

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG

INSTITUTO DE CIÊNCIAS EXATAS - ICEX

DEPARTAMENTO DE ESTATÍSTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Modelagem Bayesiana Semi-paramétrica via Misturas

Nívea Bispo da Silva

Belo Horizonte

2017

Nívea Bispo da Silva

**Modelagem Bayesiana Semi-paramétrica
via Misturas**

Tese apresentada ao Programa de Pós-Graduação em
Estatística da Universidade Federal de Minas Gerais
como parte dos requisitos para a obtenção do grau
de Doutora em Estatística.

Orientador: Prof. Dr. Marcos Oliveira Prates

Co-orientador: Prof. Dr. Flávio Bambirra Gonçalves

Belo Horizonte

2017

Modelagem Bayesiana Semi-paramétrica via Misturas

Esta versão da tese contém as correções e alterações sugeridas pela banca durante a defesa da versão original do trabalho realizada em 07 de agosto de 2017.

Banca Examinadora:

- Prof. Dr. Marcos Oliveira Prates (orientador) - UFMG
- Prof. Dr. Flávio Bambirra Gonçalves (co-orientador) - UFMG
- Profa. Dra. Rosangela Helena Loschi - UFMG
- Prof. Dr. Vinícius Diniz Mayrink - UFMG
- Prof. Dr. Dani Gamerman - UFRJ
- Prof. Dr. Celso Rômulo Barbosa Cabral - UFAM

Agradecimentos

A jornada foi longa, cansativa, mas ao olhar para trás diria citando o sábio Fernando Pessoa que *"tudo vale a pena quando a alma não é pequena"*.

Dedico este trabalho às pessoas que são o meu porto seguro: Meus pais José e Petronília, que me ensinaram a ser perseverante e determinada e a jamais desistir das coisas que acredito. Minhas irmãs, Luciene e Marília, que sempre tinham uma palavra de incentivo e ao meu sobrinho João Lucas, que chegou para dar mais luz e cor às nossas vidas. Agradeço-lhes por todo amor, orações e por sempre apoiarem as minhas decisões. Saber que posso contar com o carinho e compreensão de vocês me faz mais forte e segura de que alcançarei o que almejo, caminhando sempre com humildade e sem medo dos obstáculos que precisarei ultrapassar.

Agradeço a Deus por ser presença constante em minha vida, conduzindo-me sempre e não me deixando desistir.

Aos meus orientadores Marcos Prates e Flávio Gonçalves, agradeço por toda paciência, dedicação e por todos os momentos de discussão e aprendizado. Foi uma honra poder trabalhar com duas mentes brilhantes como vocês. Espero que a nossa parceria continue e renda muitos frutos!

Aos professores da Pós pelo conhecimento que adquiri durante as disciplinas. Às secretárias Rogeria e Rose por toda atenção, em especial à Rogeria pelas palavras de fé. À Maísa pelo maravilhoso café e por ser sempre tão prestativa!

Aos amigos que conquistei (Claudia T., Marília Souza, Mariese Alves, Silvia Lemos, Renata Fonseca, Cintia Mota, Relva Moinho e muitos outros que agora me fugiram da mente, mas que não deixam de ser importantes) ao longo desses anos e com quem compartilhei momentos únicos e inesquecíveis. Poder contar com o carinho e amizade de vocês foi e sempre será muito importante para mim.

Aos amigos que ganhei em BH e com quem compartilhei bons momentos. Agradeço-lhes pela amizade, conselhos e pelo respeito mútuo que existe entre nós. Brigadão de coração a Zaida Cornejo, Francisco William, Wagner Pinheiro, Mariana Araújo, Paulo Cerqueira e Rumenick Pereira. Agradeço também aos colegas do grupo *Stats4Good* pela

oportunidade de poder "*aprender, ensinar e discutir*" Estatística. Foi uma experiência riquíssima e que levarei sempre comigo. Brigadão também às mineiras Grazi e Débora, com quem morei, pela convivência e respeito. Conviver com todos vocês foi muito bom!

Aos membros da banca, professores Doutores Dani Gamerman (UFRJ), Celso Rômulo Cabral (UFAM), Rosangela Loschi (UFMG) e Vinícius Mayrink (UFMG), agradeço pelas ricas contribuições dadas ao texto e que serviram para enriquecê-lo ainda mais.

À Capes e Fapemig pelo apoio financeiro durante a execução deste trabalho e ao Departamento de Estatística da UFMG pela estrutura e suporte oferecidos aos longo destes anos.

Obrigada a todos que direta ou indiretamente me ajudaram a tornar possível essa conquista!

*"A mente que se abre a novas ideias jamais voltará ao seu tamanho original."
(Albert Einstein)*

Resumo

A modelagem estatística baseada em misturas finitas de distribuições é uma área de pesquisa em crescente ascensão. Devido à sua flexibilidade e ao avanço de métodos computacionais nas duas últimas décadas, esse tipo de modelagem tem se tornado bastante atrativo tanto do ponto de vista prático quanto teórico, pois permite que densidades com estruturas complexas sejam aproximadas usando uma estrutura mais simples. Além disso, os modelos estatísticos baseados em misturas conseguem capturar propriedades específicas dos dados como multimodalidade, assimetria, cauda pesada e heterogeneidade decorrente de fatores não-observados. Há na literatura inúmeros trabalhos sobre modelagem estatística baseada em misturas finitas de distribuições normais e muitos autores mostraram que esse tipo de mistura fornece uma base simples e eficaz para estimação de densidades e modelagem de populações heterogêneas. Entretanto, em problemas práticos onde há a presença de *outliers* nos dados, a distribuição normal pode ter suas estimativas para média e variância seriamente afetadas. Neste sentido há uma recente propagação de modelos baseados em misturas com componentes não-normais onde as distribuições assumidas para os componentes da mistura são, por exemplo, *t* de Student, Slash, Skew-Normal, Skew-*t*, dentre outras. Neste trabalho uma modelagem semi-paramétrica baseada em misturas finitas de distribuições *t* de Student será introduzida. A especificação do modelo proposto considera estruturas separadas para as modas e o comportamento de cauda, o que flexibiliza a estimação de densidades. Além disso, a estrutura de cauda na abordagem apresentada será estimada sem que haja a necessidade de se estimar parâmetros de grau de liberdade, cuja estimação é conhecida por ser difícil e custosa computacionalmente. Uma extensão do modelo no contexto de regressão linear também é apresentada para as situações onde os erros do modelo possuem multimodalidade, assimetria e caudas pesadas. A abordagem proposta é avaliada através de estudos de simulação e aplicações a conjuntos de dados reais, onde um algoritmo MCMC é proposto e implementado para amostrar das distribuições *a posteriori*.

Palavras-chave: *misturas finitas de distribuições, distribuições de caudas pesadas, regressão linear, inferência Bayesiana.*

Abstract

Statistical modeling based on finite mixture distributions is a growing research area. Due to its flexibility and the advance of computational methods in the last two decades, this type of modeling has become quite attractive both from a practical and a theoretical point of view, since it allows densities with complex structures to be approximated using a simpler structure. In addition, statistical models based on finite mixtures can capture specific data properties such as multimodality, asymmetry, heavy tail and heterogeneity due to unobserved factors. Numerous studies on statistical modeling based on finite mixtures of normal distributions have been published in the literature, and many authors have shown that this type of mixture provides a simple and effective basis for estimating densities and modeling heterogeneous populations. However, in practical problems where there are outliers in the data, the normal distribution may have its estimates for mean and variance severely affected. In this sense there is a recent propagation of models based on mixtures with non-normal components where the assumed distributions for the components of the mixture are, for example, Student-t, Slash, Skew-Normal, Skew-t, among others. In this work a semi-parametric model based on finite mixtures of t distributions will be introduced. The proposed model specification considers separate structures for the modes and tail behavior, which makes density estimation more flexible. In addition, the tail structure in the presented approach will be estimated without the need to estimate degree of freedom parameters, whose estimation is known to be difficult and computationally costly. An extension of the model in the linear regression context is also presented for situations where model errors have multimodality, asymmetry and heavy tails. The proposed approach is evaluated through simulation studies and applications to real data sets, where an MCMC algorithm is proposed and implemented to sample from the posterior distributions.

Keywords: *Finite mixtures, Heavy tail distributions, Linear regression, Bayesian inference.*

Sumário

Lista de Figuras	viii
Lista de Tabelas	xii
Lista de Abreviaturas	xiv
1 Introdução	1
1.1 Modelos de Mistura Finita	5
1.1.1 Distribuições Mistura de escala da Normal	9
1.1.2 Inferência em modelos de mistura finita	12
1.1.2.1 Identificabilidade em modelos de mistura finita	13
1.1.2.2 Número de componentes em modelos de mistura	15
1.1.3 Sobre Misturas de distribuições t de Student	16
2 Modelo Proposto	20
2.1 Motivação	20
2.2 O modelo	22
2.3 Inferência Bayesiana	24
2.3.1 Aspectos Computacionais	25
2.4 Extensão para modelos de regressão linear com misturas nos erros	27
3 Estudos de Simulação para o Modelo Proposto	29
3.1 Dados vindos de uma distribuição t de Student	31
3.2 Dados vindos de uma mistura de distribuições t de Student unimodal	34
3.3 Dados vindos de uma mistura usual de distribuições t de Student	36
3.4 Dados vindos da mistura de distribuições t de Student proposta	38
3.5 Modelo de regressão com misturas nos erros	40
3.6 Modelando assimetria e caudas pesadas	45
3.6.1 Dados vindos de uma distribuição skew-t	46
3.6.2 Dados vindos de uma mistura de distribuições skew-t	48

4	Aplicações a dados reais	51
4.1	Comprimento de peixes	52
4.2	Velocidade das galáxias	56
4.3	Concentração de hemoglobina em atletas	60
4.4	Pesquisa nacional de exames de saúde e nutrição - EUA	65
5	Considerações Finais	71
A	Distribuição de Y no modelo proposto	73
B	Distribuições Condicionais Completas	75
C	Outros resultados de Simulação	78
D	ACF e traços para as cadeias da <i>log</i> densidade <i>a posteriori</i>	96
	Referências Bibliográficas	99

Lista de Figuras

1.1	Distribuição empírica para o comprimento de 256 peixes.	6
1.2	Densidades de várias misturas de normais univariadas com $\sigma_1^2 = \sigma_2^2 = 1$. (a) $\mu_1 = -1, \mu_2 = 1$ e $w_1 = 0.5$; (b) $\mu_1 = -3, \mu_2 = 1$ e $w_1 = 0.6$; (c) $\mu_1 = -2, \mu_2 = 1$ e $w_1 = 0.85$; (d) $\mu_1 = -2, \mu_2 = 0.5, \mu_3 = 3, w_1 = 0.55$ e $w_2 = 0.35$	8
1.3	Algumas densidades da classe de distribuições SMN	11
1.4	Comparação entre a Mistura de t's com (a) $\mathbf{w} = (0.33, 0.34, 0.33)$, (b) $\mathbf{w} = (0.60, 0.30, 0.10)$ e (c) $\mathbf{w} = (0.1, 0.3, 0.6)$, e as densidades da t de Student com menores distâncias de Kolmogorov-Smirnov.	17
1.5	Comparação entre densidades t de Student e Misturas de t's com $\nu =$ (1.5, 2.5, 5) e (a) $\mathbf{w} = (0.43, 0.45, 0.12)$, (b) $\mathbf{w} = (0.02, 0.45, 0.53)$ e (c) $\mathbf{w} = (0.01, 0.1, 0.89)$	18
3.1	Comparação entre as curvas verdadeira, baseada no EMV e no modelo ajustado quando (a) n=100; (b) n=500 e (c) n= 2500.	33
3.2	Comparação entre as curvas verdadeira, baseada no EMV e no modelo ajustado quando (a) n=100; (b) n=500 e (c) n= 2500.	33
3.3	Histograma com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes considerando uma amostra de tamanho 2500. . .	35
3.4	Histograma para uma amostra Monte Carlo de tamanho 2500 da distân- tância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV do modelo verdadeiro, com bandas de 95% de confi- ança (cinza) e DKS's calculadas para diferentes modelos ajustados. . .	35
3.5	Histograma com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes considerando uma amostra de tamanho 2500. . .	37
3.6	Histograma para uma amostra Monte Carlo de tamanho 2500 da distân- cia de Kolmogorov-Smirnov (DKS) entre a função de distribuição real e a baseada no EMV para o modelo proposto, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.	37

3.7	Histograma para o modelo usado na geração dos dados e curvas para diferentes ajustes considerando uma amostra de tamanho 2500.	39
3.8	Histograma para uma amostra Monte Carlo de tamanho 2500 Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 99% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.	40
3.9	Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas para a amostra de tamanho 2500 e $J = 2$.	41
3.10	Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas a amostra de tamanho 2500 e $J = 3$. . .	44
3.11	Histograma para 5000 valores gerados a partir de uma Skew-t e seus respectivos ajustes em comparação à densidade real.	48
3.12	Histograma para 5000 valores gerados a partir de uma mistura de Skew-t e seus respectivos ajustes em comparação à densidade real (azul)	50
4.1	Histograma e boxplot para o comprimento de 256 peixes	52
4.2	Histograma para a distribuição empírica do comprimento dos peixes com as curvas ajustadas em cada modelo.	53
4.3	Comportamento de cauda para o comprimento dos peixes com as curvas ajustadas em cada modelo.	54
4.4	Histograma e boxplot para a velocidade de 86 galáxias.	56
4.5	Histograma para a distribuição empírica da velocidade de 82 galáxias com as curvas ajustadas em cada modelo	57
4.6	(a) Histograma e boxplot para a concentração de hemoglobina g/dl e (b) Histograma dos resíduos ordinários para o modelo de regressão linear. .	60
4.7	Resíduos do modelo de regressão com respectivas curvas ajustadas para os dados AIS.	62
4.8	Resíduos do modelo nas caudas da distribuição com respectivas curvas ajustadas para os dados AIS.	62
4.9	Boxplot para o peso em kg em função das variáveis (a) sexo e diabetes e (b) faixa etária.	66
4.10	(a) Histograma e boxplot para o peso em kilogramas e (b) Histograma dos resíduos ordinários para o modelo de regressão linear.	66
4.11	Resíduos do modelo com respectivas curvas ajustadas para os dados NHANES.	68
4.12	Resíduos do modelo nas caudas da distribuição com respectivas curvas ajustadas para os dados NHANES.	69

C.1	Comparação entre as curvas de uma densidade t de Student com $\nu = 4$ (azul), densidade da t baseada no EMV (verde) e densidade estimada para a mistura de 3 t's com $\nu = (1.5, 2.5, 5)$ (vermelha) quando (a) $n=100$; (b) $n=500$ e (c) $n= 2500$	79
C.2	Histograma para o tamanho de amostra 1000 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.	80
C.3	Histograma para o tamanho de amostra 500 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.	81
C.4	Histograma para tamanhos de amostra 500 e 1000 da amostra Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.	81
C.5	Histograma para o tamanho de amostra 1000 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.	83
C.6	Histograma para o tamanho de amostra 500 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.	85
C.7	Histograma para tamanho amostral 1000 da amostra Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.	85
C.8	Histograma para o tamanho de amostra 1000 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.	86
C.9	Histograma para o tamanho de amostra 500 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.	87
C.10	Histograma para tamanho amostral 1000 da amostra Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.	87
C.11	Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas para a amostra de tamanho 1000.	90
C.12	Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas para a amostra de tamanho 500.	90
C.13	Histograma para 1000 valores gerados a partir de uma Skew-t (0, 1, 1.5, 2.5) e seus respectivos ajustes em comparação à densidade real	93
C.14	Histograma para 1000 valores gerados a partir de uma mistura de Skew-t e seus respectivos ajustes em comparação à densidade real	95
D.1	ACF e traço para a cadeia da <i>log</i> densidade <i>a posteriori</i> para a amostra de tamanho 2500 gerada com $J = 1$ e $K = 2$ - modelo C.	96

D.2	ACF e traço para a cadeia da <i>log</i> densidade <i>a posteriori</i> para a amostra de tamanho 2500 gerada com $J = 2$ e $K = 1$ - modelo B.	97
D.3	ACF e traço para a cadeia da <i>log</i> densidade <i>a posteriori</i> para a amostra de tamanho 2500 gerada com $J = 2$ e $K = 2$ - modelo C.	97
D.4	ACF e traço para a cadeia da <i>log</i> densidade <i>a posteriori</i> para a amostra de tamanho 2500 gerada com $J = 2$ e $K = 2$ - modelo de regressão. . .	97
D.5	ACF e traço para a cadeia da <i>log</i> densidade <i>a posteriori</i> para a amostra de tamanho 2500 gerada com $J = 3$ e $K = 2$ - modelo de regressão. . .	98

Lista de Tabelas

3.1	Resultados <i>a posteriori</i> para a mistura de <i>t</i> 's considerando diferentes tamanhos amostrais e $\nu = 2$ na geração da <i>t</i> de Student	32
3.2	Resultados <i>a posteriori</i> para a mistura de <i>t</i> 's considerando diferentes tamanhos amostrais e $\nu = 3$ na geração da <i>t</i> de Student	32
3.3	Resultados <i>a posteriori</i> * para uma mistura de distribuições <i>t</i> de Student com $K = 2$	34
3.4	Resultados <i>a posteriori</i> para a mistura de <i>t</i> de Student com $J = 2$ e $K = 1$ e tamanho de amostra 2500	36
3.5	Resultados <i>a posteriori</i> para o modelo proposto considerando tamanho de amostra 2500	39
3.6	Resultados <i>a posteriori</i> para o modelo de regressão com distribuição de mistura para os erros considerando $n=2500$ e $J = 2$	42
3.7	Resultados <i>a posteriori</i> para o modelo de regressão com distribuição de mistura para os erros considerando $n=2500$ e $J = 3$	44
3.8	viés, variância e erro quadrático médio para a esperança e variância de <i>Y</i> em diferentes distribuições de mistura quando os dados são gerados de uma distribuição skew- <i>t</i>	47
3.9	viés, variância e erro quadrático médio para a esperança e variância de <i>Y</i> em diferentes distribuições de mistura quando os dados são gerados de uma mistura de skew- <i>t</i>	49
4.1	Crítério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D referentes ao comprimento de 256 peixes.	53
4.2	Resultados <i>a posteriori</i> e intervalo HPD de 95% para o comprimento dos peixes.	55
4.3	Crítério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D referentes a velocidade das galáxias.	57
4.4	resultados <i>a posteriori</i> e intervalo HPD de 95% para a velocidade das galáxias	59

4.5	Critério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D e critério BIC para diferentes valores de J no modelo E	61
4.6	Resultados <i>a posteriori</i> * para o modelo de regressão com distribuição de mistura nos erros para os dados do estudo AIS.	64
4.7	Critério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D e critério BIC para diferentes valores de J no modelo E	67
4.8	Resultados <i>a posteriori</i> * para o modelo de regressão com distribuição de mistura para os erros nos dados do estudo NHANES.	70
C.1	Resultados <i>a posteriori</i> para a mistura de t's considerando diferentes tamanhos amostrais e $\nu = 4$ na geração da t de Student	79
C.2	Resultados <i>a posteriori</i> * para uma mistura de distribuições t de Student com $K = 2$ e tamanho de amostra 1000.	82
C.3	Resultados <i>a posteriori</i> * para uma mistura de distribuições t de Student com $K = 2$ e tamanho de amostra 500.	82
C.4	Resultados <i>a posteriori</i> * para a mistura de t de Student com $J = 2$ e $K = 1$ e tamanho de amostra 1000	84
C.5	Resultados <i>a posteriori</i> * para a mistura de t de Student com $J = 2$ e $K = 1$ e tamanho de amostra 500	84
C.6	Resultados <i>a posteriori</i> * para os parâmetros do modelo proposto considerando tamanho de amostra 1000	88
C.7	Resultados <i>a posteriori</i> * para os parâmetros do modelo proposto considerando tamanho de amostra 500	88
C.8	Resultados <i>a posteriori</i> para o modelo de regressão com distribuição de mistura para os erros considerando $n=1000$	91
C.9	Resultados <i>a posteriori</i> para o modelo de regressão com distribuição de mistura para os erros considerando $n=500$	91
C.10	viés, variância e erro quadrático médio para a esperança e variância de Y em diferentes distribuições de mistura quando os dados são gerados de uma distribuição skew-t	92
C.11	viés, variância e erro quadrático médio para a esperança e variância de Y em diferentes distribuições de mistura quando os dados são gerados de uma mistura de skew-t.	94

Lista de Abreviaturas

BDMCMC	Monte Carlo via Cadeias de Markov com processos de nascimento e morte
DIC	Deviance Information Criterion
EM	Expectation-Maximization
ECM	Algoritmo EM Condicional
fdp	função densidade de probabilidade
MCMC	Monte Carlo via Cadeias de Markov
NI	Normal Independente
RJMCMC	Monte Carlo via Cadeias de Markov com Saltos Reversíveis
SMN	Mistura de escala da Normal
SMSN	Mistura de escala da skew-normal
SSMEM	Algoritmo <i>stepwise</i> dividir e unir via algoritmo EM

Capítulo 1

Introdução

A modelagem estatística baseada em misturas finitas de distribuições é uma área de pesquisa em crescente ascensão nos últimos anos. Com o avanço de métodos computacionais nas duas últimas décadas esse tipo de modelagem tem se tornado bastante atrativo tanto do ponto de vista prático quanto teórico, pois permite que densidades com estruturas complexas e que não são bem modeladas por alguma família paramétrica padrão, sejam aproximadas usando uma estrutura mais simples (BÖHNING *et al.*, 2014). Além disso, modelos de mistura fornecem uma alternativa paramétrica para métodos não-paramétricos de estimação de densidades, como por exemplo, a estimação por Kernels (STEPHENS, 1997).

