi: 10.1111/1467-842x.00072. URL <http://dx.doi.org/10.1111/1467-842X.00072>. Citado na pág. 6, 44
- Gupta e Kundu(2001a)** Rameshwar D. Gupta e Debasis Kundu. Exponentiated exponential family: An alternative to gamma and weibull distributions. *Biometrical Journal*, 43(1):117–130. ISSN 1521-4036. doi: 10.1002/1521-4036(200102)43:1<117::aid-bimj117>

- 3.0.co;2-r. URL [http://dx.doi.org/10.1002/1521-4036\(200102\)43:1<117::AID-BIMJ117>3.0.CO;2-R](http://dx.doi.org/10.1002/1521-4036(200102)43:1<117::AID-BIMJ117>3.0.CO;2-R). Citado na pág. 6, 44
- Gupta e Kundu(2001b)** Rameshwar D. Gupta e Debasis. Kundu. Generalized exponential distribution: different method of estimations. *Journal of Statistical Computation and Simulation*, 69(4):315–337. ISSN 1563-5163. doi: 10.1080/00949650108812098. URL <http://dx.doi.org/10.1080/00949650108812098>. Citado na pág. 6, 44
- Gupta e Kundu(2003)** Rameshwar D. Gupta e Debasis Kundu. Closeness of gamma and generalized exponential distribution. *Communications in Statistics - Theory and Methods*, 32(4):705–721. ISSN 1532-415X. doi: 10.1081/sta-120018824. URL <http://dx.doi.org/10.1081/STA-120018824>. Citado na pág. 6
- Gupta e Kundu(2004)** Rameshwar D. Gupta e Debasis Kundu. Discriminating between gamma and generalized exponential distributions. *Journal of Statistical Computation and Simulation*, 74(2):107–121. ISSN 1563-5163. doi: 10.1080/0094965031000114359. URL <http://dx.doi.org/10.1080/0094965031000114359>. Citado na pág. 6
- Gupta e Kundu(2007)** Rameshwar D. Gupta e Debasis Kundu. Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11):3537–3547. ISSN 0378-3758. doi: 10.1016/j.jspi.2007.03.030. URL <http://dx.doi.org/10.1016/j.jspi.2007.03.030>. Citado na pág. 6
- Hartigan e Wong(1979)** John A Hartigan e Manchek A Wong. A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108. Citado na pág. 25, 76
- Henze(1986)** Norbert Henze. A probabilistic representation of the skew-normal distribution. *Scandinavian journal of statistics*, 13(4):271–275. Citado na pág. 9
- Jarvis e Patrick(1973)** R.A. Jarvis e E.A. Patrick. Clustering using a similarity measure based on shared near neighbors. *IEEE Transactions on Computers*, C-22(11):1025–1034. ISSN 0018-9340. doi: 10.1109/t-c.1973.223640. URL <http://dx.doi.org/10.1109/T-C.1973.223640>. Citado na pág. 2
- KumarPatidar et al.(2012)** Anil KumarPatidar, Jitendra Agrawal e Nishchol Mishra. Analysis of different similarity measure functions and their impacts on shared nearest neighbor clustering approach. *International Journal of Computer Applications*, 40(16):1–5. ISSN 0975-8887. doi: 10.5120/5061-7221. URL <http://dx.doi.org/10.5120/5061-7221>. Citado na pág. 2
- Lehmann(1953)** ERIC L Lehmann. The power of rank tests. *The Annals of Mathematical Statistics*, 24(1):23–43. Citado na pág. 5, 18
- Lin et al.(2007a)** Tsung I Lin, Jack C Lee e Wan J Hsieh. Robust mixture modeling using the skew t distribution. *Statistics and Computing*, 17(2):81–92. Citado na pág. 1, 19, 36
- Lin et al.(2007b)** Tsung I Lin, Jack C Lee, Shu Y Yen et al. Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, 17(3):909. Citado na pág. 2, 19
- Marin et al.(2005)** Jean-Michel Marin, Kerrie Mengersen e Christian P. Robert. Bayesian modelling and inference on mixtures of distributions. *Bayesian Thinking - Modeling and Computation*, 25:459–507. ISSN 0169-7161. doi: 10.1016/s0169-7161(05)25016-2. URL [http://dx.doi.org/10.1016/S0169-7161\(05\)25016-2](http://dx.doi.org/10.1016/S0169-7161(05)25016-2). Citado na pág. 70, 71