Modelos de mistura são utilizados em diversos contextos: O modelo Poisson inflacionado de zeros, por exemplo, é definido como uma mistura onde um dos componentes possui ponto de massa em zero. Há também modelos com estrutura Markoviana que conseguem capturar dependência de longa duração e heterocedasticidade condicional comuns na análise de séries temporais ou ainda os modelos com classes latentes que envolvem variáveis latentes discretas (BÖHNING *et al.*, 2007; FRÜHWIRTH-SCHNATTER, 2006). Além do que, modelos estatísticos baseados em misturas conseguem capturar propriedades específicas dos dados como multimodalidade, assimetria, caudas pesadas e heterogeneidade decorrente de fatores não-observados, podendo, devido a sua grande flexibilidade, ser aplicados em diferentes áreas como Biologia, Biometria, Genética, Medicina, Marketing, dentre outras (MCLACHLAN; PEEL, 2000).

Uma das primeiras aplicações envolvendo modelos de misturas finitas foram propostas pelo astrônomo e matemático Newcomb (1886) e pelo matemático Pearson (1894). Newcomb (1886) considerou em seu trabalho uma mistura de distribuições normais para modelar dados com *outliers*. Já Pearson (1894) utilizou uma mistura de duas distribuições normais com componentes heterocedásticos para modelar o comprimento de caranguejos na baía de Nápoles. A partir da publicação desses dois artigos a modelagem baseada em misturas finitas de distribuições tornou-se popular e ao longo dos últimos anos várias publicações relacionadas ao tema surgiram em diferentes áreas.

Titterington et al. (1985) fazem uma vasta revisão sobre modelos de mistura englobando desde aplicações nas mais diversas áreas até os aspectos matemáticos desse tipo de modelagem. Até meados dos anos 1980, quando foi publicado o livro de Titterington et al. (1985), a maior parte das aplicações utilizava mistura de distribuições normais, seguidas de misturas de distribuições lognormais e misturas em modelos de regressão. Referências mais recentes sobre modelos de mistura finita incluem os livros de Lindsay (1995), Böhning (2000), McLachlan e Peel (2000), Frühwirth-Schnatter (2006) e Mengersen et al. (2011), além de dois editoriais sobre o tema (BÖHNING et al., 2007; BÖHNING et al., 2014).

Na literatura há uma gama de publicações sobre modelagem estatística baseada em misturas finitas de distribuições normais, começando pelos trabalhos de Newcomb (1886) e Pearson (1894). Muitos autores, incluindo Everitt (1984), McLachlan e Basford (1988), Diebolt e Robert Diebolt e Robert (1994), Nobile (1994), Richardson e Green (1997), McLachlan e Peel (2000), dentre outros, mostraram que misturas de distribuições normais fornecem uma base simples e eficaz para estimação de densidades e modelagem de populações heterogêneas. Entretanto, em muitos problemas práticos onde há presença de observações atípicas (*outliers*) nos dados e/ou dados com caudas pesadas, a distribuição normal pode ter suas estimativas para média e variância seriamente afetadas por ser mais sensível a dados com estas características, podendo, ainda, requerer um número maior de componentes de forma a conseguir aproximar bem a distribuição que gerou os dados, o que pode não ser parcimonioso em termos de modelagem. Neste sentido há uma recente propagação de modelos baseados em misturas com componentes não-normais, tanto no contexto onde a distribuição assumida para os componentes da mistura é simétrica, quanto assimétrica, que buscam modelar de maneira mais flexível, e portanto parcimoniosa, dados de natureza contínua que apresentam *outliers*, caudas pesadas e/ou certa assimetria. No contexto de modelos simétricos a distribuição t de Student tornou-se uma alternativa à distribuição normal para lidar com observações atípicas já que apresenta caudas pesadas, o que pode proporcionar ajustes mais robustos. No contexto de modelos assimétricos, misturas de distribuições Skew-Normal e Skew-t têm ganhado bastante espaço nos últimos anos. É importante mencionar, contudo, que apesar de flexíveis, misturas de distribuições t de Student ou ainda misturas de distribuições Skew-t possuem problemas relacionados à estimação do parâmetro de grau de liberdade, que é conhecido na literatura por ser difícil de estimar. Alguns trabalhos importantes utilizando misturas de distribuição t de Student foram propostos por Peel e McLachlan (2000), Lin et al. (2004) e McLachlan et al. (2006). Em misturas de distribuições assimétricas alguns dos recentes trabalhos foram propostos por Lin et al. (2007a), Lin et al. (2007b), Cabral et al. (2008), Frühwirth-Schnatter e Pyne (2010), Cabral et al. (2012b), Prates et al. (2013) e Lee e McLachlan (2014).

Dentro da modelagem de misturas finitas a classe de modelos de regressão linear tem sido amplamente utilizada para investigar a relação entre variáveis provenientes de grupos latentes desconhecidos. Os primeiros trabalhos a assumirem uma distribuição de misturas finitas para modelar os erros do modelo de regressão foram propostos por [Bartolucci e Scaccia \(2005\)](#) e [Soffritti e Galimberti \(2011\)](#), que consideraram a distribuição dos erros como sendo uma mistura finita de distribuições normais univariadas e multivariadas, respectivamente. Duas abordagens mais recentes foram propostas por [Galimberti e Soffritti \(2014\)](#) e [Benites et al. \(2016\)](#), onde os autores propõem modelar de maneira mais flexível e robusta os erros do modelo de regressão quando estes apresentam multimodalidade, caudas pesadas e/ou assimetria. [Galimberti e Soffritti \(2014\)](#) assumiram que os erros do modelo seguem uma mistura finita de distribuições t de Student multivariadas, enquanto que [Benites et al. \(2016\)](#) consideraram que a distribuição dos erros pertence à classe mistura de escala da Skew-Normal (SMSN). Apesar de serem flexíveis em suas formulações gerais as duas abordagens apresentam limitações decorrentes de restrições impostas nos respectivos modelos como forma de minimizar problemas de otimização. Uma destas limitações refere-se à estimação do parâmetro de grau de liberdade, que por conveniência computacional é assumido não variar entre os componentes da mistura. Em ambos os trabalhos o algoritmo EM foi utilizado para fazer inferência por máxima verossimilhança.

Neste trabalho uma modelagem semi-paramétrica baseada em misturas finitas de distribuições t de Student é introduzida. A abordagem proposta considera em sua especificação estruturas separadas para a modelagem de multimodas/assimetria e comportamento de cauda e diferentemente da mistura com componentes t de Student inicialmente apresentada em [Peel e McLachlan \(2000\)](#), a metodologia considerada neste trabalho estima a estrutura de cauda sem que haja necessidade de se estimar parâmetros de grau de liberdade, uma vez que a estimação deste parâmetro é difícil e computacionalmente custosa. Em linhas gerais o modelo proposto contempla uma mistura finita em dois níveis, onde o primeiro nível contém uma mistura com J componentes e o segundo nível uma mistura com K componentes. A mistura no nível J tem como objetivo modelar unicamente o comportamento multimodal e/ou assimétrico dos dados e cada moda j será modelada a partir de uma mistura finita de K componentes com densidade t de Student. A estrutura proposta tem como objetivo proporcionar uma flexibilidade na estimação de densidades semelhante à modelagem não-paramétrica, uma vez que cada moda existente nos dados será modelada a partir da mistura com K componentes t de Student. Uma extensão da metodologia proposta é apresentada no contexto de modelos de regressão linear, onde assume-se que os erros do modelo de regressão seguem a mistura de distribuições introduzida. A inferência para o modelo proposto é feita sob o paradigma Bayesiano e um algoritmo MCMC é implementado para amostrar da distri-

buição *a posteriori* do modelo. Uma contribuição da implementação desenvolvida que decorre diretamente da classe geral proposta no trabalho é a possibilidade de se obter amostras *a posteriori* para diferentes subclasses do modelo de mistura com componentes t de Student, obtendo-se inclusive a cadeia para o parâmetro de grau de liberdade, que é amostrado em um passo de Metropolis-Hastings. A modelagem proposta é avaliada através de diferentes cenários de simulação desenhados de maneira a contemplar as várias subclasses de modelos englobadas na classe geral proposta neste trabalho. Além de avaliar a performance do modelo proposto e sua extensão para regressão através dos estudos simulados são também consideradas aplicações a quatro conjuntos de dados reais.

O trabalho está organizado da seguinte forma: na Seção 1.1 a metodologia sobre misturas finitas de distribuições será apresentada. A classe de distribuições mistura de escala da Normal, cuja representação estocástica será assumida para definir a densidade de mistura proposta neste trabalho será brevemente discutida na subseção 1.1.1. Na subseção 1.1.2 será abordada a questão da inferência em modelos de mistura finita. Dois tópicos relevantes e que têm forte influência na estimação dos parâmetros do modelo de mistura referem-se à identificabilidade do modelo e ao número de componentes da mistura. A suposição de que existe uma caracterização única para qualquer classe de modelos será discutida na subseção 1.1.2.1. A subseção 1.1.3 aborda alguns aspectos relacionados à mistura finita de distribuições t de Student. A modelagem proposta no presente trabalho será apresentada no Capítulo 2. A Seção 2.1 discute a motivação para este trabalho e apresenta os principais modelos até então existentes na literatura, que podem ser vistos como casos particulares do modelo proposto. A modelagem proposta será apresentada e justificada na Seção 2.2. Os principais aspectos referentes à inferência, via paradigma Bayesiano, e modelagem serão discutidos na Seção 2.3. A Seção 2.4 traz uma extensão do modelo proposto no contexto de modelos de regressão linear. Nesta extensão assumiremos que os erros do modelo de regressão seguem uma mistura finita de distribuições t de Student derivada do modelo apresentado na Seção 2.2. No Capítulo 3 são realizados estudos de simulação para avaliar a metodologia proposta em diferentes cenários. As quatro aplicações com dados reais serão apresentadas no Capítulo 4, onde compararemos o modelo proposto com a mistura de distribuições t de Student já existente na literatura. Por fim, no Capítulo 5 são apresentados os comentários finais sobre o trabalho aqui desenvolvido e algumas propostas para desenvolvimento futuro serão discutidas.

1.1 Modelos de Mistura Finita

Devido à sua flexibilidade em descrever estruturas de dados bastante dissonantes, os modelos de mistura finita têm sido cada vez mais explorados como uma alternativa semi-paramétrica para modelar diferentes distribuições desconhecidas (MCLACHLAN; PEEL, 2000). Em particular, as distribuições de mistura finita conseguem capturar propriedades específicas dos dados como multimodalidade, assimetria, caudas pesadas e heterogeneidade não-observada, além de possuírem uma vantagem em comparação à modelagem não-paramétrica no que se refere ao ganho de precisão nas estimativas dos parâmetros do modelo.

Um dos primeiros trabalhos envolvendo modelos de mistura finita foram propostos por Newcomb (1886) e Pearson (1894), sendo o trabalho de Pearson o que impulsionou a modelagem usando misturas finitas de distribuições. Pearson considerou uma mistura de duas distribuições normais com médias e variâncias diferentes para modelar o comprimento de 1000 caranguejos amostrados na baía de Nápoles. Os dados apresentavam uma leve assimetria à direita e para tentar acomodá-la Pearson considerou uma mistura com componentes heterocedásticos. A metodologia proposta sugeria a presença de duas subespécies distintas nos dados analisados e os resultados obtidos comprovaram a existência de duas espécies de caranguejo. A estimação dos parâmetros do modelo foi feita a partir do método dos momentos.

Para entender como distribuições de mistura finita podem surgir de forma natural em um problema estatístico, considere uma população de interesse onde determinada característica aleatória \mathbf{Y} é observada. A distribuição de mistura surge, em geral, quando uma heterogeneidade devido a fatores não-observados se faz presente na população em questão. Um exemplo de como isso ocorre na prática são os dados analisados em Titterington et al. (1985), referentes ao comprimento de 256 peixes (Figura 1.1). A distribuição empírica destes dados apresenta várias modas e uma possível explicação é que os peixes pertencem a grupos de idade distintos, onde cada grupo possui tamanhos n_1, n_2, \dots, n_J , respectivamente. Devido a heterogeneidade entre os grupos, a distribuição de probabilidade para o comprimento dos peixes será diferente em cada grupo, com o parâmetro θ que indexa tal distribuição diferindo entre os grupos. Em geral a heterogeneidade em uma dada amostra ocorre sempre que a média da característica aleatória \mathbf{Y} difere entre os elementos observados. No exemplo dos peixes a heterogeneidade não-observada surge pois a idade do peixe é uma variável difícil de ser medida. Assim, o comprimento do peixe pode ser um indicativo de qual grupo de idade ele pertence e ao considerar uma densidade de mistura é possível capturar tal heterogeneidade (FRÜHWIRTH-SCHNATTER, 2006).

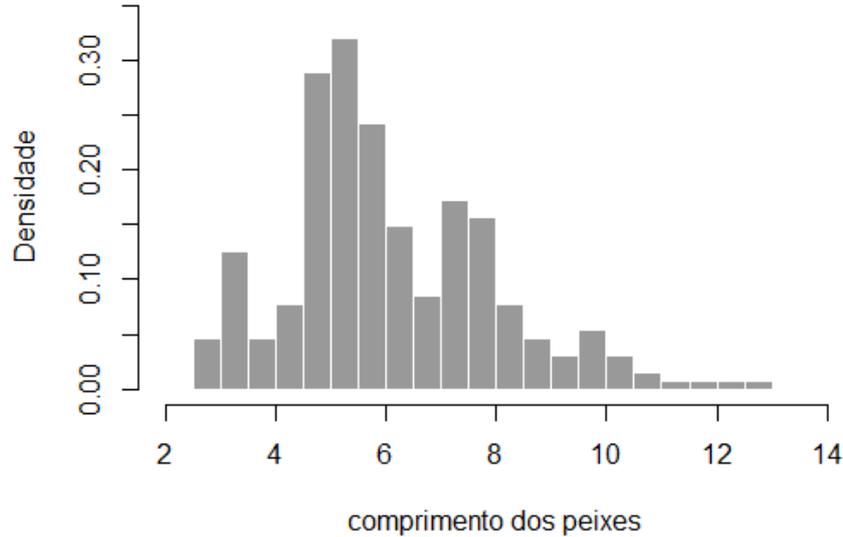


Figura 1.1: Distribuição empírica para o comprimento de 256 peixes.

Definição 1.1.1 : *Uma mistura finita de distribuições é qualquer combinação linear convexa de outras distribuições:*

$$\pi(y) = \sum_{j=1}^J w_j \pi_j(y), \quad \forall y \in \mathcal{Y}, \quad (1.1)$$

onde $\pi_1(\cdot), \pi_2(\cdot), \dots, \pi_J(\cdot)$ são funções densidades de probabilidade com respeito à medida de Lebesgue, denominadas componentes da mistura e w_1, w_2, \dots, w_J os pesos. O vetor $\mathbf{w} = (w_1, w_2, \dots, w_J)$ dos pesos assume valores em um simplex unitário $\xi_J \subset (\mathbb{R}^+)^J$.

Um modelo de mistura finita pode ser representado através da expressão dada em (1.1), sendo também possível caracterizá-lo a partir de uma construção hierárquica ao considerar uma variável latente Z . Desta maneira, a distribuição de uma variável aleatória \mathbf{Y} que tem densidade dada por (1.1) admite a seguinte representação hierárquica:

$$\begin{aligned} (\mathbf{Y}|Z = j) &\sim \pi_j, \quad \forall j = 1, \dots, J, \\ \mathbb{P}(Z = j) &= w_j. \end{aligned}$$

Em muitas situações $\pi_j(\cdot), \forall j = 1, \dots, J$, pertence a uma determinada família paramétrica $\mathcal{P} = \{\pi : \pi(y|\theta); y \in \mathcal{Y} \subset \mathbb{R}^n, \theta \in \Theta\}$, indexada por um parâmetro $\theta \in \Theta$, de maneira que:

$$\pi(y|\boldsymbol{\eta}) = \sum_{j=1}^J w_j \pi(y|\theta_j), \quad \forall y \in \mathcal{Y}, \quad (1.2)$$

onde $\boldsymbol{\eta} = (\theta_1, \dots, \theta_J, \mathbf{w}) \in \Theta^J \times \xi_J$.

O teorema a seguir (DASGUPTA, 2008, Seção 33.1) motiva matematicamente o uso de misturas como modelos probabilísticos em qualquer espaço Euclidiano de dimensão finita.

Teorema 1 : *Sejam $f(y)$ e $g(\cdot)$ funções densidade de probabilidade em \mathbb{R}^n , com $1 \leq n \leq \infty$. Assuma que f e g são contínuas. Então, dado $\varepsilon > 0$ e um conjunto compacto $\mathcal{C} \subset \mathbb{R}^n$, existe $J \in \mathbb{N}$, $\mu_j \in \mathbb{R}^n$, $\sigma_j > 0$ e uma mistura finita da forma:*

$$\pi(y) = \sum_{j=1}^J w_j \frac{1}{\sigma_j^n} g\left(\frac{y - \mu_j}{\sigma_j}\right), \quad (1.3)$$

tal que $\sup_{y \in \mathcal{C}} |\pi(y) - f(y)| < \varepsilon$.

Segundo DasGupta (2008), o Teorema acima sugere que qualquer densidade contínua f pode ser bem aproximada (exceto, possivelmente, nas caudas) por uma mistura finita de distribuições na família de locação-escala.

Uma propriedade importante na modelagem de misturas finitas é que a representação de misturas como uma combinação convexa de distribuições implica que os seus momentos também serão combinações convexas de $\pi_j(\cdot)$. Desta maneira:

$$\begin{aligned} \mathbb{E}[\varphi(Y^m|\boldsymbol{\eta})] &= \int_{\mathcal{Y}} \varphi(y^m) \pi(y^m|\boldsymbol{\eta}) dy \\ &= \int_{\mathcal{Y}} \varphi(y^m) \sum_{j=1}^J w_j \pi_j(y|\theta_j) dy \\ &= \sum_{j=1}^J w_j \int_{\mathcal{Y}} \varphi(y^m) \pi_j(y|\theta_j) dy \\ &= \sum_{j=1}^J w_j \mathbb{E}_k[\varphi(Y^m)|\theta_j] \end{aligned}$$

onde $\varphi(\cdot)$ é uma função integrável de \mathbf{Y} .

Segundo Frühwirth-Schnatter (2006), a propriedade mais interessante de uma mistura finita de distribuições é que a forma da densidade é bastante flexível. A Figura 1.2 apresenta algumas das diferentes formas que uma mistura de distribuições normais com variâncias iguais, mas médias e pesos da mistura diferentes pode apresentar.

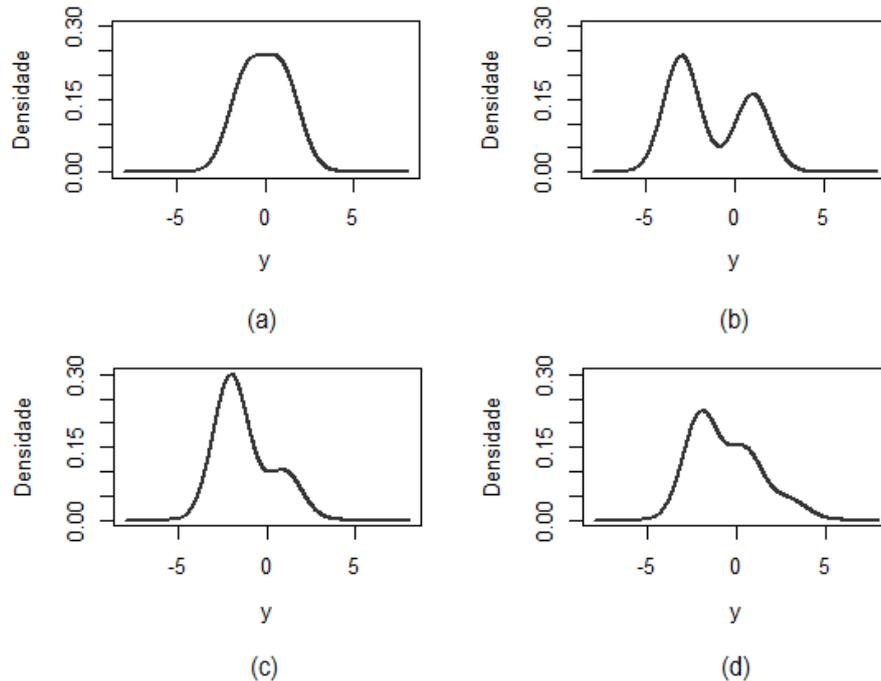


Figura 1.2: Densidades de várias misturas de normais univariadas com $\sigma_1^2 = \sigma_2^2 = 1$. (a) $\mu_1 = -1, \mu_2 = 1$ e $w_1 = 0.5$; (b) $\mu_1 = -3, \mu_2 = 1$ e $w_1 = 0.6$; (c) $\mu_1 = -2, \mu_2 = 1$ e $w_1 = 0.85$; (d) $\mu_1 = -2, \mu_2 = 0.5, \mu_3 = 3, w_1 = 0.55$ e $w_2 = 0.35$.

Devido à sua flexibilidade em modelar populações heterogêneas, as misturas de distribuições normais têm sido uma das mais utilizadas para estimação de densidades, não sendo muito difícil motivar o seu uso tanto em termos práticos quanto teóricos. Na prática muitas medidas e processos naturais tendem a ser normalmente distribuídos, dessa maneira populações que contêm subpopulações de tais medidas tenderão a ter densidades semelhantes a uma mistura finita de distribuições normais (NGUYEN, 2015). Entretanto, é importante ressaltar que o uso da distribuição normal pode ser inadequado em situações onde há presença de observações atípicas nos dados ou ainda quando os mesmos possuem caudas pesadas ou certa assimetria (LIN et al., 2007b; SVENSÉN; BISHOP, 2005; LIN et al., 2004; PEEL; MCLACHLAN, 2000; STEPHENS, 2000a). Nestes casos a mistura com componentes normais não deixa de ser uma boa aproximação para a verdadeira distribuição geradora dos dados, contudo ela poderá requerer um número maior de componentes na mistura de forma a conseguir aproximar bem a densidade de interesse. Mais componentes implica em ter médias mais próximas, o que pode levar a um possível problema de identificabilidade, além da perda de parcimônia (FRÜHWIRTH-SCHNATTER, 2006). Nesse sentido faz-se necessária a utilização de modelos robustos à presença de *outliers* ou caudas pesadas nos dados, de maneira a fornecer estimativas mais eficientes para os parâmetros do modelo em estudo.

Uma forma de ampliar a família paramétrica de modelos de mistura finita em situações onde os dados apresentam valores extremos e caudas pesadas é assumir, por

exemplo, uma mistura finita de distribuições t de Student. Alguns trabalhos nesta direção foram propostos por [Veaux e Krieger \(1990\)](#), [Peel e McLachlan \(2000\)](#), [Lin et al. \(2004\)](#), [Svensén e Bishop \(2005\)](#) e [McLachlan et al. \(2006\)](#). [Veaux e Krieger \(1990\)](#) consideraram uma mistura de duas distribuições normais e propuseram substituir a componente normal por uma componente t de Student com grau de liberdade fixo e igual a três. Além disso, assumiram a mediana como parâmetro de locação e a mediana do desvio absoluto como parâmetro de escala. [Peel e McLachlan \(2000\)](#) e [Lin et al. \(2004\)](#) assumiram misturas finitas de distribuições t multivariadas e utilizaram métodos MCMC e EM condicional, respectivamente, para fazer inferência dos modelos. O trabalho de [Peel e McLachlan \(2000\)](#) foi o pioneiro em modelagem de misturas finitas utilizando a distribuição t de Student. [Svensén e Bishop \(2005\)](#) desenvolveram uma metodologia Bayesiana baseada em mistura de distribuições t de Student e utilizaram inferência variacional para fazer inferência sobre o modelo proposto. Por fim, [McLachlan et al. \(2006\)](#) fazem uma breve revisão sobre modelos de mistura finita utilizando distribuições t de Student e consideram uma mistura de distribuições t para analisar dados de cluster que possuem grandes dimensões. Além da distribuição t de Student, outras distribuições capazes de acomodar dados de natureza contínua que apresentam *outliers* e caudas pesadas podem ser utilizadas ([MAYRINK; GONÇALVES, 2017](#); [PRATES et al., 2013](#); [NASCIMENTO et al., 2012](#); [WIPER et al., 2001](#)). Uma extensão paramétrica do modelo normal é apresentada a seguir.

1.1.1 Distribuições Mistura de escala da Normal

Esta classe de distribuições foi introduzida por [Andrews e Mallows \(1974\)](#) como uma extensão paramétrica do modelo normal que visa tratar dados com valores extremos preservando sua estrutura simétrica. A classe de distribuições mistura de escala da normal (SMN) é definida a seguir.

Definição 1.1.2 *Uma variável aleatória contínua Y tem distribuição na família mistura de escala da Normal se ela admite a seguinte representação estocástica:*

$$Y = \mu + [\kappa(U)]^{\frac{1}{2}}W, \quad U \perp W, \quad (1.4)$$

onde μ é um parâmetro de locação, U é uma variável aleatória não-negativa com densidade $h(\cdot|\boldsymbol{\nu})$, indexada pelo parâmetro $\boldsymbol{\nu}$, $\kappa(U)$ é uma função positiva de U e $W \sim \mathcal{N}(0, \sigma^2)$, onde $\mathcal{N}(\cdot)$ denota a distribuição Normal.

Se assumirmos $\kappa(U) = U^{-1}$, a distribuição de Y pertencerá à classe de distribuições Normal independente (NI), proposta por [Lange e Sinsheimer \(1993\)](#). Distribuições

na classe NI são capazes de acomodar caudas pesadas, uma vez que esta classe inclui como casos particulares as versões multivariadas das distribuições t de Student (\mathcal{T}), Pearson tipo VII (\mathcal{PVII}), Slash (\mathcal{S}), Normal contaminada (\mathcal{NC}), entre outras.

Considerando-se a representação estocástica em (1.4), tem-se que, condicional a U , $Y|U = u \sim \mathcal{N}(0, u^{-1}\sigma^2)$. Dessa forma, integrando a densidade conjunta de Y e U em relação a U é possível obter a densidade marginal de Y :

$$f(y|\mu, \sigma^2, \boldsymbol{\nu}) = \int_0^\infty f_{\mathcal{N}}\left(\frac{y - \mu}{\sqrt{u^{-1}\sigma^2}}\right) h(u|\boldsymbol{\nu}) du.$$

Na densidade acima $f_{\mathcal{N}}(\cdot)$ representa a densidade da normal, U e $h(\cdot|\boldsymbol{\nu})$ são chamados de fator de escala e distribuição de mistura, respectivamente. É válido ressaltar que casos particulares de distribuições na classe SMN são determinados pela variável U , ou seja:

- Se U é uma variável degenerada em 1, ou seja, se $P(U = 1) = 1$, então $Y \sim \mathcal{N}(\mu, \sigma^2)$;
- Se $U|\nu \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, com $\nu > 0$, e $\mathcal{G}(\cdot)$ denotando a distribuição gama, então temos que $Y \sim \mathcal{T}(\mu, \sigma^2, \nu)$, com fdp dada por:

$$f_Y(y; \mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\sigma^2\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(y - \mu)^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}},$$

onde μ , σ e ν representam, respectivamente, os parâmetros de locação, escala e graus de liberdade.

- Se $U|\nu, \delta \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\delta}{2}\right)$, com $\nu, \delta > 0$, então $Y \sim \mathcal{PVII}(\mu, \sigma^2, \nu, \delta)$, com fdp dada por:

$$f_Y(y; \mu, \sigma^2, \nu, \delta) = \frac{1}{\text{B}(\nu/2, 1/2)\sqrt{\sigma^2\delta}} \left(1 + \frac{(y - \mu)^2}{\sigma^2\delta}\right)^{-\frac{\nu+1}{2}},$$

onde μ , σ e (ν, δ) representam, respectivamente, os parâmetros de locação, escala e forma (quando $\nu = \delta$, caímos no caso anterior); $\text{B}(\cdot)$ representa a função beta.

- Se $U|\nu \sim \mathcal{B}(\nu, 1)$, com $\nu > 0$, e $\mathcal{B}(\cdot)$ denotando a distribuição Beta, então temos que $Y \sim \mathcal{S}(\mu, \sigma^2, \nu)$, com fdp dada por:

$$f_Y(y; \mu, \sigma^2, \nu) = \nu \int_0^1 u^{\nu-1} \phi\left(\frac{y - \mu}{u^{-1/2}\sigma}\right),$$

onde μ , σ^2 e ν são parâmetros de locação, escala e graus de liberdade, respectivamente.

- Se U é uma variável aleatória discreta do tipo:

$$U = \begin{cases} \gamma, & \text{com probabilidade } \nu \\ 1, & \text{com probabilidade } 1 - \nu \end{cases},$$

então $Y \sim \mathcal{NC}(\mu, \sigma^2, \nu, \gamma)$, com fdp dada por:

$$f_Y(y; \mu, \sigma^2, \nu, \gamma) = \nu \phi(y|\mu, \gamma^{-1}) + (1 - \nu)\phi(y).$$

A Figura 1.3 mostra algumas das densidades pertencentes à classe de distribuições SMN, assumindo média zero e variância dois para todas as distribuições. Além disso, foi assumido $\nu = 3$ para a distribuição t de Student e $\nu = 1.25$ para a Slash, de forma que ambas pudessem ser comparáveis em termos da magnitude do referido parâmetro. Na normal contaminada foram assumidos $\nu = 0.5$ e $\gamma = 3$.

Ao observar inicialmente apenas a Figura 1.3(a) pode-se chegar a uma conclusão errônea no que se refere ao comportamento de cauda das densidades consideradas. Na referida figura a distribuição Normal aparenta possuir cauda mais pesada em comparação às demais distribuições. No entanto, ao analisar o comportamento de cauda das quatro densidades na Figura 1.3(b), é possível notar que as distribuições t de Student e Slash, respectivamente, são as que possuem caudas mais pesadas.

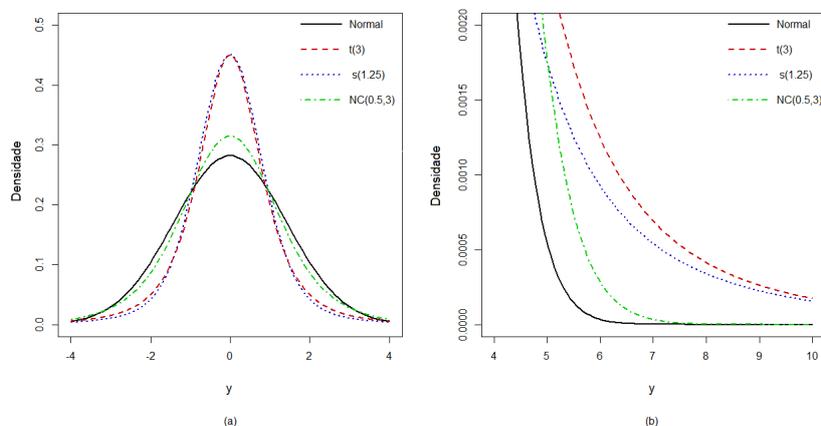


Figura 1.3: Algumas densidades da classe de distribuições SMN

Por ser uma classe de distribuições capaz de acomodar *outliers* e/ou observações influentes, a SMN tem sido largamente utilizada na literatura. [Fernandez e Steel \(2000\)](#), por exemplo, fizeram inferência Bayesiana e também assumiram que a distribuição para os termos do erro no modelo de regressão pertencia à classe SMN. [Peel e McLachlan \(2000\)](#) utilizaram a classe SMN para descrever uma modelagem robusta baseada em misturas finitas de distribuições t multivariadas, enquanto que [Castro et al. \(2014\)](#) utilizaram esta classe para modelar os erros em um modelo de regressão para dados

parcialmente censurados. Mais recentemente [Gonçalves et al. \(2015\)](#) apresentaram uma metodologia para seleção de modelo Bayesiano na classe dos modelos de regressão com distribuições de cauda pesada pertencentes à classe SMN. O método proposto pelos autores considera uma mistura finita de distribuições para modelar uma variável latente onde cada componente da mistura corresponde a um modelo possível dentro da classe SMN. Vale ressaltar que outras distribuições simétricas, membros da família SMN, podem ser usadas para descrever a distribuição marginal de \mathbf{Y} ([BARNDORFF-NIELSEN, 1978](#); [CHOY; CHAN, 2008](#)).

Além da classe SMN é também possível assumir que a distribuição marginal de \mathbf{Y} pertence à família mistura de escala da Skew-Normal (SMSN), proposta por [Branco e Dey \(2001\)](#). A classe SMSN contempla além das distribuições assimétricas (Skew-Normal, Skew-t, Skew-Slash e Skew-Normal contaminada), as distribuições pertencentes à classe SMN. Alguns trabalhos nesta direção foram propostos por [Azzalini \(2005\)](#), [Arellano-Valle e Azzalini \(2006\)](#), [Cabral et al. \(2008\)](#), [Azzalini e Genton \(2008\)](#), [Basso et al. \(2010\)](#), [Cabral et al. \(2012b\)](#) e [Prates et al. \(2013\)](#).

1.1.2 Inferência em modelos de mistura finita

Modelos de mistura finita fornecem uma extensão simples e bastante flexível de modelos estatísticos clássicos, entretanto o preço a ser pago por essa flexibilidade é que a inferência nessa classe de modelos requer um maior esforço computacional ([FRÜHWIRTH-SCHNATTER, 2006](#)). O primeiro método considerado para estimação dos parâmetros em modelos de mistura finita foi o método dos momentos, utilizado por [Pearson \(1894\)](#). Quando comparado a outros métodos, como por exemplo, o de máxima verossimilhança, o método dos momentos pode ser bastante ineficiente. Até meados da década de 1970 quando [Dempster et al. \(1977\)](#) propuseram o algoritmo EM (Expectation-Maximization), a estimação por máxima verossimilhança era realizada a partir da maximização da função de verossimilhança por algum método numérico, como o método de Newton-Raphson ou o método do gradiente.

O uso do algoritmo EM para estimar os parâmetros no modelo de misturas finitas, especialmente mistura de distribuições normais, foi demonstrado por [McLachlan e Basford \(1988\)](#). [Redner e Walker \(1984\)](#) fazem uma revisão e discutem a formulação e propriedades teóricas e práticas do algoritmo EM para modelos de mistura finita, em particular misturas de densidades na família exponencial. [Feng e McCulloch \(1996\)](#) estabeleceram um resultado de consistência para modelos de mistura finita onde apresentam condições para os quais os estimadores de máxima verossimilhança irão convergir para um ponto arbitrário em um subconjunto do espaço paramétrico que representa o

verdadeiro modelo, permitindo que os estimadores se aproximem de um ponto limite do espaço paramétrico. Apesar da estimação por máxima verossimilhança ser uma das mais utilizadas em modelos de mistura finita, ela possui algumas dificuldades práticas. Uma dessas dificuldades é que pode ser difícil encontrar o máximo global numericamente. Além do mais, podem ocorrer falhas de convergência, especialmente quando o tamanho da amostra é pequeno ou os componentes do modelo de mistura não são bem separados (FRÜHWIRTH-SCHNATTER, 2006).

Além da inferência por máxima verossimilhança, a estimação dos parâmetros sob a perspectiva Bayesiana, via MCMC, tem ganhado bastante espaço nos últimos anos. Uma das razões para se adotar a abordagem Bayesiana em modelos de mistura finita é que o uso de distribuições *a priori* próprias introduz, em geral, um efeito suavizador na função de verossimilhança, reduzindo o risco de se obter falsas modas (FRÜHWIRTH-SCHNATTER, 2006). O artigo de Diebolt e Robert (1994) é um dos primeiros trabalhos sobre estimação Bayesiana em modelos de mistura finita utilizando MCMC. Além dos métodos MCMC, a inferência em modelos de mistura finita sob o paradigma Bayesiano pode também ser conduzida através de métodos de inferência variacional. Wang e Titterington (2003), por exemplo, provaram que, para modelos de mistura envolvendo componentes da mistura conhecidos, o estimador variacional de Bayes converge localmente para o estimador de máxima verossimilhança. Uma interessante discussão sobre inferência Bayesiana em misturas finitas de distribuições pode ser encontrada nos trabalhos de Jasra et al. (2005), Marin et al. (2005) e Lee et al. (2008).

Dois pontos importantes a serem destacados na modelagem de misturas finitas, e que têm forte influência na estimação dos parâmetros do modelo, referem-se à identificabilidade do modelo e ao número de componentes da mistura, discutidos a seguir.

1.1.2.1 Identificabilidade em modelos de mistura finita

A suposição de identificabilidade, ou seja, da existência de uma caracterização única para qualquer classe de modelos, é de extrema importância na teoria e prática estatística e quando a mesma não é satisfeita os processos de estimação podem não ser bem definidos (TITTERINGTON et al., 1985). É importante destacar que a identificabilidade é uma propriedade relativa ao modelo e não a algum método específico de estimação. Contudo, se um modelo não é identificável, a inferência pode ser dificultada. A definição matemática de identificabilidade é dada a seguir.

Definição 1.1.3 : *Uma família paramétrica \mathcal{P}_θ é dita ser identificável se distintos valores de θ correspondem a distribuições distintas. Ou seja, se $\theta \neq \theta'$, para $\theta, \theta' \in \Theta$, implica em $\pi_\theta \neq \pi_{\theta'}$.*

Definição 1.1.4 : *Define-se como classe das misturas finitas da família \mathcal{P} , a família*

$$\mathcal{H} = \left\{ \pi : \pi(y|\boldsymbol{\eta}) = \sum_{j=1}^J w_j \pi_j(y|\theta_j); \quad \pi_j(y|\theta_j) \in \mathcal{P}, \quad J \in \mathbb{N} \right\}. \quad (1.5)$$

Em misturas finitas a identificabilidade é definida de maneira ligeiramente diferente da definição usualmente conhecida. Em geral, é necessário distingui-la entre dois tipos: A não-identificabilidade conhecida como "*label-switching*", causada pela invariância da distribuição de mistura ao reclassificar os seus componentes; e a não-identificabilidade causada pelo potencial sobreajuste do modelo de mistura.

A não-identificabilidade causada pelo problema de "*label-switching*" foi notada inicialmente por [Redner e Walker \(1984\)](#) e ocorre quando os componentes da mistura são permutados. Já a não-identificabilidade causada por um potencial sobreajuste do modelo foi abordada inicialmente por [Crawford \(1994\)](#), que mostrou que qualquer mistura com $J - 1$ componentes define um subconjunto não-identificável no espaço paramétrico Θ^J que contempla as misturas finitas de J componentes.

O conceito de identificabilidade em misturas finitas de distribuições pode ser formalmente definido como:

Definição 1.1.5 : *Suponha que $\pi(y|\boldsymbol{\eta})$ e $\pi'(y|\boldsymbol{\eta}')$ são membros da classe \mathcal{H} dada em (1.5), isto é:*

$$\begin{aligned} \pi(y|\boldsymbol{\eta}) &= \sum_{j=1}^J w_j \pi_j(y|\theta_j), \\ \pi'(y|\boldsymbol{\eta}') &= \sum_{j=1}^{J'} w'_j \pi'_j(y|\theta'_j), \end{aligned}$$

e que $\pi(y|\boldsymbol{\eta}) \equiv \pi'(y|\boldsymbol{\eta}')$ se e somente se $J = J'$, e é possível permutar os índices dentro das respectivas somas, tal que $w_j = w'_j$ e $\pi_j(\cdot) = \pi'_j(\cdot)$. Então a classe \mathcal{H} é identificável ([TITTERINGTON et al., 1985](#)).

[Teicher \(1963\)](#) apresentou uma condição suficiente para que a classe de todas as misturas finitas da família de funções de distribuição acumulada $\mathcal{F} = \{F_1, \dots, F_J\}$ seja identificável. Segundo Teicher a identificabilidade ocorre se existirem J valores reais $\{y_1, \dots, y_J\}$ para os quais o $\det(F_j(y_l)) \neq 0$, para $l \leq J$. O autor também estabeleceu a identificabilidade para toda mistura finita de distribuições normais unidimensionais, além das misturas finitas de distribuições gama. [Yakowitz e Spragins \(1968\)](#) apresentaram, em contrapartida, uma condição suficiente e necessária para que a classe de todas as misturas finitas seja identificável. Segundo os autores a família \mathcal{F} induz misturas finitas identificáveis se ela for um conjunto linearmente independente em \mathbb{R} . Mais

recentemente, [Holzmann et al. \(2006\)](#) apresentaram resultados gerais de identificabilidade para misturas finitas de distribuições elípticas e mostraram, em particular, que misturas finitas de distribuições t de Student multivariadas (o resultado também vale para o caso univariado) são identificáveis, mesmo quando o parâmetro de grau de liberdade varia entre os componentes da mistura. Uma discussão sobre identificabilidade em misturas finitas de distribuições pode ser encontrada em [McLachlan e Peel \(2000\)](#) e [Frühwirth-Schnatter \(2006\)](#).

1.1.2.2 Número de componentes em modelos de mistura

Um problema comum em modelos de mistura finita é que pode ser difícil escolher um número adequado de componentes para a mistura. Na prática o mais usual, principalmente quando a estimação dos parâmetros é feita por máxima verossimilhança, é assumir um número fixo de componentes e a partir de algum critério de informação, como por exemplo o AIC e o BIC, testar, para diferentes valores de J , qual seria o número de componentes mais adequado ([LEROUX, 1992](#); [ROEDER](#); [WASSERMAN, 1997](#)). Contudo, estudos de simulação mostram que o AIC tende a superestimar a escolha do número de componentes, enquanto o BIC tende a subestimá-los ([CUTLER](#); [WINDHAM, 1994](#)). Outras opções de escolha foram propostas ao longo dos últimos anos. [Carlin e Chib \(1995\)](#) estimaram as verossimilhanças marginais dos J componentes e usaram o fator de Bayes para testar o modelo contendo J componentes contra o modelo com $J+1$ componentes. [Zhang e Cheng \(2004\)](#), por exemplo, utilizaram o teste de Kolmogorov-Smirnov estendido para construir um procedimento baseado no algoritmo EM e assim determinar um número ótimo de componentes. Outros autores sugeriram métodos baseados em bootstrap não-paramétrico, medidas de distância, testes de homogeneidade ponderados e validação cruzada para escolha do número de componentes ([MCLACHLAN, 1987](#); [JAMES et al., 2001](#); [SUSKO, 2003](#); [MILOSLAVSKY](#); [LAAN, 2003](#)).

Uma alternativa às abordagens existentes para escolha do número de componentes no modelo de mistura é considerá-lo desconhecido e estimá-lo. Esta opção, contudo, é uma questão bastante delicada em termos de inferência. A maior parte das metodologias até então existentes na literatura utilizam métodos MCMC para estimar o número de componentes da mistura. A primeira metodologia foi proposta por [Richardson e Green \(1997\)](#) e utiliza o método *reversible jump* MCMC (RJMCMC) para amostrar da distribuição conjunta de todos os parâmetros, incluindo o número de componentes.

Alternativamente ao método proposto por [Richardson e Green \(1997\)](#), [Stephens \(2000a\)](#) propôs um método que se baseia na construção de um processo Markoviano (em tempo contínuo) de nascimento e morte (BDMCMC) para estimar o número de componentes. Segundo o autor, métodos MCMC baseados nesse tipo de processo têm

sido largamente utilizados na literatura de processos pontuais para simular realizações de processos pontuais que são difíceis de serem simulados diretamente. Uma extensão do método proposto por Stephens (2000a) foi introduzida por Cappé et al. (2002). Os autores mostram neste trabalho que os algoritmos em tempo contínuo não são restritos à estrutura de nascimento e morte, podendo ser generalizados de forma a incluir outros tipos de saltos em tempo contínuo, como por exemplo, os movimentos de dividir/combinar utilizados Richardson e Green (1997).