- Martínez-Flórez et al.(2013)** Guillermo Martínez-Flórez, Barry C Arnold, Heleno Bolfarine e Héctor W Gómez. The multivariate alpha-power model. *Journal of Statistical Planning and Inference*, 143(7):1244–1255. Citado na pág. 18
- McLachlan e Krishnan(2008)** Geoffrey McLachlan e Thriyambakam Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, second edição. Citado na pág. 20
- McLachlan e Peel(2000)** Geoffrey McLachlan e David Peel. *Finite mixture models*. John Wiley & Sons. Citado na pág. 1, 18, 70, 71, 72
- McLachlan e Basford(1988)** Geoffrey J McLachlan e Kaye E Basford. *Mixture models: Inference and applications to clustering*. Marcel Dekker, New York, United States. Citado na pág. 1, 18, 25, 70
- Mengersen et al.(2011)** Kerrie Mengersen, Christian Robert e Mike Titterton. *Mixtures: estimation and applications*. John Wiley & Sons. Citado na pág. 1, 18, 70
- Mohamed et al.(2014)** Mohamed Mohmoud Mohamed, Elsayed Saleh e Shaymaa Mohamed Helmy. Bayesian prediction under a finite mixture of generalized exponential lifetime model. *Pakistan Journal of Statistics and Operation Research*, 10(4):417. ISSN 1816-2711. doi: 10.18187/pjsor.v10i4.620. URL <http://dx.doi.org/10.18187/pjsor.v10i4.620>. Citado na pág. 70
- Mudholkar e Srivastava(1993)** G.S. Mudholkar e D.K. Srivastava. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2): 299–302. ISSN 0018-9529. doi: 10.1109/24.229504. URL <http://dx.doi.org/10.1109/24.229504>. Citado na pág. 44
- Nassar e Eissa(2005)** M. M. Nassar e Fathy H. Eissa. Bayesian estimation for the exponentiated weibull model. *Communications in Statistics - Theory and Methods*, 33(10): 2343–2362. ISSN 1532-415X. doi: 10.1081/sta-200031447. URL <http://dx.doi.org/10.1081/STA-200031447>. Citado na pág. 44
- O’Hagan e Leonard(1976)** A O’Hagan e Tom Leonard. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63(1):201–203. Citado na pág. 18
- Pewsey(2000)** Arthur Pewsey. Problems of inference for azzalini’s skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870. Citado na pág. 9
- Pewsey(2006)** Arthur Pewsey. Some observations on a simple means of generating skew distributions. Em *Advances in Distribution Theory, Order Statistics, and Inference*, páginas 75–84. Springer. Citado na pág. 9
- Pewsey et al.(2012)** Arthur Pewsey, Héctor W. Gómez e Heleno Bolfarine. Likelihood-based inference for power distributions. *TEST*, 21(4):775–789. ISSN 1863-8260. doi: 10.1007/s11749-011-0280-0. URL <http://dx.doi.org/10.1007/s11749-011-0280-0>. Citado na pág. 6, 8, 18
- Prates et al.(2013)** Marcos Oliveira Prates, VH Lachos e C Cabral. mixsmn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, 54 (12):1–20. URL <http://www.jstatsoft.org/v54/i12/>. Citado na pág. 33

- R Core Team(2015)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL <https://www.R-project.org/>. Citado na pág. 3, 12, 25, 40, 55, 75, 76, 82, 85
- Raqab e Madi(2005)** Mohamed Z. Raqab e Mohamed T. Madi. Bayesian inference for the generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 75(10):841–852. ISSN 1563-5163. doi: 10.1080/00949650412331299166. URL <http://dx.doi.org/10.1080/00949650412331299166>. Citado na pág. 44
- Raqab(2002)** Mohammad Z. Raqab. Inferences for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference*, 104(2):339–350. ISSN 0378-3758. doi: 10.1016/s0378-3758(01)00246-4. URL [http://dx.doi.org/10.1016/S0378-3758\(01\)00246-4](http://dx.doi.org/10.1016/S0378-3758(01)00246-4). Citado na pág. 44
- Roberts(1966)** Charles Roberts. A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61(316):1184–1190. Citado na pág. 18
- Schwarz et al.(1978)** Gideon Schwarz et al. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464. Citado na pág. 27
- Shawky e Bakoban(2009)** AI Shawky e RA Bakoban. On finite mixture of two-component exponentiated gamma distribution. *Journal of Applied Sciences Research*, 5(10):1351–1369. Citado na pág. 70
- Spiegelhalter et al.(2002)** David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin e Angelika van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64(4):583–639. ISSN 1467-9868. doi: 10.1111/1467-9868.00353. URL <http://dx.doi.org/10.1111/1467-9868.00353>. Citado na pág. 55