Tentando solucionar o problema de estimar o número de componentes por máxima verossimilhança, via algoritmo EM, Wang et al. (2004) propuseram o algoritmo *stepwise* EM (SSMEM). Segundo os autores a principal dificuldade do EM em modelos de mistura é que o número de componentes precisa ser previamente especificado para que o algoritmo funcione de maneira adequada. Uma abordagem mais recente foi apresentada por Nobile e Fearnside (2007). A metodologia proposta pelos autores pode ser vista como uma versão do RJMCMC caracterizada por um espaço de estados que contém apenas o número de componentes da mistura e as variáveis de alocação (variáveis latentes que determinam em que componente da mistura está a i -ésima observação), de maneira que as transições ocorram entre espaços de estados discretos contendo diferentes tamanhos, diferentemente das metodologias baseadas em métodos MCMC anteriormente mencionadas que assumem que o espaço de estados possui dimensão variável.

1.1.3 Sobre Misturas de distribuições t de Student

Nesta seção serão abordados alguns aspectos relacionados à mistura finita de distribuições t de Student. Como anteriormente mencionado, as misturas com componentes t de Student surgiram como uma alternativa robusta à mistura de distribuições normais em situações onde os dados apresentam observações atípicas ou caudas pesadas. Nestas situações algumas questões acerca da modelagem utilizada podem ser levantadas:

- i. Se os dados vierem de uma distribuição simétrica unimodal, justifica a utilização de um modelo de mistura para modelar comportamento de cauda?

Supondo que a distribuição que gerou os dados seja simétrica unimodal e possua caudas pesadas, é possível avaliar através de um estudo empírico se justifica utilizar a estrutura de um modelo de mistura para modelar comportamento de cauda. O estudo pode ser conduzido fazendo uso, por exemplo, da distância de Kolmogorov-Smirnov

que mede o quanto duas distribuições de probabilidade unidimensionais subjacentes diferem. Neste caso as funções de distribuição da t de Student e de uma mistura finita de distribuições t de Student são comparadas. A ideia consiste em avaliar se existe pelo menos uma distribuição t com determinada configuração de parâmetros cuja distância de Kolmogorov-Smirnov entre esta configuração e a densidade de mistura assumida seja suficientemente pequena.

A título de ilustração considere uma mistura de distribuições t de Student com $K = 3$ componentes e a seguinte configuração de parâmetros: $\mu = 0$, $\sigma^2 = (0.7, 1, 1.3)$, $\nu = (1.5, 2.5, 3.5)$ e três diferentes configurações de peso $\mathbf{w} = (0.33, 0.34, 0.33)$, $\mathbf{w} = (0.6, 0.3, 0.1)$ e $\mathbf{w} = (0.1, 0.3, 0.6)$. A comparação entre a densidade de mistura e uma única t será baseada na situação genérica em que ambos os parâmetros ν e σ^2 da distribuição t de Student são estimados. Diferentes configurações da distribuição t de Student são consideradas em grids definidos simultaneamente para ν e σ^2 . A distância de Kolmogorov-Smirnov entre a densidade de mistura e cada uma das possíveis combinações de parâmetros na densidade da t mostrou que haverá pelo menos uma distribuição t que se aproxima da mistura, inclusive nas caudas da distribuição, para as três configurações de peso assumidas na densidade de mistura (Figura 1.4). Os resultados obtidos a partir do estudo empírico indicam que existe uma densidade t de Student que se aproxima da densidade de mistura e neste caso não justifica utilizar um modelo de mistura para modelar comportamento de cauda em dados que possuem distribuição unimodal, visto que o modelo mais simples fornece bons resultados.

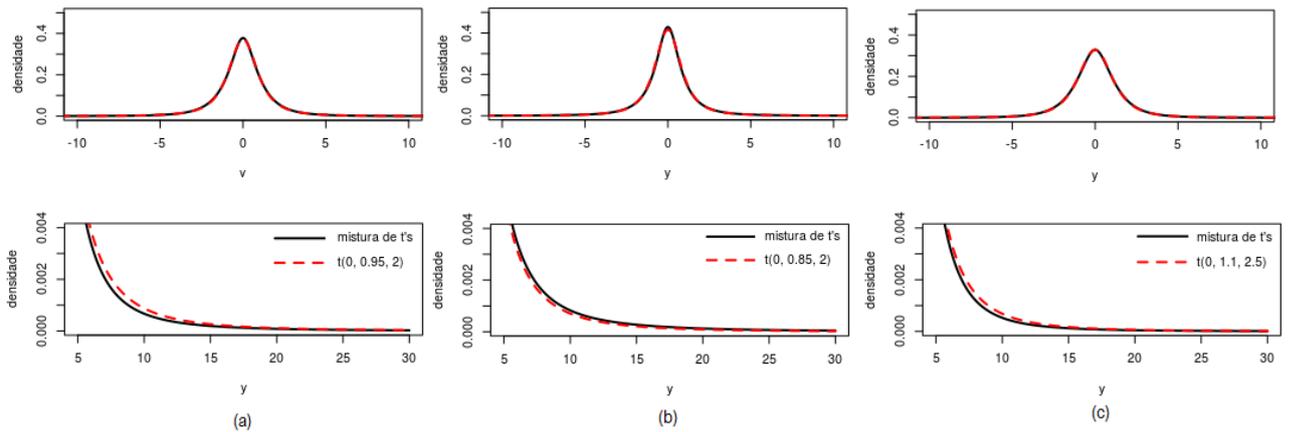


Figura 1.4: Comparação entre a Mistura de t 's com (a) $\mathbf{w} = (0.33, 0.34, 0.33)$, (b) $\mathbf{w} = (0.60, 0.30, 0.10)$ e (c) $\mathbf{w} = (0.1, 0.3, 0.6)$, e as densidades da t de Student com menores distâncias de Kolmogorov-Smirnov.

- ii. Em situações onde são utilizadas misturas de distribuições t de Student com graus de liberdade fixos, a mistura é capaz de aproximar uma densidade t de Student com grau de liberdade definido no intervalo fixado para os graus de liberdade da mistura?

É possível mostrar empiricamente que misturas finitas de distribuições t de Student com graus de liberdade fixos conseguem aproximar bem uma distribuição t com grau de liberdade arbitrário. Em particular, é essencial que o parâmetro ν na densidade da t esteja no intervalo de variação dos graus de liberdade assumidos nas componentes da mistura. A Figura 1.5 mostra três densidades t de Student com $\mu = 0$, $\sigma^2 = 1$ e graus de liberdade 2, 3 e 4, respectivamente, e a densidade de mistura assumindo $\mu = 0$, $\sigma^2 = 1$, $\nu = (1.5, 2.5, 5)$ com as respectivas combinações de peso \mathbf{w} que melhor aproximam as três densidades da t, segundo a distância de Kolmogorov-Smirnov. O estudo consiste em avaliar se existe pelo menos uma combinação de pesos na densidade de mistura cuja distância de Kolmogorov-Smirnov em relação à densidade da t é suficientemente pequena. Neste caso um grid é definido para o vetor \mathbf{w} e todas as possíveis combinações dentro deste grid são consideradas, de maneira que a soma dos pesos seja igual a um. Cada configuração possível em \mathbf{w} induz uma densidade de mistura diferente e para cada uma delas é calculada a distância de Kolmogorov-Smirnov em relação a cada uma das três densidade da t de Student.

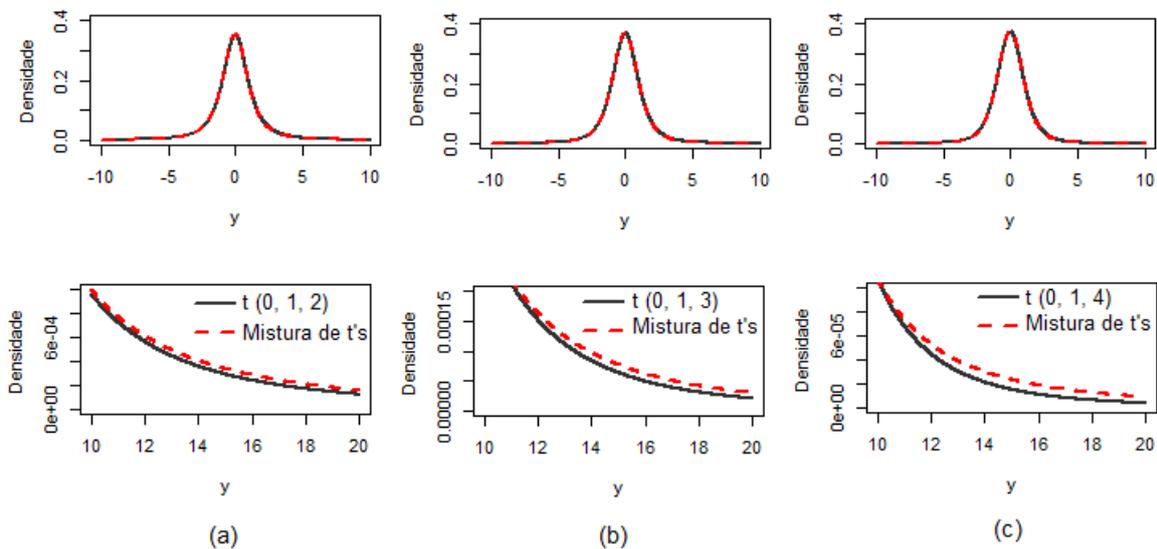


Figura 1.5: Comparação entre densidades t de Student e Misturas de t's com $\nu = (1.5, 2.5, 5)$ e (a) $\mathbf{w} = (0.43, 0.45, 0.12)$, (b) $\mathbf{w} = (0.02, 0.45, 0.53)$ e (c) $\mathbf{w} = (0.01, 0.1, 0.89)$.

Um tópico que requer atenção quando a distribuição t de Student é adotada

para modelar dados com observações atípicas e caudas pesadas refere-se à estimação do parâmetro de grau de liberdade, que é conhecida na literatura por ser difícil e custosa (FERNANDEZ; STEEL, 1999; FONSECA et al., 2008; VILLA et al., 2014). Em modelos de regressão com erros t de Student, por exemplo, quando $\nu \rightarrow 0$ a função de verossimilhança tende a infinito. E, mesmo quando o espaço paramétrico está restrito a uma região onde a função de verossimilhança é limitada, o estimador de máxima verossimilhança pode não existir com probabilidade positiva (FERNANDEZ; STEEL, 1999, Teorema 5). Liu e Rubin (1995) mostraram que a convergência do algoritmo EM na estimação de uma única componente t é bastante lenta quando ν é assumido desconhecido, e neste caso sugerem a utilização do algoritmo EM Condicional com um passo E adicional (ECME) na tentativa de obter uma convergência mais rápida. No caso de mistura com componentes t de Student, Peel e McLachlan (2000) justificam que a implementação do algoritmo ECME para estimação de ν não é simples e, em decorrência desta dificuldade computacional, este parâmetro é em geral escolhido de maneira adaptativa fazendo uso da verossimilhança perfilada, onde a verossimilhança para ν é calculada em um grid de pontos previamente definido, sendo geralmente estimado um único ν para todo componente j da mistura. Sob o paradigma Bayesiano há algumas propostas de distribuição *a priori* para ν . Lin et al. (2004), por exemplo, adotaram uma distribuição *a priori* uniforme contínua ($\mathcal{U}(0, h)$, h suficientemente grande) para estimação de ν no contexto de modelos de mistura. Fonseca et al. (2008) propuseram duas distribuições *a priori* objetivas de Jeffrey's para estimação de ν no contexto de modelos de regressão. Vallejos e Steel (2013) mostraram, entretanto, que uma das distribuições *a priori* proposta por Fonseca et al. (2008) não fornece uma distribuição *a posteriori* própria e dão uma condição necessária para sua existência. Cabral et al. (2012a) propuseram um modelo baseado em misturas finitas para modelar a distribuição conjunta do erro e efeito aleatórios em um modelo de efeitos mistos e adotaram uma distribuição *a priori* hierárquica para ν . Os autores assumiram que $\nu \sim \exp(\lambda)$, com λ seguindo uma distribuição uniforme contínua no segundo nível de hierarquia. Recentemente Villa et al. (2014) propuseram uma *priori* objetiva para ν que se baseia em uma função de perda construída a partir da divergência de Kullback-Leibler. Há também o trabalho de Martins et al. (2014) onde os autores propõem uma família geral de distribuições *a priori* para parâmetros de flexibilidade que inclui parâmetros de comportamento da cauda.

Uma vez que a estimação do parâmetro de grau de liberdade é problemática e computacionalmente cara, faz-se necessário buscar estratégias de modelagem que consigam remediar o problema sem perder robustez e flexibilidade. No próximo capítulo uma abordagem baseada em misturas finitas com componentes t de Student que separa a modelagem de modas/assimetria e comportamento de cauda será apresentada.

Capítulo 2

Modelo Proposto

2.1 Motivação

É comum observar conjuntos de dados que apresentam multimodalidade, assimetria e caudas pesadas. A Figura 1.1, apresentada na Seção 1.1, representa um típico conjunto de dados onde estas características são simultaneamente observadas. A distribuição empírica para o comprimento dos peixes possui um comportamento multimodal com uma leve assimetria na cauda à direita. A multimodalidade neste caso pode ser decorrente de uma heterogeneidade não-observada, possivelmente explicada pela existência de diferentes grupos de idade entre os peixes. O comprimento do peixe pode ser um indicativo de qual grupo de idade ele pertence, e neste caso a assimetria à direita pode sugerir que peixes com comprimentos maiores são em geral mais velhos, o que justificaria a baixa frequência observada na cauda da distribuição. Para dados com esse tipo de característica, é possível definir um modelo que seja flexível, parcimonioso e capaz de capturar simultaneamente multimodalidade, assimetria e caudas pesadas?

Como anteriormente mencionado, a modelagem estatística baseada em misturas finitas possui grande flexibilidade em descrever dados com diferentes estruturas, sendo portanto bastante atrativa na estimação de densidades e modelagem de populações heterogêneas. Entretanto, vale ressaltar que para capturar estruturas como as da Figura 1.1 é necessário que o modelo de mistura adotado estime a estrutura de cauda e como pontuado no final do capítulo anterior, esta nem sempre é uma tarefa fácil na prática.

Um possível modelo que pode responder a pergunta acima seria assumir, por exemplo, que a distribuição dos dados segue uma mistura finita de normais com componentes heterocedásticos, como proposto por [Pearson \(1894\)](#):

$$\pi(y|\boldsymbol{\eta}) = \sum_{j=1}^J w_j f_{\mathcal{N}}(y|\mu_j, \sigma_j^2),$$

onde $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{w})$ e $f_{\mathcal{N}}(\cdot)$ denota a fdp de uma distribuição normal.

Uma mistura de normais é capaz de lidar com a heterogeneidade não-observada nos dados capturando, por exemplo, a multimodalidade existente. Porém, se os dados apresentarem observações atípicas e caudas pesadas, este modelo pode ser inapropriado (LIN et al., 2007a; LIN et al., 2004; PEEL; MCLACHLAN, 2000). A ocorrência simultânea de multimodalidade, assimetria e caudas pesadas nos dados pode requerer um número maior de componentes na mistura de normais, podendo causar problemas de identificabilidade, além de deixar o modelo menos parcimonioso.

Uma alternativa mais flexível seria assumir, por exemplo, que a distribuição dos dados segue uma mistura finita de distribuições t de Student como proposto nos trabalhos de Peel e McLachlan (2000) e Lin et al. (2004):

$$\pi(y|\boldsymbol{\eta}) = \sum_{j=1}^J w_j f_{\mathcal{T}}(y|\mu_j, \sigma_j^2, \nu_j),$$

onde $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\nu}, \boldsymbol{w})$ e $f_{\mathcal{T}}(\cdot)$ denota a fdp de uma distribuição t de Student.

A mistura com componentes t de Student remedia o problema de lidar com observações atípicas e caudas pesadas nos dados, visto que a distribuição t contém um parâmetro adicional responsável por modelar o comportamento de cauda da distribuição. Outra opção também flexível seria assumir que as componentes do modelo são modeladas a partir de distribuições assimétricas, como a Skew-t, por exemplo. Trabalhos nesta direção foram propostos por Lin et al. (2007a), Lin et al. (2007b) e Frühwirth-Schnatter e Pyne (2010). No entanto, como apontado por Frühwirth-Schnatter e Pyne (2010), apesar da extensão do modelo para o contexto assimétrico parecer bastante natural, a estimação de alguns parâmetros também resulta em problemas computacionais.

Apesar da abordagem proposta inicialmente por Peel e McLachlan (2000) apresentar flexibilidade em sua formulação e assumir ν 's diferentes para cada componente j da mistura, há a restrição de que $\nu_1 = \dots = \nu_J = \nu$, sem a qual não há garantias de que haja convergência no processo de estimação. É válido ressaltar que assumir o mesmo grau de liberdade para todas as componentes da mistura pode ser bastante restritivo, visto que um único ν pode não dar conta de modelar a estrutura de cauda nas diferentes componentes do modelo. Outro ponto importante a ser destacado é que o número de componentes necessário para acomodar multimodalidade e assimetria pode ser diferente do número de componentes para modelar estrutura de cauda. Pensando em uma estrutura de modelo flexível, parcimoniosa e sem graves problemas de estimação, propomos o modelo de mistura definido a seguir.

2.2 O modelo

Definição 2.2.1 : Seja \mathbf{Y} um vetor $n \times 1$ de variáveis resposta. Considere $\mathbf{w} = (w_1, \dots, w_J)^\top$ e $\dot{\mathbf{w}}_j = (\dot{w}_{j1}, \dots, \dot{w}_{jK})^\top$ vetores de peso com dimensão J e K , respectivamente; $\boldsymbol{\mu} = (\mu_1, \dots, \mu_J)^\top$ e $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_J^2)^\top$ vetores de dimensão J , e $\boldsymbol{\nu} = (\nu_1, \dots, \nu_K)^\top$ um vetor de dimensão K , com ν_k fixo e conhecido $\forall k = 1, \dots, K$.

Propomos o seguinte modelo de mistura:

$$f_Y(y) = \sum_{j=1}^J w_j \sum_{k=1}^K \dot{w}_{jk} f_{\mathcal{T}}(y|\eta_j, \nu_k), \quad (2.1)$$

onde $\eta_j = (\mu_j, \sigma_j^2)$, $\forall j = 1, \dots, J$.

Por questões computacionais utilizamos a seguinte representação hierárquica (já com as distribuições *a priori*):

$$(Y_i | Z_{ij} = 1, U_i = u_i, \mu_j, \sigma_j^2) \stackrel{ind}{\sim} \mathcal{N}(\mu_j, \sigma_j^2 u_i^{-1}), \quad \forall i, \quad (2.2)$$

$$(U_i | \dot{Z}_{ijk} = 1, \nu_k) \stackrel{ind}{\sim} \mathcal{G}\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right), \quad \forall i, j, k \quad (2.3)$$

$$\mathbf{Z}_i | \mathbf{w} \stackrel{ind}{\sim} \mathcal{M}(1, w_1 \dots w_J), \quad \forall i \quad (2.4)$$

$$\dot{\mathbf{Z}}_{ij} | \dot{\mathbf{w}}_j \stackrel{ind}{\sim} \mathcal{M}(1, \dot{w}_{j1} \dots \dot{w}_{jK}), \quad \forall i, j \quad (2.5)$$

$$\mathbf{w} \sim \mathcal{D}(\alpha_1, \dots, \alpha_J), \quad (2.6)$$

$$\dot{\mathbf{w}}_j \sim \mathcal{D}(\dot{\alpha}_{j1}, \dots, \dot{\alpha}_{jK}), \quad \forall j \quad (2.7)$$

$$(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \sim \mathcal{NGI}(\mu_0, \tau, \dot{\alpha}, \dot{\beta}). \quad (2.8)$$

Nas expressões (2.4)-(2.8), $\mathcal{M}(\cdot)$, $\mathcal{D}(\cdot)$ e $\mathcal{NGI}(\cdot)$ denotam as distribuições multinomial, dirichlet e normal gama-inversa, respectivamente; $\mathbf{U} = (U_1, \dots, U_n)^\top$ representa um vetor de variáveis aleatórias com dimensão $n \times 1$; $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iJ})^\top$ é um vetor de variáveis latentes com dimensão J e $\dot{\mathbf{Z}}_{ij} = (\dot{Z}_{ij1}, \dots, \dot{Z}_{ijK})^\top$ outro vetor de variáveis latentes com dimensão K , $\forall i = 1, \dots, n$, e $\forall j = 1, \dots, J$. Na estrutura proposta \mathbf{Z}_i e $\dot{\mathbf{Z}}_{ij}$ são variáveis indicadoras que definem em qual componente da mistura em J e K encontra-se o i -ésimo indivíduo.

Buscando resolver os problemas anteriormente levantados, o modelo definido em (2.1) contempla, em linhas gerais, uma mistura finita em dois níveis onde considera estruturas separadas para as modas/assimetria e o comportamento de cauda, de modo que:

$$\mathbf{Y}|\mathbf{U} \sim \sum_{j=1}^J w_j f_{\mathcal{N}}(\mu_j, \sigma_j^2 \mathbf{u}^{-1}), \quad (2.9)$$

$$\mathbf{U} \sim \sum_{k=1}^K \dot{w}_k f_{\mathcal{G}}\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right). \quad (2.10)$$

A representação estocástica para o modelo proposto é baseada na representação da classe SMN apresentada na Seção 1.1.1. A prova da equivalência para (2.1) e sua forma hierárquica em (2.9)-(2.10), encontra-se no Apêndice A. O modelo proposto pode ser justificado da seguinte maneira:

- (i) A mistura a nível J tem como objetivo modelar unicamente o comportamento multimodal e/ou assimétrico dos dados. Ao propor a mistura neste nível é bem razoável fixar o número de componentes baseado em uma análise empírica do histograma dos dados;
- (ii) A modelagem separada da cauda evita uma superparametrização devido à necessidade de várias componentes de moda;
- (iii) Misturar K graus de liberdade distintos, sendo livres para escolher os pesos, flexibiliza o modelo no sentido de capturar distintas e variadas estruturas de cauda;
- (iv) A estrutura de cauda na abordagem proposta é estimada através dos pesos da mistura em K , sem que haja a necessidade de se estimar o parâmetro ν . Neste sentido, o aumento de K penaliza pouco o modelo e o fato dos ν 's serem fixos evita problemas de estimação com este parâmetro;
- (v) Na Seção 1.1.3 do capítulo anterior foi mostrado que com escolhas razoáveis dos ν 's, é possível aproximar bem uma variedade de estruturas de cauda.

Uma das vantagens da abordagem proposta está na flexibilidade do modelo em capturar as diferentes estruturas presentes nos dados sem que haja sub ou sobreparametrização do modelo nem problemas na estimação de determinado parâmetro. Além disso, a ideia de inserir uma mistura de distribuições gama para os U_i 's em (2.10) tem como objetivo dar uma característica não-paramétrica ao modelo, uma vez que escolhas particulares desta distribuição levam a casos específicos da distribuição marginal de \mathbf{Y} , como discutido na Seção 1.1.1. Outra vantagem é que a classe de modelos proposta é bastante geral e engloba como casos particulares os modelos de mistura baseados em componentes t de Student existentes na literatura, ou seja:

1. Se $J = K = 1$ então $\mathbf{Y} \sim \mathcal{T}(\mu, \sigma^2, \nu)$;
2. Se $J = 1$ e $K \geq 2$ então $\mathbf{Y} \sim \sum_{k=1}^K \dot{w}_k \mathcal{T}(\mu, \sigma^2, \nu_k)$;
3. Se $J \geq 2$ e $K = 1$ então $\mathbf{Y} \sim \sum_{j=1}^J w_j \mathcal{T}(\mu_j, \sigma_j^2, \nu)$, onde $\nu_1 = \dots = \nu_J = \nu$;
4. Se $J = K \geq 2$ com $\dot{\mathbf{w}} = (\dot{w}_1, \dots, \dot{w}_J)^\top = \mathbb{1}_{J \times K}$, onde $\mathbb{1}_{J \times K}$ é uma matriz identidade, então $\mathbf{Y} \sim \sum_{j=1}^J w_j \mathcal{T}(\mu_j, \sigma_j^2, \nu_j)$;
5. Se $J \geq 2$ e $K \geq 2$ com $J = K$ (ou $J \neq K$) e $\dot{\mathbf{w}} \neq \mathbb{1}_{J \times K}$, então \mathbf{Y} segue de (2.1).

Uma vez que a classe geral apresentada em (2.1) considera uma mistura finita com componentes t de Student em sua estrutura, é possível então fazer uso dos resultados existentes relacionados a este modelo. Como mencionado na subseção 1.1.2.1, [Holzmann et al. \(2006\)](#) mostraram que misturas finitas de distribuições t de Student são identificáveis. Sendo assim, o modelo teórico proposto neste trabalho carrega esta propriedade.

2.3 Inferência Bayesiana

A inferência para o modelo proposto em (2.2)-(2.8) será realizada via MCMC, mais especificamente, via Gibbs Sampling. Desta maneira, obtém-se uma amostra adequada da distribuição *a posteriori* de todas as componentes desconhecidas do modelo.

O seguinte esquema de blocagem será assumido para o Gibbs:

$$(\mathbf{w}, \dot{\mathbf{w}}_j), \quad (\mathbf{U}, \mathbf{Z}, \dot{\mathbf{Z}}), \quad (\boldsymbol{\mu}, \boldsymbol{\sigma}^2), \quad (2.11)$$

onde $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top$ e $\dot{\mathbf{Z}} = (\dot{\mathbf{Z}}_{1j}, \dots, \dot{\mathbf{Z}}_{nj})^\top$ são matrizes de dimensão $n \times J$ e $n \times K$, respectivamente.

A blocagem escolhida em (2.11) visa minimizar o número de blocos para os quais a amostragem direto da condicional completa ainda é possível. Quanto menor o número de blocos, menor será a autocorrelação da cadeia e mais rápida a sua convergência.

As distribuições condicionais completas de todos os blocos são derivadas a partir da densidade conjunta de todos os componentes do modelo, que é dada por:

$$\begin{aligned} \pi(\mathbf{Y}, \mathbf{U}, \mathbf{Z}, \dot{\mathbf{Z}}, \mathbf{w}, \dot{\mathbf{w}}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \nu) &\propto \pi(\mathbf{Y}|\mathbf{Z}, \mathbf{U}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \pi(\mathbf{U}|\mathbf{Z}, \dot{\mathbf{Z}}, \nu) \pi(\dot{\mathbf{Z}}|\mathbf{Z}, \dot{\mathbf{w}}_j) \pi(\mathbf{Z}|\mathbf{w}) \times \\ &\times \pi(\mathbf{w}) \pi(\dot{\mathbf{w}}) \pi(\boldsymbol{\mu}, \boldsymbol{\sigma}^2). \end{aligned} \quad (2.12)$$

É possível amostrar diretamente de todas as condicionais completas. Detalhes sobre cada condicional completa podem ser vistos no Apêndice B. Na subseção a seguir serão abordados alguns aspectos práticos referentes à implementação computacional da metodologia proposta.

2.3.1 Aspectos Computacionais

O MCMC proposto requer certo cuidado em alguns aspectos, de forma a garantir sua eficiência. Para isso foram adotadas algumas estratégias que serão descritas a seguir.

Para inicializar os vetores \mathbf{w} , $\boldsymbol{\mu}$ e σ^2 será utilizado o pacote `mixsmsn` (PRATES *et al.*, 2013), que implementa rotinas para estimação por máxima verossimilhança, via algoritmo EM, em modelos de mistura finita com componentes pertencendo à classe SMSN. A matriz de pesos $\dot{\mathbf{w}}$ será inicializada assumindo que todos os seus elementos possuem probabilidade $1/K$. A inicialização do vetor U_i será feita pela distribuição *a priori*, dado a grade previamente definida para $\boldsymbol{\nu}$. Os valores assumidos para os hiperparâmetros das distribuições *a priori* de $\boldsymbol{\mu}$ e σ^2 serão $\mu_0 = \bar{Y}$, $\tau = 0.005$, $\hat{\alpha} = 1$ e $\hat{\beta} = 1.5$, respectivamente. Já o valor do hiperparâmetro α nas distribuições *a priori* de \mathbf{w} e $\dot{\mathbf{w}}$ será assumido igual a 1.

Por questões de custo computacional, guardamos apenas a média das variáveis indicadoras para calcular os respectivos pesos *a posteriori*.

Dentro do esquema de blocagem assumido em (2.11), o passo mais complexo é o que amostra do bloco $(\mathbf{U}, \mathbf{Z}, \dot{\mathbf{Z}})$. A estratégia utilizada para amostrar de $\pi(\mathbf{U}, \mathbf{Z}, \dot{\mathbf{Z}}|\cdot)$ foi utilizar a seguinte fatoração:

$$\pi(\mathbf{U}, \mathbf{Z}, \dot{\mathbf{Z}}|\cdot) \propto \pi(\mathbf{U}|\mathbf{Z}, \dot{\mathbf{Z}}, \boldsymbol{\nu})\pi(\dot{\mathbf{Z}}|\mathbf{Z}, \dot{\mathbf{w}}_j)\pi(\mathbf{Z}|\mathbf{w}).$$

A fatoração acima sugere o seguinte esquema de amostragem:

- i. Obtém-se uma amostra da condicional completa de $(\mathbf{Z}|\cdot)$;
- ii. Em seguida, amostra-se da condicional completa de $(\dot{\mathbf{Z}}|\cdot)$;
- iii. Por fim, amostramos da condicional completa de $(\mathbf{U}|\cdot)$.

O algoritmo a seguir sumariza o MCMC proposto.

ALGORITMO

1. Inicialize o algoritmo escolhendo $\mathbf{w}^{(0)}, \dot{\mathbf{w}}_j^{(0)}, \boldsymbol{\mu}_j^{(0)}, \boldsymbol{\sigma}_j^{2(0)}$ e $U_i^{(0)}$, para $i = 1, \dots, n$; $j = 1, \dots, J$, $k = 1, \dots, K$.

2. Para $m = 1, 2, \dots, M$:

(2.1) Para $i = 1, \dots, n$, gere:

$$Z_i^{(m)} \sim \text{Mult}(1, \tilde{p}_{i1}, \dots, \tilde{p}_{iJ}),$$

$$\dot{Z}_i^{(m)} \sim \text{Mult}(1, p_{ij1}, \dots, p_{ijK}), \quad \forall j = 1, \dots, J$$

(2.2) Gere $\mathbf{w}^{(m)}$ e $\dot{\mathbf{w}}_j^{(m)}$ de:

$$(\mathbf{w}|\cdot) \sim \text{Dir}(\alpha_1 + n_1, \dots, \alpha_J + n_J),$$

$$(\dot{\mathbf{w}}_j|\cdot) \sim \text{Dir}(\hat{\alpha}_{j1} + n_{j1}, \dots, \hat{\alpha}_{jK} + n_{jK}), \quad \forall j = 1, \dots, J$$

(2.3) Gere $(U_i|\cdot)^{(m)}$ de $\pi(U_i|\cdot) \sim \mathcal{G}\left(\frac{\nu_k+1}{2}, \frac{\nu_k}{2} + \frac{(y_i-\mu_j)^2}{2\sigma_j^2\gamma_k}\right)$,
 $\forall i = 1, \dots, n_{jk}$ e $\forall k = 1, \dots, K$;

(2.4) Para $j = 1, \dots, J$, gere $(\mu_j^{(m)}, \sigma_j^{2(m)}|\cdot) \sim \mathcal{NGI}(\mu_0^*, \tau^*, \alpha^*, \beta^*)$ da seguinte maneira:

$$(\boldsymbol{\sigma}^2|\cdot) \sim \mathcal{GI}(\alpha^*, \beta^*),$$

$$(\boldsymbol{\mu}|\cdot) \sim \mathcal{N}\left(\mu_0^*, \frac{\boldsymbol{\sigma}^2}{\tau^*}\right).$$

Para evitar o problema de *label-switching*, a geração deste passo é feita via *rejection sampling*. Enquanto o vetor de médias gerado não estiver ordenado, o algoritmo irá rejeitá-lo e continuará amostrando da condicional acima.

2.4 Extensão para modelos de regressão linear com misturas nos erros

A literatura sobre estimação de densidades e modelagem de populações heterogêneas utilizando modelos de misturas finitas já é bastante consolidada. Por outro lado, um grande número de aplicações buscam relacionar uma variável aleatória \mathbf{Y} de interesse com um conjunto de covariáveis com o intuito de investigar a relação entre variáveis provenientes de vários grupos latentes homogêneos desconhecidos.

Uma extensão natural de um modelo independente e identicamente distribuído de misturas seria considerar um conjunto de covariáveis para modelar linearmente a média. Tal extensão, no entanto, pode não ser suficiente para explicar a heterogeneidade dos dados. Uma solução é considerar uma mistura finita para modelar os erros do modelo de regressão. Com a mistura nos erros é possível capturar o efeito de regressores categóricos relevantes que, por algum motivo, não foram considerados no modelo.

Os primeiros trabalhos nesta direção foram propostos por [Bartolucci e Scaccia \(2005\)](#) e [Soffritti e Galimberti \(2011\)](#), que consideraram o problema no contexto univariado e multivariado, respectivamente, e assumiram uma mistura de distribuições normais para modelar os erros do modelo de regressão. Uma abordagem mais flexível e que permite modelar dados que apresentam observações atípicas e caudas pesadas foi apresentada por [Galimberti e Soffritti \(2014\)](#). Os autores propuseram uma mistura finita de distribuições t de Student multivariadas para modelar os erros. Mais recentemente, [Benites et al. \(2016\)](#) consideraram que a distribuição dos erros pertence à classe SMSN. Ambos os trabalhos fizeram inferência do modelo por máxima verossimilhança, via algoritmo EM.

Como extensão do modelo de mistura proposto na Seção 2.2, assumiremos um regressor linear para modelar a média da variável resposta e nosso modelo de mistura apresentado em (2.1) para modelar os erros do modelo de regressão.

Definição 2.4.1 : *Seja \mathbf{Y} um vetor $n \times 1$ de variáveis resposta, \mathbf{X} uma matriz de variáveis explicativas de dimensão $n \times p$ e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ um vetor $p \times 1$ de coeficientes da regressão. A extensão proposta para o modelo de regressão é definida como:*

$$Y_i = \mu_j^* + \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad (2.13)$$

onde $\mu_j^* = \mu_j + \beta_0$, sendo $\beta_0 = \sum_{j=1}^J w_j \mu_j^*$. Assim, $\epsilon_i \sim \sum_{j=1}^J w_j \sum_{k=1}^K \dot{w}_{jk} \mathcal{T}(\mu_j, \sigma_j^2, \nu_k)$, para $i = 1, \dots, n$.

Para garantir a identificabilidade do modelo, impõe-se que $\sum_{j=1}^J w_j \mu_j = 0$. O mo-

delo completo é obtido pelas equações (2.3)-(2.8), fazendo-se $Y_i = \epsilon_i$. Desta maneira, dado (2.13), a fdp de \mathbf{Y} é dada por:

$$f(y_i) = \sum_{j=1}^J w_j \sum_{k=1}^K \dot{w}_{jk} f_{\mathcal{T}}(y_i | \tilde{\mu}_{ij}, \sigma_j^2, \nu_k), \quad \forall i = 1, \dots, n, \quad (2.14)$$

where $\tilde{\mu}_{ij} = \mu_j^* + \mathbf{X}_i^\top \boldsymbol{\beta}$.

As condicionais completas para a extensão apresentada podem ser vistas no Apêndice B.

Capítulo 3

Estudos de Simulação para o Modelo Proposto

Neste capítulo são apresentados os resultados obtidos em seis estudos de simulação conduzidos sob diferentes configurações de cenários com o objetivo de avaliar a performance da abordagem proposta na seção 2.2 do capítulo 2. Cada configuração assumida refere-se a um submodelo contemplado na classe geral de modelos apresentada neste trabalho. Como anteriormente mencionado, a metodologia proposta estima a estrutura de cauda no modelo de mistura sem precisar estimar o parâmetro de grau de liberdade e a flexibilidade trazida por essa estratégia será avaliada dentro de cada um dos estudos realizados. Neste sentido, para todos os estudos de simulação foram feitas escolhas aleatórias para o vetor de parâmetros ν , objetivando avaliar o impacto destas escolhas na estimação da estrutura de cauda. Em situações práticas, contudo, uma possível estratégia para escolher de maneira apropriada o vetor (grade) de ν 's é utilizar a divergência de Kullback-Liebler (DKL), calculada, neste caso, com base na densidade da normal em relação a densidade da t de Student para diferentes graus de liberdade (o pesquisador definirá o ν mínimo e máximo que deseja trabalhar e a partir desta escolha calculará a DKL para diferentes valores de ν abrangendo o intervalo de interesse). A grade de interesse seria então definida a partir de uma sequência de valores que contemplaria a divergência máxima entre as duas densidades e uma divergência mínima (DKL próxima de zero) a partir da qual seria difícil distinguir as duas densidades. Assim, ao fixar o número de componentes para a mistura em K é possível encontrar quais valores de ν corresponderiam à sequência gerada pelas divergências mínima e máxima mencionadas, obtendo-se então a grade de interesse.

Para o primeiro estudo de simulação consideramos que os dados vêm de uma distribuição t de Student assumindo diferentes graus de liberdade na geração dos dados. Neste cenário o objetivo geral é mostrar que uma distribuição t com grau de liberdade arbitrário consegue ser bem aproximada por uma mistura de distribuições t de Student

com graus de liberdade fixos. No segundo estudo simulado os dados possuem distribuição unimodal ($J = 1$) e consideramos uma mistura com K componentes t de Student para geração dos mesmos. Com este cenário objetivamos comparar o ajuste baseado em uma t de Student com grau de liberdade estimado, com o modelo que gerou os dados e que estima a estrutura de cauda com base nos pesos estimados para cada componente da mistura em K . O terceiro estudo considera que os dados vêm de um modelo de mistura com componentes t de Student onde os mesmos graus de liberdade são assumidos para cada componente. Neste estudo comparamos o ajuste do modelo que estima um único ν (por máxima verossimilhança, via algoritmo EM) com o modelo proposto que estima a estrutura de cauda a partir dos pesos $\hat{\mathbf{w}}$. No quarto estudo os dados são gerados a partir do modelo que considera $J \neq K$. Para este estudo assumimos que a estrutura de cauda é diferente para cada moda j . No quinto estudo simulado avaliamos a performance da extensão proposta no contexto de regressão. Assumimos que os erros do modelo seguem a distribuição de mistura apresentada neste trabalho. Além dos estudos que contemplam os submodelos derivados da classe geral, também apresentamos resultados de um estudo onde avaliamos a flexibilidade do modelo proposto para modelar dados vindos de uma distribuição assimétrica.

A performance dos modelos será avaliada a partir dos resultados *a posteriori* obtidos via MCMC. Para os estudos de 1 a 5 (seções 3.1 a 3.5) foi considerada uma cadeia de tamanho 100000 onde as primeiras 15000 iterações foram descartadas (*burn-in*). O *lag* utilizado foi definido a partir do tamanho efetivo da amostra (ESS), calculado com base na *log* densidade *a posteriori* do modelo completo. O *lag* que retornar o maior ESS será o utilizado. A amostra final *a posteriori* foi definida levando em consideração o *burn-in* e o *lag*. Todos os estudos foram realizados no *software* R (R Core Team, 2017).

Um estudo de Monte Carlo (MC) também foi conduzido com o objetivo de definir uma métrica que nos permitisse avaliar a distância das curvas ajustadas pelos modelos em relação à curva real (modelo que gerou os dados). A definição da métrica foi baseada na distância de Kolmogorov-Smirnov (DKS) e para os estudos de simulação das seções 3.2 a 3.4 foram geradas 2500 réplicas de MC para esta distância. Dentro do MC foram geradas a cada iteração amostras de tamanho n do modelo de interesse. A partir da amostra obtida calculamos a DKS da seguinte maneira:

$$DKS = \max_i |F_i^{real} - F_i^{emv}|, \quad (3.1)$$

onde F_i^{real} representa a fda do modelo que gerou os dados e F_i^{emv} a fda do modelo que gerou os dados plugando as estimativas de máxima verossimilhança para μ e σ^2 calculadas com base no modelo verdadeiro. Como exemplo, se os dados foram gerados a partir de uma distribuição t de Student com parâmetros μ , σ^2 e ν (fixo em todos os cenários), encontramos as estimativas para estes parâmetros no modelo t . Para os

demais cenários foi utilizada a mesma ideia.

As curvas ajustadas pelos modelos serão comparadas através do histograma para a amostra MC da DKS com respectivas bandas de 95% de confiança. Para esta comparação a DKS calculada para cada modelo ajustado foi baseada na expressão (3.1), substituindo a F_i^{emv} pela fda do modelo usado para ajuste. Além da DKS também calculamos a distância de variação total (DVT) que mede a distância entre duas funções densidade de probabilidade. Para o cálculo desta distância consideramos que:

$$DVT = \frac{1}{2} \sum_i |f_i^{emv} - \hat{f}_i|.$$

Os modelos que obtiverem os menores valores para a DKS e a DVT em comparação aos seus respectivos valores reais, serão aqueles com melhor aproximação para a distribuição que gerou os dados.

3.1 Dados vindos de uma distribuição t de Student

Esta configuração contempla o modelo mais simples dentre todos que serão considerados nos estudos de simulação. Com este estudo objetivamos mostrar que misturas de distribuições t de Student com graus de liberdade fixos conseguem aproximar bem distribuições t com qualquer grau de liberdade, como mostrado empiricamente na Seção 1.1.3 do Capítulo 1.

Dois cenários foram assumidos para geração dos dados considerando diferentes tamanhos de amostra. No primeiro cenário os dados foram gerados a partir de uma distribuição t de Student com $\mu = 0$, $\sigma^2 = 1$ e $\nu = 2$. O cenário dois também assumiu uma distribuição t de Student na geração, no entanto, consideramos $\mu = 0$, $\sigma^2 = 1$ e $\nu = 3$. A seguinte mistura de distribuições t de Student foi utilizada para ajuste em cada cenário e tamanho de amostra considerados:

$$\sum_{k=1}^3 \dot{w}_k \mathcal{T}(\mu, \sigma^2, \nu_k),$$

onde $\nu = (1.5, 2.5, 5)$.

Os resultados *a posteriori* obtidos para os dois cenários de geração são apresentados nas Tabelas 3.1 e 3.2. A média *a posteriori* para μ e σ^2 é comparada com o EMV de μ e σ^2 no modelo que gerou os dados. Observamos que, em geral, as médias *a posteriori* para μ apresentaram resultados próximos do EMV para todos os tamanhos

de amostra considerados. O efeito do tamanho da amostra nos resultados *a posteriori* pode ser observado em relação ao parâmetro σ^2 . Estes tendem a ter médias *a posteriori* similares ao EMV de σ^2 para tamanhos de amostra maiores. Constatamos que as probabilidades *a posteriori* para \dot{w} atribuem maiores pesos aos graus de liberdade próximos do verdadeiro. Por exemplo, na Tabela 3.1 observa-se que as maiores probabilidades são referentes às componentes do modelo de mistura com $\nu = 1.5$ e $\nu = 2.5$. O mesmo ocorre na Tabela 3.2. Estes resultados corroboram com o estudo empírico apresentado na Seção 1.1.3 do Capítulo 1, onde mostramos que uma mistura de distribuições t de Student com grau de liberdade fixos conseguem aproximar bem uma única t com grau de liberdade arbitrário.

Tabela 3.1: Resultados *a posteriori* para a mistura de t's considerando diferentes tamanhos amostrais e $\nu = 2$ na geração da t de Student

n	EMV	μ	HPD	EMV	σ^2	HPD	\dot{w}
100	-0.0686	-0.0673	[-0.277, 0.164]	0.6673	0.8527	[0.486, 1.247]	(0.243, 0.374, 0.383)
500	-0.0428	-0.0425	[-0.152, 0.076]	1.0270	1.0252	[0.773, 1.270]	(0.546, 0.297, 0.157)
2500	-0.0486	-0.0493	[-0.097, -0.001]	0.9253	0.9625	[0.852, 1.085]	(0.456, 0.335, 0.209)

Tabela 3.2: Resultados *a posteriori* para a mistura de t's considerando diferentes tamanhos amostrais e $\nu = 3$ na geração da t de Student

n	EMV	μ	HPD	EMV	σ^2	HPD	\dot{w}
100	0.1252	0.1248	[-0.164, 0.406]	1.3586	1.3971	[0.860, 1.956]	(0.165, 0.345, 0.490)
500	-0.0321	-0.0308	[-0.131, 0.085]	1.0206	1.0179	[0.807, 1.235]	(0.156, 0.362, 0.482)
2500	-0.0467	-0.0473	[-0.094, -0.002]	0.9646	0.9858	[0.866, 1.101]	(0.045, 0.509, 0.446)

Nas Figuras 3.1 e 3.2 é possível observar que a curva ajustada a partir do modelo de mistura se aproxima da curva real e da baseada no EMV à medida que o tamanho da amostra aumenta. Resultados para a geração de uma t com grau de liberdade 4 podem ser vistos no Apêndice C.

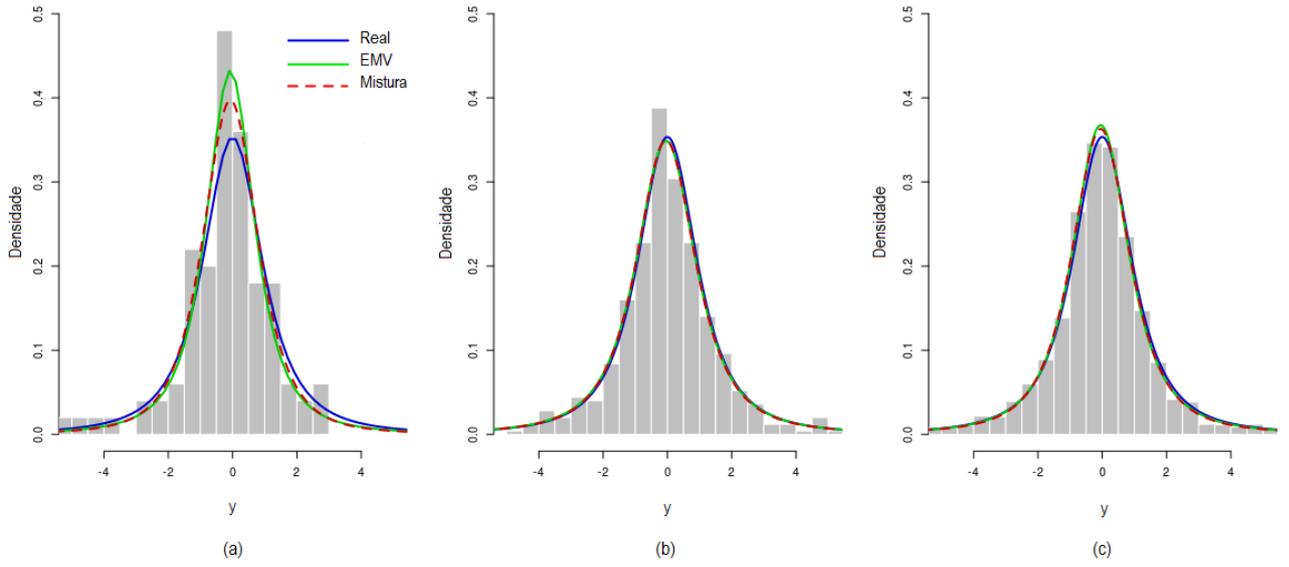


Figura 3.1: Comparação entre as curvas verdadeira, baseada no EMV e no modelo ajustado quando (a) $n=100$; (b) $n=500$ e (c) $n= 2500$.

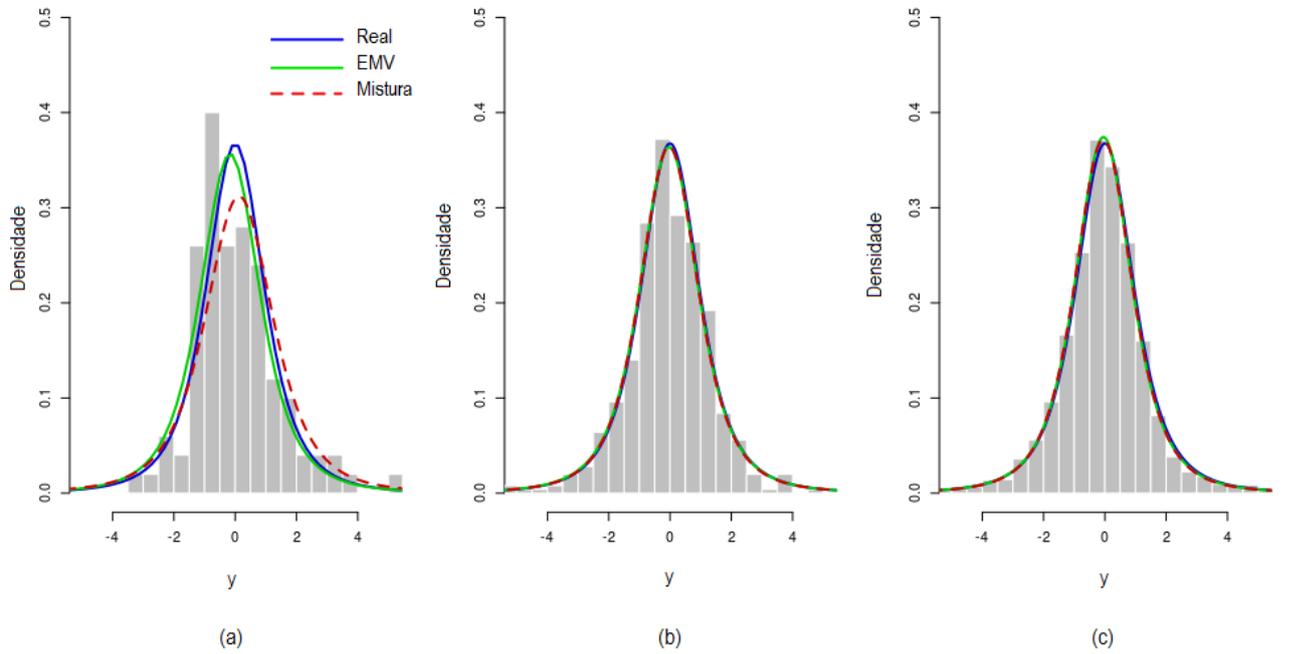


Figura 3.2: Comparação entre as curvas verdadeira, baseada no EMV e no modelo ajustado quando (a) $n=100$; (b) $n=500$ e (c) $n= 2500$.

3.2 Dados vindos de uma mistura de distribuições t de Student unimodal

Nesta configuração a distribuição dos dados é unimodal ($J = 1$) e os geramos a partir de uma mistura com K componentes t de Student assumindo ν 's diferentes para cada componente. Neste cenário a estrutura de cauda é estimada com base nos pesos *a posteriori* obtidos para cada componente da mistura em K . Para o ajuste do modelo de mistura assumimos duas estruturas diferentes para a cauda da distribuição com o intuito de avaliar a capacidade do modelo em identificar a verdadeira estrutura de cauda.

Para geração dos dados consideramos uma amostra de tamanho 2500 com $\mu = 0$, $\sigma^2 = 1$, $\boldsymbol{\nu} = (2.5, 4)$ e $\boldsymbol{w} = (0.5, 0.5)$. Os modelos assumidos para ajuste foram:

- A. única t de Student estimando ν por máxima verossimilhança;
- B. Modelo proposto com $K = 2$ e $\boldsymbol{\nu}$ verdadeiro;
- C. Modelo proposto com $K = 2$ e $\boldsymbol{\nu} = (2.1, 5)$;
- D. Modelo proposto com $K = 3$ e $\boldsymbol{\nu} = (1.8, 3.5, 6)$.

A Tabela 3.3 traz os resultados *a posteriori* para os modelos ajustados. Em todos os modelos ajustados as médias *a posteriori* para μ e σ^2 são comparadas com os EMV's de μ e σ^2 no modelo que gerou os dados. Os respectivos EMV's foram calculados fixando $\boldsymbol{\nu}$ e \boldsymbol{w} em seus valores reais. Observamos que em todos os quatro ajustes a média *a posteriori* de μ coincidiu com o EMV. Para o parâmetro σ^2 nos modelos B, C e D, notamos que seus respectivos intervalos HPD contêm o verdadeiro valor. Em relação às probabilidades *a posteriori* para \boldsymbol{w} nos modelos onde assumimos estruturas diferentes para o vetor de ν 's, é possível observar que a estrutura de cauda foi estimada de maneira satisfatória. O modelo D, por exemplo, atribuiu maiores probabilidades às duas últimas componentes da mistura que se referem aos valores 3.5 e 6 no vetor de ν 's, e estes valores são os que mais se aproximam do ν real.

Tabela 3.3: Resultados *a posteriori** para uma mistura de distribuições t de Student com $K = 2$

Modelos	EMV	$\boldsymbol{\mu}$	HPD	EMV	σ^2	HPD	\boldsymbol{w}	DVT
A		0.021	-		1.016	-	-	0.0072
B	0.021	0.021	[-0.029, 0.070]	1.033	1.025	[0.919, 1.124]	(0.546, 0.454)	0.0016
C		0.021	[-0.026, 0.069]		1.058	[0.946, 1.166]	(0.403, 0.597)	0.0058
D		0.021	[-0.031, 0.066]		1.059	[0.954, 1.169]	(0.193, 0.571, 0.236)	0.0047

$DVT_{real} = 0.0098$ *exceto para o modelo A.

Na Figura 3.3 observamos que apenas o modelo C apresentou um ajuste da curva diferente dos demais. Ao compará-lo com os modelos B e D, nota-se que ele possui a

maior DVT em relação aos demais modelos. Vale ressaltar que dos quatro modelos ajustados, o modelo A foi o que apresentou a maior DVT em relação à densidade que gerou os dados. A Figura 3.4 apresenta o histograma referente à amostra de Monte Carlo para a distância de Kolmogorov-Smirnov. É possível observar que os modelos B e C são os que apresentam as menores DKS's em relação ao valor real (verde). No Apêndice C encontram-se resultados complementares deste estudo para os tamanhos de amostra 500 e 1000. A função de autocorrelação e traço para a cadeia da *log* densidade *a posteriori* do modelo C podem ser vistas no Apêndice D.

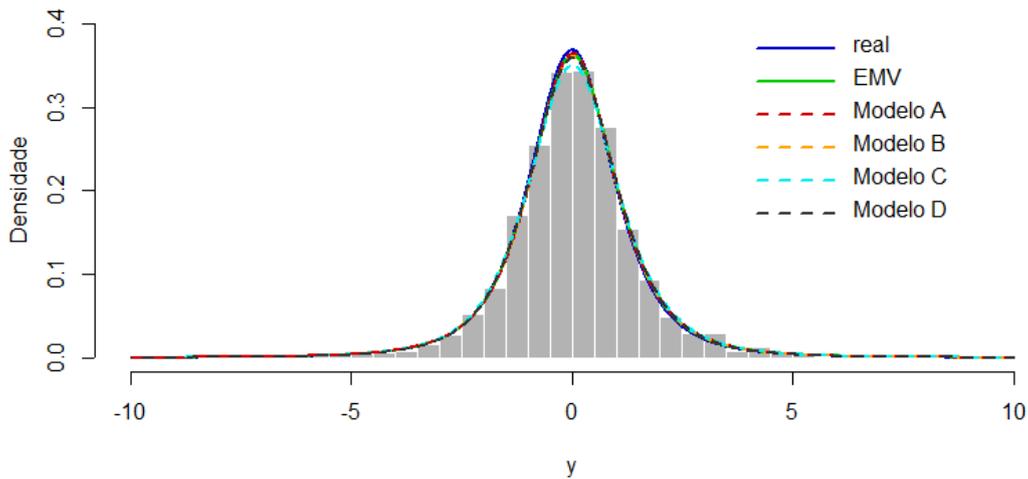


Figura 3.3: Histograma com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes considerando uma amostra de tamanho 2500.

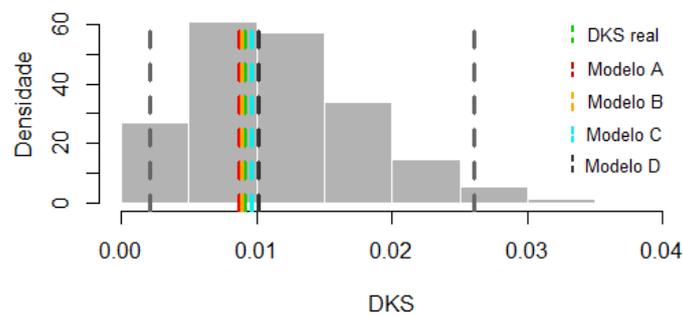


Figura 3.4: Histograma para uma amostra Monte Carlo de tamanho 2500 da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV do modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.

3.3 Dados vindos de uma mistura usual de distribuições t de Student

Para esta configuração os dados foram gerados a partir de uma mistura de distribuições t de Student considerando ν 's iguais para cada componente da mistura. Esta subclasse de modelos ($J \geq 2$, $K = 1$) é usualmente assumida nos trabalhos que utilizam mistura de distribuições t de Student em decorrência dos problemas enfrentados na estimação do parâmetro de grau de liberdade.

Para geração dos dados consideramos uma amostra de tamanho 2500 com $J = 2$, $K = 1$, $\boldsymbol{\mu} = (-1, 2.5)$, $\boldsymbol{\sigma}^2 = (1, 0.75)$, $\nu = 2.5$ e $\boldsymbol{w} = (0.65, 0.35)$. Os seguintes modelos foram utilizados para ajuste:

- A. Mistura com componentes t de Student estimando $\nu_1 = \nu_2 = \nu$ (pacote `mixsmsn`);
- B. Modelo proposto com $J = 2$, $K = 2$ e $\boldsymbol{\nu} = (1.9, 4)$;
- C. Modelo proposto com $J = 2$, $K = 3$ e $\boldsymbol{\nu} = (2.1, 3.5, 5)$.

Os resultados *a posteriori* apresentados na Tabela 3.4 sugerem que os modelos B e C conseguiram ajustar de maneira satisfatória a estrutura adotada na geração dos dados. Os resultados *a posteriori* para $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}^2$ são comparadas com os respectivos EMV's de $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}^2$, calculados com base no modelo que gerou os dados. Observa-se que as médias *a posteriori* para $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}^2$ foram similares às estimativas obtidas no modelo A, que ajustou o modelo verdadeiro e estimou $\nu = 2.67$. Ao avaliar as probabilidades *a posteriori* para $\hat{\boldsymbol{w}}$, observa-se que na primeira moda a maior probabilidade foi atribuída a $\nu = 4$ no modelo B e $\nu = 3$ e 5 no modelo C. Para a segunda moda as probabilidades invertem e os maiores pesos são atribuídos aos graus de liberdade menores.

Tabela 3.4: Resultados *a posteriori* para a mistura de t de Student com $J = 2$ e $K = 1$ e tamanho de amostra 2500

Modelos	EMV	$\boldsymbol{\mu}$	HPD	EMV	$\boldsymbol{\sigma}^2$	HPD	\boldsymbol{w}	$\hat{\boldsymbol{w}}$	DVT
A		-1.040	-	1.047	-	-	0.639	-	0.0071
		2.522	-	0.793	-	-	0.361	-	
B	-1.023	-1.043	[-1.131, -0.968]	1.067	1.132	[0.975, 1.281]	0.633	(0.153, 0.847)	0.0113
	2.526	2.528	[2.432, 2.619]	0.738	0.776	[0.610, 0.945]	0.367	(0.665, 0.335)	
C	-1.046	-1.046	[-1.118, -0.959]	1.125	1.125	[0.967, 1.279]	0.635	(0.258, 0.372, 0.370)	0.0128
		2.526	[2.428, 2.616]	0.812	0.812	[0.640, 0.986]	0.365	(0.600, 0.214, 0.186)	

$$DVT_{real} = 0.0126$$

A curva ajustada em todos os modelos forneceu uma boa aproximação dos dados, no entanto o modelo A foi o que apresentou a menor DVT em relação à verdadeira distribuição que gerou os dados. Este resultado era esperado, uma vez que este modelo

ajustou a mesma estrutura considerada na geração dos dados (Figura 3.5). Em contrapartida, ao avaliarmos o histograma da amostra MC para a DKS, observa-se que os modelos B e C são o que mais se aproximam da DKS real (verde), como mostra a Figura 3.6.

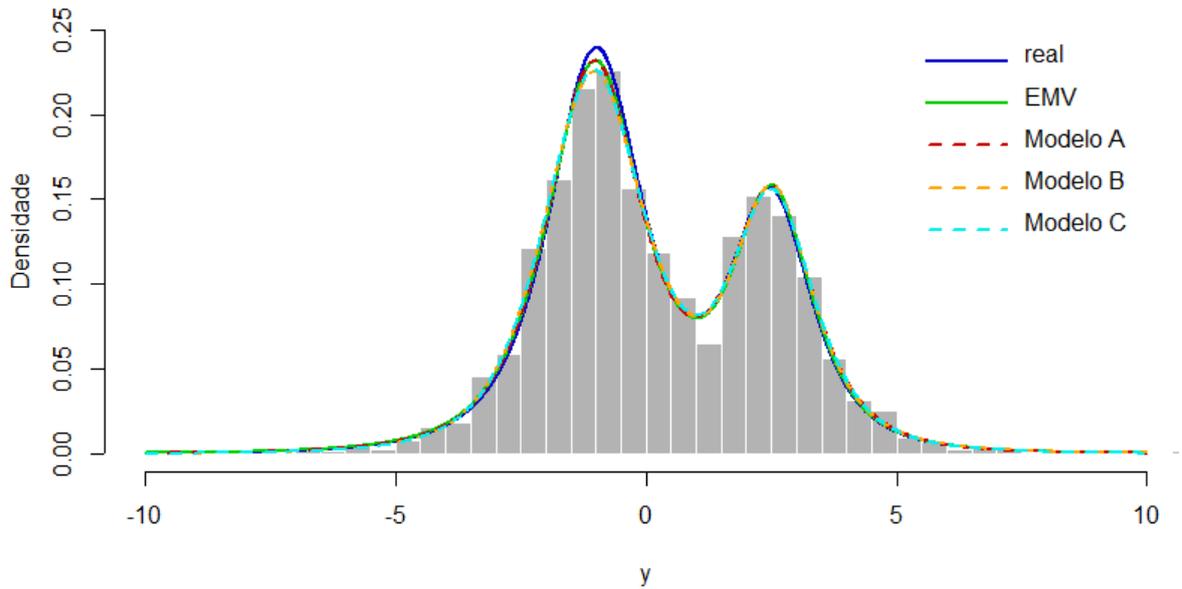


Figura 3.5: Histograma com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes considerando uma amostra de tamanho 2500.

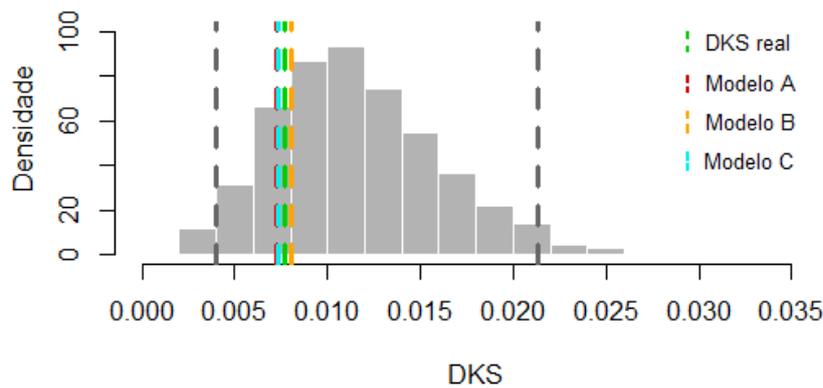


Figura 3.6: Histograma para uma amostra Monte Carlo de tamanho 2500 da distância de Kolmogorov-Smirnov (DKS) entre a função de distribuição real e a baseada no EMV para o modelo proposto, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.

A função de autocorrelação e traço para a cadeia da *log* densidade *a posteriori* do modelo B podem ser vistas no Apêndice D.

3.4 Dados vindos da mistura de distribuições t de Student proposta

Para esta configuração assumimos que a estrutura de cauda pode ser modelada por uma mistura de K componentes, como definido na Seção 2.2 do Capítulo anterior. Neste caso, ao invés de considerar que a estrutura em cada moda j possui o mesmo comportamento, sugerimos que pode existir uma diferença entre eles, e portanto é plausível modelar tal comportamento separadamente para cada moda. Com este intuito, uma amostra de tamanho 2500 foi gerada, considerando $J = 2$, $K = 2$, $\boldsymbol{\mu} = (-1, 2.5)$, $\boldsymbol{\sigma}^2 = (1, 0.75)$, $\boldsymbol{\nu} = (2.3, 3.5)$, $\boldsymbol{w} = (0.65, 0.35)$ e $\dot{\boldsymbol{w}} = \begin{pmatrix} 0.65 & 0.35 \\ 0.35 & 0.65 \end{pmatrix}$. Como forma de avaliar se diferentes escolhas para a grade de $\boldsymbol{\nu}$'s influenciam nas médias *a posteriori*, consideramos os seguintes modelos para ajuste:

- A. Mistura de t's de Student estimando $\nu_1 = \nu_2 = \nu$ (pacote `mixsmsn`);
- B. Modelo proposto com $K = 2$ e $\boldsymbol{\nu}$ verdadeiro;
- C. Modelo proposto com $K = 2$ e $\boldsymbol{\nu} = (2.1, 4.5)$;
- D. Modelo proposto com $K = 3$ e $\boldsymbol{\nu} = (1.9, 3, 5)$.

A Tabela 3.5 apresenta os resultados *a posteriori* obtidos. O Modelo A estimou o parâmetro $\nu = 2.36$. Em relação aos modelos B, C e D observa-se que todos os intervalos HPD contêm o valor real de $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}^2$. Além disso, nota-se que a escolha dos grid's para $\boldsymbol{\nu}$ nos modelos C e D não afetou os resultados, que apresentaram médias *a posteriori* para $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}^2$ bem próximas aos seus respectivos EMV's (calculado com base no modelo verdadeiro). As probabilidades *a posteriori* para o peso \boldsymbol{w} apresentaram estimativas próximas dos valores reais. Em relação ao peso $\dot{\boldsymbol{w}}$, observamos que as grades utilizadas nos modelos C e D conseguiram capturar a estrutura de cauda utilizada na geração dos dados. Para o modelo C as probabilidades *a posteriori* foram maiores quando $\nu = 2.1$, enquanto que no modelo D, $\nu = 1.9$ e 3 apresentaram as maiores probabilidades. Estes resultados são condizentes com as probabilidades obtidas no modelo B, onde o modelo verdadeiro foi ajustado. Resultados para amostras de tamanho 500 e 1000 podem ser vistos no Apêndice C, onde avaliamos se o tamanho da amostra influencia nos resultados *a posteriori*.

Tabela 3.5: Resultados *a posteriori* para o modelo proposto considerando tamanho de amostra 2500

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	w	\hat{w}	DVT
A		-1.014	-		0.961	-	0.631	-	0.0153
		2.443	-		0.826	-	0.369	-	
B	-0.996	-0.996	[-1.065, -0.916]	1.047	1.015	[0.883, 1.178]	0.643	(0.730, 0.270)	0.0057
	2.499	2.481	[2.390, 2.572]	0.740	0.747	[0.577, 0.904]	0.357	(0.799, 0.201)	
C		-0.995	[-1.071, -0.915]		1.024	[0.875, 1.195]	0.640	(0.629, 0.371)	0.0066
		2.481	[2.389, 2.575]		0.721	[0.576, 0.902]	0.360	(0.839, 0.161)	
D		-0.995	[-1.075, -0.922]		1.031	[0.857, 1.180]	0.639	(0.396, 0.321, 0.283)	0.0068
		2.481	[2.384, 2.567]		0.725	[0.554, 0.886]	0.361	(0.628, 0.238, 0.134)	

$$DVT_{real} = 0.00634$$

A Figura 3.7 apresenta o histograma dos dados com a curva real e seus respectivos ajustes. Observa-se que os mesmos foram bem similares. Através do histograma para a distância de Kolmogorov-Smirnov é possível verificar que as menores distâncias em relação à distância real são para os modelos B e C (Figura 3.8). Este resultado corrobora com os valores encontrados para a DVT. O traço para a cadeia da *log* densidade *a posteriori* no modelo C sugere que houve convergência da cadeia (ver Apêndice D).

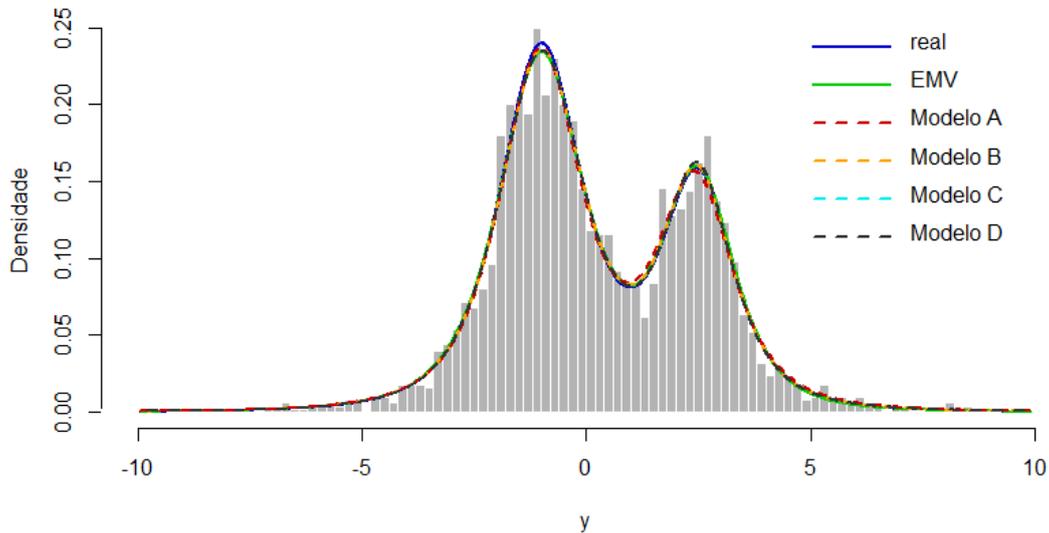


Figura 3.7: Histograma para o modelo usado na geração dos dados e curvas para diferentes ajustes considerando uma amostra de tamanho 2500.

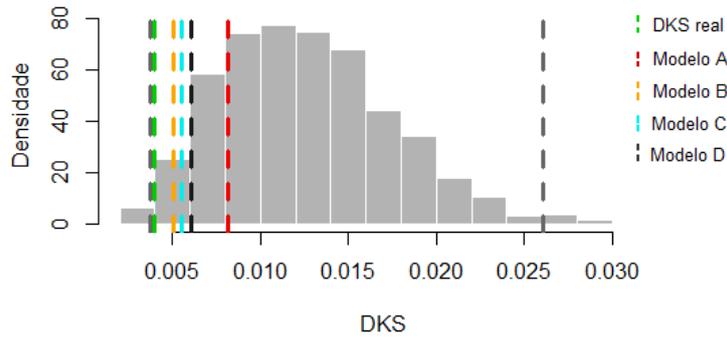


Figura 3.8: Histograma para uma amostra Monte Carlo de tamanho 2500 Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 99% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.

3.5 Modelo de regressão com misturas nos erros

Os estudos de simulação conduzidos nesta seção tiveram como objetivo avaliar a performance da extensão proposta no contexto onde os erros do modelo de regressão seguem a mistura de distribuições apresentada neste trabalho. Para este cenário foram geradas amostras de tamanho 2500. Resultados com tamanhos amostrais 500 e 1000 podem ser vistos no Apêndice C, onde avaliamos o impacto do tamanho amostral nos resultados *a posteriori* obtidos.

i) Erros com distribuição bimodal

Para este cenário assumimos que os erros do modelo de regressão possuem uma distribuição bimodal. Os erros foram gerados supondo $J = 2$, $K = 2$, $\sigma^2 = (1, 0.75)$, $\nu = (2.3, 3.5)$, $w = (0.65, 0.35)$, $\dot{w} = \begin{pmatrix} 0.65 & 0.35 \\ 0.35 & 0.65 \end{pmatrix}$ e $\mu = (-1.88, 3.5)$. Consideramos um modelo com 2 variáveis explicativas, sendo $X_1 \sim \mathcal{N}(0, 1)$ e $X_2 \sim \mathcal{Ber}(0.6)$, com $\beta_0 = 1$ e $\beta = (-2, 1)$. Como consequência dos valores escolhidos para μ e β_0 , temos que $\mu^* = (-0.88, 4.5)$. Vale ressaltar que os valores para w e μ foram escolhidos de forma que a $\sum_{j=1}^2 w_j \mu_j = 0$.

Como forma de avaliar se a grade escolhido para ν afeta os resultados *a posteriori*, assumimos 3 diferentes modelos para ajuste. Os modelos ajustados foram:

- A. $J = K = 2$ com ν verdadeiro;
- B. $J = K = 2$ com $\nu = (2.05, 4.5)$;
- C. $J = 2, K = 3$ com $\nu = (1.9, 3, 5)$.

A Figura 3.9 apresenta o histograma dos erros com a curva real e as ajustadas. Observa-se que não há diferença substancial na estimação da densidade para os três modelos ajustados. As curvas ajustadas conseguem acompanhar bem o comportamento da curva real. A função de autocorrelação e o traço para a cadeia da *log* densidade *a posteriori* no modelo B podem ser vistos no Apêndice D.

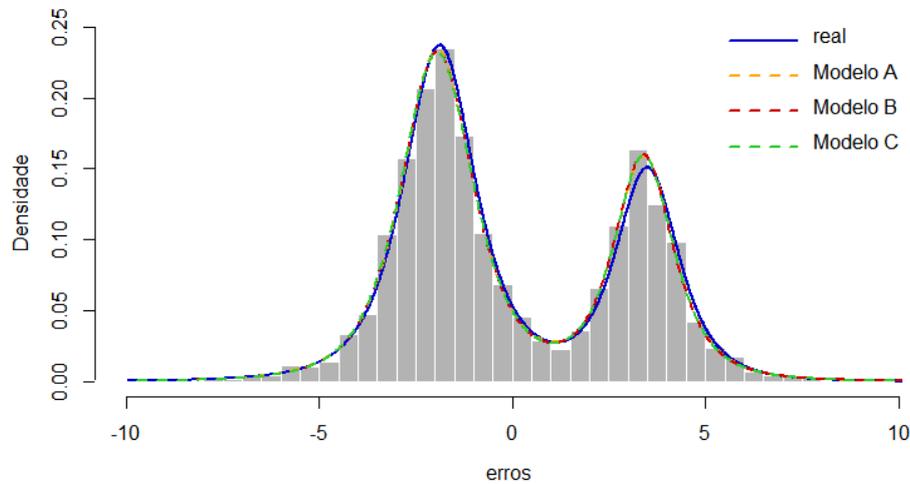


Figura 3.9: Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas para a amostra de tamanho 2500 e $J = 2$.

Na Tabela 3.6 apresentamos os resultados *a posteriori* obtidos para cada um dos parâmetros com seus respectivos intervalos de alta densidade *a posteriori* (HPD) considerando 95% de probabilidade. Observa-se que as médias *a posteriori* para μ , μ^* , σ^2 e β foram similares para os três modelos ajustados, e similares aos valores reais. Os intervalos HPD incluem os valores verdadeiros, exceto o parâmetro β_0 que apresentou média *a posteriori* superior ao valor real nos três modelos. No modelo B as probabilidades *a posteriori* para $\hat{\nu}_j$ na primeira moda são próximas dos valores reais. Já na segunda moda é dado maior peso para $\nu = 2.05$. No modelo C a maior probabilidade *a posteriori* foi dada para $\nu = 3$ na moda 1 e para $\nu = 1.9$ na moda 2. Embora os modelos B e C assumam estruturas diferentes para a cauda, os resultados obtidos são próximos do modelo A que ajustou a estrutura verdadeira. Além disso, os pesos *a posteriori* estimados para cada valor de ν indicam que os modelos foram capaz de identificar o valor de ν mais próximo do real.

Tabela 3.6: Resultados *a posteriori* para o modelo de regressão com distribuição de mistura para os erros considerando $n=2500$ e $J = 2$

Modelos	μ	HPD	μ^*	HPD	σ^2	HPD	β_0	HPD	β	HPD	w	\hat{w}
A	-1.938	[-2.001, -1.872]	-0.819	[-0.917, -0.730]	0.997	[0.868, 1.126]	1.120	[1.033, 1.204]	-2.022	[-2.083, -1.964]	0.638	(0.558, 0.442)
	3.411	[3.330, 3.495]	4.531	[4.431, 4.626]	0.702	[0.585, 0.828]			0.997	[0.884, 1.105]	0.362	(0.727, 0.273)
B	-1.945	[-2.013, -1.877]	-0.822	[-0.917, -0.734]	1.015	[0.862, 1.146]	1.122	[1.033, 1.204]	-2.025	[-2.084, -1.967]	0.636	(0.489, 0.511)
	3.401	[3.315, 3.487]	4.523	[4.428, 4.624]	0.693	[0.573, 0.828]			1.002	[0.894, 1.113]	0.364	(0.710, 0.290)
C	-1.947	[-2.017, -1.883]	-0.822	[-0.920, -0.738]	1.021	[0.897, 1.153]	1.124	[1.039, 1.213]	-2.025	[-2.080, -1.963]	0.636	(0.255, 0.463, 0.282)
	3.401	[3.321, 3.491]	4.525	[4.425, 4.624]	0.705	[0.580, 0.839]			1.003	[0.891, 1.114]	0.364	(0.503, 0.291, 0.206)

ii) Erros com distribuição trimodal

Neste cenário assumimos que os erros do modelo de regressão possuem uma distribuição trimodal. Os erros foram gerados supondo $J = 3$, $K = 2$, $\boldsymbol{\sigma}^2 = (1, 0.81, 0.64)$, $\boldsymbol{\nu} = (2.5, 5)$, $\boldsymbol{w} = (0.5, 0.35, 0.15)$, $\dot{\boldsymbol{w}} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$ e $\boldsymbol{\mu} = (-1.33, 1, 5)$. Consideramos um modelo com 2 variáveis explicativas, sendo $X_1 \sim \mathcal{N}(0, 1)$ e $X_2 \sim \mathcal{U}(0, 1)$, com $\beta_0 = 1$ e $\boldsymbol{\beta} = (-2, 1)$. Como consequência dos valores escolhidos para $\boldsymbol{\mu}$ e β_0 , temos que $\boldsymbol{\mu}^* = (-0.33, 2, 6)$. Lembrando que os valores para \boldsymbol{w} e $\boldsymbol{\mu}$ foram escolhidos de forma que a $\sum_{j=1}^3 w_j \mu_j = 0$.

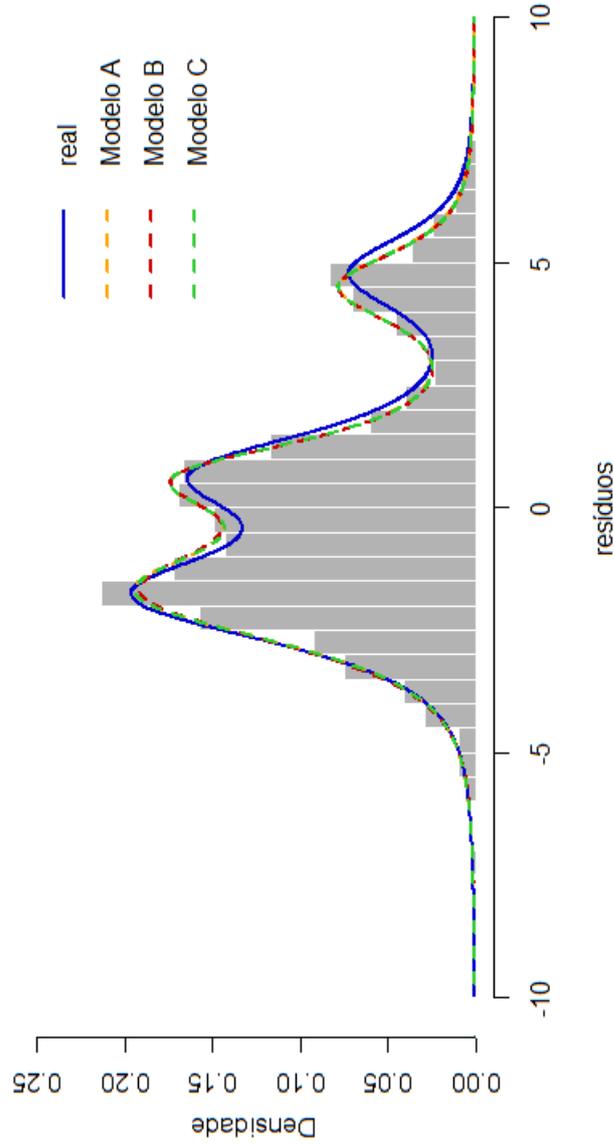
Assim como na configuração anterior, assumimos 3 diferentes modelos para ajuste. Como consideramos na matriz de pesos $\dot{\boldsymbol{w}}$ que cada valor de ν contribui com igual proporção em cada moda, esperamos que os modelos B e C sejam capazes de identificar qual estrutura de cauda é mais representativa em cada moda. Os modelos ajustados foram:

- A. $J = K = 2$ com ν verdadeiro;
- B. $J = K = 2$ com $\nu = (2.05, 6)$;
- C. $J = 2, K = 3$ com $\nu = (1.9, 3, 8)$.

Na Tabela 3.10 apresentamos os resultados *a posteriori* obtidos para cada um dos parâmetros com seus respectivos intervalos HPD considerando 95% de probabilidade. Como os resíduos apresentam duas modas relativamente próximas, isso pode causar certo confundimento na classificação e como consequência nas médias *a posteriori* dos parâmetros, em especial para o parâmetro $\boldsymbol{\mu}$. É possível observar que os intervalos HPD para $\boldsymbol{\mu}$ não incluem os valores reais deste parâmetro. Uma possível explicação é que este parâmetro sofre influência de β_0 , que também apresentou médias *a posteriori* superestimadas e também não incluiu o valor real em seu intervalo HPD. As probabilidades *a posteriori* para o vetor de pesos \boldsymbol{w} foram bem estimadas. Em relação às probabilidades *a posteriori* para $\dot{\boldsymbol{w}}$, nota-se que para os modelos B e C conseguiram capturar a estrutura de cauda, dando maiores pesos aos valores de ν que mais se aproximam dos valores reais. Estes resultados vão na direção esperada, ou seja, em ambos os modelos a estrutura de cauda foi satisfatoriamente capturada. Uma observação importante é que quanto mais refinado for o grid definido para $\boldsymbol{\nu}$, melhores são os resultados obtidos *a posteriori* para a estrutura de cauda. No entanto, em termos de aproximação das curvas, nota-se que parece não haver uma substancial diferença nos ajustes, como pode ser observado na Figura 3.10.

Tabela 3.7: Resultados *a posteriori* para o modelo de regressão com distribuição de mistura para os erros considerando $n=2500$ e $J = 3$

Modelos	μ^\top	HPD	$\mu^{*\top}$	HPD	$\sigma^{2\top}$	HPD	β_0	HPD	β	HPD	w	\hat{w}
A	-1.728	[-1.872, -1.592]	-0.253	[-0.440, -0.061]	1.112	[0.867, 1.358]	1.475	[1.361, 1.595]	-2.016	[-2.081, -1.957]	0.525	(0.421, 0.579)
	0.618	[0.482, 0.759]	2.094	[1.928, 2.287]	0.648	[0.449, 0.852]			0.933	[0.733, 1.173]	0.318	(0.476, 0.524)
	4.516	[4.396, 4.626]	5.992	[5.823, 6.130]	0.593	[0.434, 0.776]					0.157	(0.600, 0.400)
B	-1.715	[-1.862, -1.556]	-0.227	[-0.429, -0.015]	1.194	[0.915, 1.476]	1.487	[1.362, 1.600]	-2.014	[-2.071, -1.950]	0.529	(0.225, 0.775)
	0.633	[0.483, 0.772]	2.121	[1.920, 2.299]	0.636	[0.441, 0.835]			0.920	[0.688, 1.148]	0.314	(0.488, 0.512)
	4.510	[4.399, 4.629]	5.998	[5.843, 6.151]	0.590	[0.409, 0.777]					0.157	(0.530, 0.470)
C	-1.728	[-1.887, -1.590]	-0.252	[-0.449, -0.038]	1.114	[0.878, 1.402]	1.476	[1.355, 1.598]	-2.016	[-2.075, -1.955]	0.525	(0.118, 0.515, 0.367)
	0.621	[0.476, 0.764]	2.097	[1.903, 2.280]	0.645	[0.453, 0.841]			0.934	[0.690, 1.158]	0.318	(0.233, 0.408, 0.359)
	4.514	[4.402, 4.631]	5.990	[5.836, 6.147]	0.585	[0.413, 0.771]					0.157	(0.289, 0.445, 0.266)

Figura 3.10: Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas a amostra de tamanho 2500 e $J = 3$.

A função de autocorrelação e o traço para a cadeia da *log* densidade *a posteriori* no modelo B podem ser vistas no Apêndice D.

3.6 Modelando assimetria e caudas pesadas

Nesta seção apresentamos os resultados do estudo de simulação conduzido com o objetivo de avaliar a eficiência e flexibilidade do modelo proposto em acomodar assimetria e caudas pesadas simultaneamente. Para isso consideramos dois cenários onde a assimetria se faz presente. No primeiro cenário os dados vêm de uma distribuição skew-t. Para o segundo cenário assumimos que os dados seguem uma mistura de distribuições skew-t.

A título de comparação, além do modelo proposto consideramos também uma mistura de normais, mistura de t's de Student estimando o mesmo ν para todos os componentes e mistura de t's de Student estimando ν 's diferentes, na tentativa de aproximar a curva da distribuição que gerou os dados. Os modelos baseados em misturas de distribuições t de Student, onde ν é estimado, foram implementados dentro do MCMC apresentado na Seção 2.3. O parâmetro ν foi amostrado em um passo de Metropolis-Hastings conforme Gonçalves et al. (2015) e a distribuição *a priori* adotada foi a proposta por Martins et al. (2014).

Para comparação dos modelos avaliamos o viés, variância e erro quadrático médio (EQM) dos resultados *a posteriori* referentes a esperança e variância de \mathbf{Y} nas distribuições de mistura utilizadas para ajuste. Como forma de avaliar qual modelo fornece uma melhor aproximação da curva ajustada globalmente e nas caudas da distribuição em relação à curva real (distribuição assumida na geração dos dados), foi definida a seguinte medida percentual de variação entre a densidade real e a densidade estimada em cada modelo:

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i^{real} - \hat{f}_i}{f_i^{real}} \right|$$

O cálculo de \bar{D} nas caudas da distribuição foi baseado no percentis 1% e 99%. Apenas valores abaixo do percentil 1% e acima do percentil 99% foram utilizados para encontrar a medida \bar{D}_{cauda} .

As amostras *a posteriori* da $E(Y)$ e $\text{Var}(Y)$ foram geradas a partir do cálculo destas quantidades para uma distribuição de mistura. Em um modelo de misturas finitas estas quantidades são dadas respectivamente por:

$$E(Y) = \sum_{j=1}^J w_j \lambda_j = \lambda_{mix},$$

$$\text{Var}(Y) = \sum_{j=1}^J w_j [(\lambda_j - \lambda_{mix})^2 + \zeta_j^2],$$

onde λ_j e ζ_j^2 são respectivamente a média e variância da distribuição assumida na componente j da mistura.

Cadeias de tamanho 50000 foram geradas e as primeiras 10000 iterações foram descartadas (*burn-in*). Como forma de considerar o mesmo tamanho para a amostra final *a posteriori*, assumimos o maior *lag* retornado entre todos os modelos ajustados, lembrando que o *lag* é definido a partir do cálculo do tamanho efetivo da amostra.

3.6.1 Dados vindos de uma distribuição skew-t

Para este cenário foram geradas 5000 observações vindas de uma distribuição Skew-t com $\mu = 0$, $\sigma^2 = 1$, $\nu = 2.5$ e $\lambda = 1.5$. Neste estudo avaliamos a flexibilidade de se utilizar um modelo de mistura com componentes simétricos para estimação de densidades que são assimétricas unimodais. Os modelos comparados foram:

- A. Mistura de Normais com $J = 2, 3$;
- B. Mistura de t's $J = 2, 3$ e estimando $\nu_1 = \dots = \nu_J = \nu$;
- C. Mistura de t's $J = 2, 3$ e estimando $\nu_1 \neq \dots \neq \nu_J$;
- D1. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.3, 5)$ fixo;
- D2. Modelo Proposto com $J = 2, 3$, $K = 3$ e $\nu = (2.3, 3, 4)$ fixo;
- D3. Modelo Proposto com $J = 2, 3$, $K = 4$ e $\nu = (2.3, 2.8, 3.5, 5)$ fixo.

A Tabela 3.8 apresenta os resultados obtidos para este cenário. As quantidades calculadas com base nos resultados *a posteriori* para a $E(Y)$ foram similares em todos os modelos, o que nos leva a concluir que em média os modelos comparados não diferem muito. Acreditamos, pelos resultados obtidos, que o impacto maior no viés, variância e EQM seja para os resultados *a posteriori* para a $\text{Var}(Y)$. Observa-se que os modelos D1, D2 e D3 são os que forneceram os menores EQM's. Além disso, estes mesmos modelos são o que possuem as menores distâncias em relação à curva real, tanto globalmente quanto nas caudas da distribuição, sendo o modelo D1 com $J = 2$, o que apresentou a menor distância dentre todos os modelos considerados. É importante ressaltar que

no modelo proposto a escolha da grade para ν é de extrema importância, uma vez que uma má escolha pode influenciar no cálculo da variância, fornecendo valores altos para o EQM. Na prática o ideal é assumir uma grade bem refinada, de maneira que uma variedade de valores de ν possam ser varridas. Os valores que não contribuirão na mistura terão as menores estimativas a posteriori para \hat{w}_j em cada moda j .

Na Figura 3.11 apresentamos os modelos que forneceram as melhores aproximações com base nas distâncias \bar{D} e \bar{D}_{cauda} . É possível observar que os modelos conseguem aproximar de maneira satisfatória a verdadeira distribuição que gerou os dados. Ressalta-se que, quanto maior o tamanho da amostra, melhor a aproximação. No Apêndice C apresentamos outros resultados deste estudo quando $n = 1000$.

Tabela 3.8: viés, variância e erro quadrático médio para a esperança e variância de Y em diferentes distribuições de mistura quando os dados são gerados de uma distribuição skew-t

Modelos	$\widehat{E(Y)}$			$\widehat{Var(Y)}$			\bar{D}	\bar{D}_{cauda}
	vício	var	EQM	vício	var	EQM		
<u>A</u>								
$J = 2$	0.01398	0.00044	0.00064	-1.07750	0.01766	1.17881	0.850	0.924
$J = 3$	0.02009	0.00041	0.00081	-1.00801	0.02442	1.04060	0.760	0.856
<u>B</u>								
$J = 2$	-0.15042	0.00027	0.02290	-2.23640	0.00570	5.0072	0.767	0.866
$J = 3$	-0.00390	0.00028	0.00029	-1.45462	0.01870	2.13472	0.683	0.772
<u>C</u>								
$J = 2$	-0.07566	0.00031	0.00602	-1.75210	0.01481	3.0848	0.740	0.834
$J = 3$	-0.00183	0.00030	0.00030	-1.42133	0.02104	2.04122	0.676	0.765
<u>D1</u>								
$J = K = 2$	-0.08902	0.00023	0.00816	-0.31350	0.05503	0.15332	0.519	0.581
$J = 3, K = 2$	-0.01314	0.00025	0.00043	-0.60871	0.13395	0.50448	0.487	0.544
<u>D2</u>								
$J = 2, K = 3$	-0.10052	0.00023	0.01033	-0.56898	0.07150	0.39524	0.572	0.641
$J = K = 3$	-0.02466	0.00024	0.00085	-0.83921	0.09876	0.80304	0.509	0.569
<u>D3</u>								
$J = 2, K = 4$	-0.07713	0.00025	0.00621	-0.42865	0.09806	0.28181	0.578	0.648
$J = 3, K = 4$	-0.01709	0.00024	0.00053	-0.90111	0.08782	0.89982	0.521	0.582

$E(Y) = 1.00348$ (média amostral: 1.02860); $var(Y) = 3.99303$ (variância amostral: 3.36196)

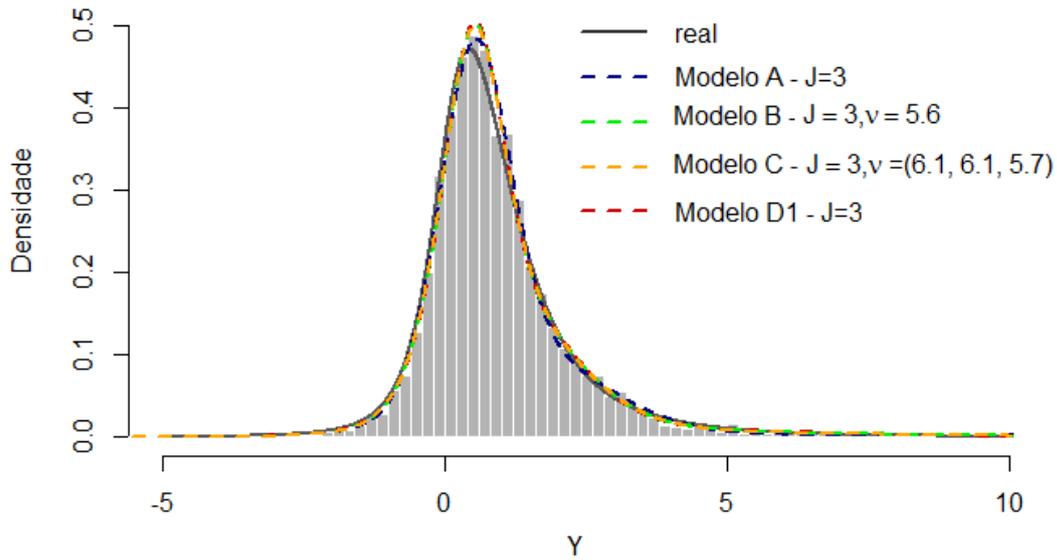


Figura 3.11: Histograma para 5000 valores gerados a partir de uma Skew-t e seus respectivos ajustes em comparação à densidade real.

3.6.2 Dados vindos de uma mistura de distribuições skew-t

Neste cenário foram geradas 5000 observações vindas de uma mistura de distribuições Skew-t com $\mu = (-1, 1)$, $\sigma^2 = (1, 0.81)$, $\nu = (2.8, 4)$, $\lambda = (-1.5, 0.8)$ e $w = (0.65, 0.35)$. Ao contrário do estudo anterior onde a distribuição geradora dos dados era unimodal, neste a distribuição é bimodal e estamos interessados em também avaliar a aproximação da densidade real baseando-se no cálculo das medidas já citadas, mas também avaliando quantos componentes são necessários para obter uma boa aproximação. Os modelos comparados neste estudo foram:

- A. Mistura de Normais;
- B. Mistura de t's estimando $\nu_1 = \dots = \nu_J = \nu$;
- C. Mistura de t's estimando $\nu_1 \neq \dots \neq \nu_J$;
- D1. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.8, 4)$;
- D2. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.3, 5)$;
- D3. Modelo Proposto com $J = 2, 3$, $K = 3$ e $\nu = (2.3, 3.5, 4.5)$.

De maneira análoga aos resultados obtidos no cenário anterior, as quantidades calculadas com base nos resultados *a posteriori* para a $E(Y)$ apresentaram resultados similares, exceto para a mistura de Normais que forneceu o menor EQM em ambos os ajustes. Para resultados *a posteriori* da $Var(Y)$, o modelo D3 foi o que obteve os menores EQM's dentre todos os modelos considerados. Ao avaliar a distância \bar{D} globalmente e nas caudas, nota-se que os modelos D2 e D3 apresentaram as menores distâncias, no entanto, o modelo D3 apresentou o menor EQM dentre todos os modelos considerados para ajuste. Desta maneira, o modelo D2 foi o que forneceu um melhor ajuste, além das menores distâncias (global e nas caudas).

Tabela 3.9: viés, variância e erro quadrático médio para a esperança e variância de Y em diferentes distribuições de mistura quando os dados são gerados de uma mistura de skew-t.

Modelos	$\widehat{E(Y)}$			$\widehat{Var(Y)}$			\bar{D}	\bar{D}_{cauda}
	vício	var (10^{-2})	EQM (10^{-2})	vício	var (10^{-2})	EQM (10^{-2})		
<u>A</u>								
$J = 2$	0.047	0.070	0.290	-0.657	0.660	43.880	0.550	0.601
$J = 5$	0.028	0.030	0.110	-0.282	2.520	10.450	0.415	0.581
<u>B</u>								
$J = 2$	0.104	0.030	0.110	0.227	24.090	29.270	0.275	0.349
$J = 3$	0.065	0.030	0.460	-0.550	2.900	33.170	0.311	0.433
<u>C</u>								
$J = 2$	0.098	0.030	0.990	0.276	31.340	38.960	0.259	0.325
$J = 3$	0.057	0.040	0.350	-0.661	1.740	45.470	0.359	0.501
<u>D1</u>								
$J = K = 2$	0.095	0.030	0.940	-0.474	1.070	23.490	0.319	0.411
$J = 3, K = 2$	0.065	0.030	0.450	-0.476	2.060	24.750	0.268	0.368
<u>D2</u>								
$J = K = 2$	0.089	0.030	0.830	1.343	4.800	185.230	0.154	0.167
$J = 3, K = 2$	0.062	0.020	0.410	0.555	12.230	43.070	1.619	1.864
<u>D3</u>								
$J = 2, K = 3$	0.094	0.030	0.920	0.110	2.190	3.410	0.157	0.176
$J = K = 3$	0.065	0.030	0.450	-0.233	4.840	10.270	0.219	0.297

$E(Y) = -0.718$ (média amostral: -0.706); $var(Y) = 4.947$ (variância amostral: 4.884)

A Figura 3.12 apresenta o histograma com as curvas ajustadas tomando como base os modelos que forneceram menor EQM e distância \bar{D} . Observamos que a curva referente ao ajuste da mistura de normais obteve uma melhor aproximação em comparação aos demais modelos, no entanto vale ressaltar que foram necessários cinco componentes para conseguir uma boa aproximação da curva.

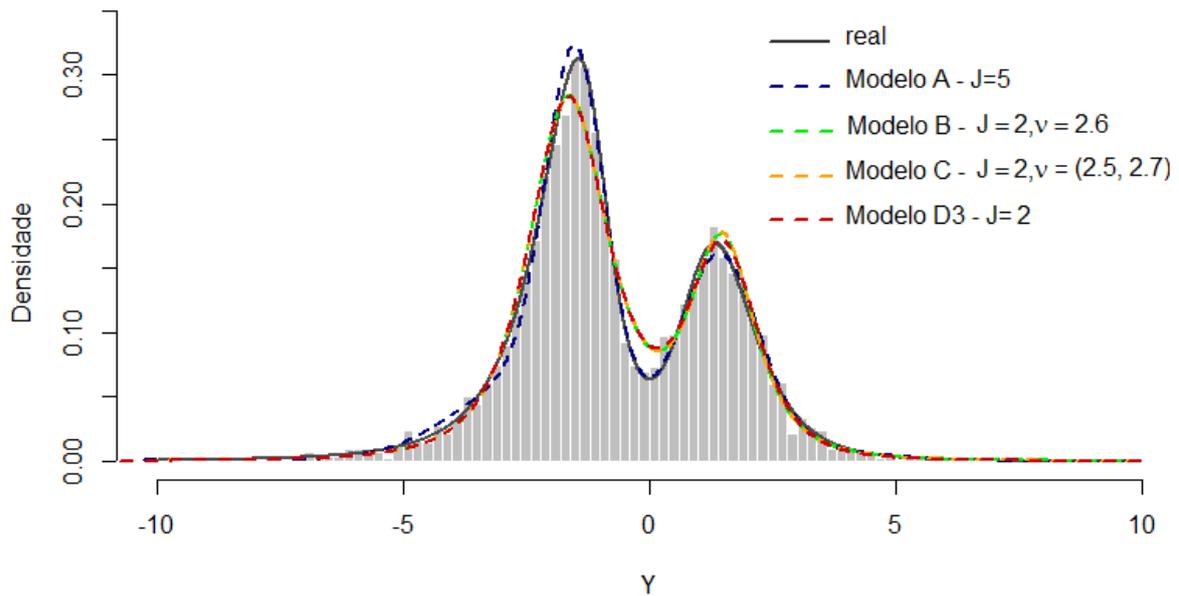


Figura 3.12: Histograma para 5000 valores gerados a partir de uma mistura de Skew-t e seus respectivos ajustes em comparação à densidade real (azul)

A partir dos cinco estudos de simulação discutidos neste capítulo é possível observar a flexibilidade da modelagem proposta em acomodar simultaneamente dados com multimodalidade, assimetria e caudas pesadas. Para o último estudo de simulação apresentado nota-se que a flexibilidade apresentada pela abordagem proposta é similar à mistura de distribuições t de Student quando diferentes ν 's são estimados para cada componente do modelo, mas com a vantagem de não precisar estimar o parâmetro de grau de liberdade. Ao varrer, com certa liberdade, o parâmetro ν em uma grade de valores, estimando a estrutura de cauda através dos pesos $\hat{\boldsymbol{w}}$, evitamos problemas de estimação com este parâmetro. Além disso, existe ainda a versatilidade de poder assumir mais de um ν para modelar a estrutura de cauda em cada moda, o que é ainda mais plausível quando há, por exemplo, assimetria nos dados.

É importante destacar que diferentes escolhas para ν foram consideradas e em todas elas os resultados obtidos foram muito similares entre os modelos, que não sofreram impacto da grade assumida. No entanto, ressalta-se que em situações práticas o pesquisador não precisará fazer escolhas aleatórias para a grade. Ao fixar o número de componentes da mistura em K (não havendo, segundo resultados obtidos nas simulações, a necessidade de assumir $K > 5$), recomendamos utilizar a estratégia mencionada no início deste capítulo, que utiliza a divergência de Kullback-Liebler para definição de uma grade adequada.

Capítulo 4

Aplicações a dados reais

Neste capítulo quatro conjuntos de dados são explorados. Para os dois primeiros conjuntos de dados será utilizada a metodologia proposta na Seção 2.2, para dados independentes e identicamente distribuídos, enquanto que os demais serão aplicados no contexto de regressão. Nas duas primeiras aplicações a metodologia proposta será comparada com os modelos de mistura finita assumindo componentes normais e t de Student existentes na literatura. Nas aplicações em regressão a extensão proposta será comparada com os modelos propostos por [Bartolucci e Scaccia \(2005\)](#) e [Galimberti e Soffritti \(2014\)](#), que assumem, respectivamente, misturas de distribuições normais univariadas e misturas de distribuições t de Student para modelar os erros. A classe de modelos apresentada por [Benites et al. \(2016\)](#), que se baseia na família de mistura de escala da skew-normal também será utilizada para comparação.

Para obter um melhor desempenho computacional optamos por amostrar o parâmetro $\boldsymbol{\mu}$ a partir da distribuição normal truncada, onde cada μ_j foi truncado em intervalos complementares que compreendiam o intervalo de variação dos dados. Os submodelos baseados em misturas de distribuições t de Student, onde o parâmetro $\boldsymbol{\nu}$ é estimado, foram implementados dentro do MCMC apresentado na Seção 2.3. O parâmetro $\boldsymbol{\nu}$ foi amostrado em um passo de Metropolis-Hastings. Em todas as aplicações assumimos diferentes valores para o número de componentes J em cada um dos modelos considerados para ajuste. A escolha de J dentro de cada modelo foi baseada no DIC ([SPIEGELHALTER et al., 2002](#)), que também foi utilizado para escolher o modelo com melhor ajuste aos dados. Para o modelo proposto foi fixado $K = 5$ em todas as aplicações e a grade escolhida foi $\boldsymbol{\nu} = (2.1, 2.3, 2.6, 3.2, 10)$, definida com base na estratégia que utiliza a divergência de Kullback-Liebler apresentada no Capítulo 3.

4.1 Comprimento de peixes

Conjunto de dados analisado em [Titterington et al. \(1985\)](#), referente ao comprimento de 256 peixes. Os dados, disponíveis no pacote `bayesmix` ([GRÜN, 2011](#)), possuem uma distribuição empírica que sugere a existência de pelo menos quatro modas. Como mencionado na seção 1.1, uma possível explicação para a multimodalidade é que os peixes pertencem a diferentes grupos etários. Como a idade do peixe é uma variável difícil de ser medida, surge aí uma heterogeneidade não-observada que é acomodada por uma mistura.

A análise descritiva dos dados apontou que o comprimento médio dos peixes é de 6.104 unidades de medida (u.m), com variância amostral igual a 3.614. A mediana dos dados é 5.625, sendo os comprimentos mínimo e máximo respectivamente iguais 2.875 a 12.625. Vale ressaltar que apenas 2.5% dos peixes possuem comprimento superior a 10.375. Além de várias modas, os dados também apresentam uma leve assimetria na cauda da direita e algumas observações atípicas, como mostra a Figura 4.1.

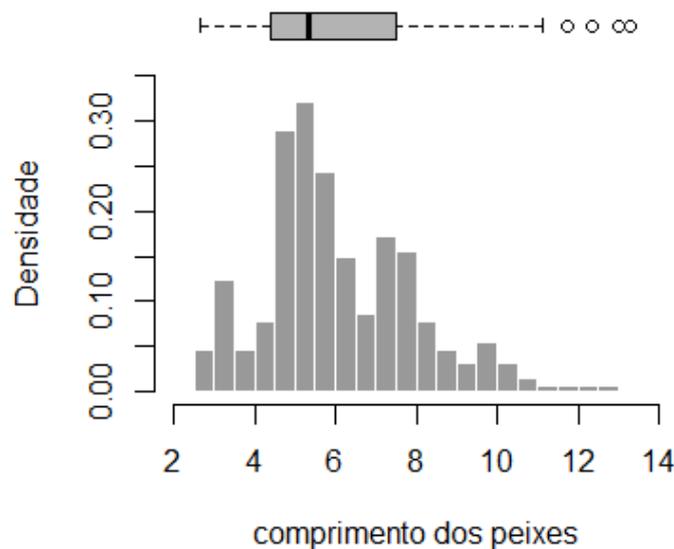


Figura 4.1: Histograma e boxplot para o comprimento de 256 peixes

Os seguintes modelos foram considerados para ajuste:

- A. Modelo Proposto com $J = 3, 4$, $K = 5$ e $\nu = (2.1, 2.3, 2.6, 3.2, 10)$ fixo;
- B. Mistura de t's com $J = 3, 4$ e estimando o mesmo $\nu \forall j$;
- C. Mistura de t's com $J = 3, 4$ e estimando ν 's diferentes $\forall j$;
- D. Mistura de Normais com $J = 3, 4$.

A Tabela 4.1 traz os resultados do DIC para os quatro modelos assumindo $J = 3$ e 4 no ajuste. Os modelos com $J = 3$ forneceram os menores DIC's, sugerindo que para estes dados é suficiente assumir um modelo de mistura contendo apenas três componentes. Os resultados obtidos para cada modelo considerando $J = 3$ são apresentados na Tabela 4.2.

Tabela 4.1: Critério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D referentes ao comprimento de 256 peixes.

Modelo	DIC	
	J=3	J=4
A	1056.9	1071.0
B	1086.7	1109.1
C	1060.2	1084.1
D	1107.4	1125.1

A Figura 4.2 traz as curvas ajustadas em cada modelo. Observa-se que o modelos A, B e C forneceram ajustes similares para as densidades. A mistura de normais preciseria, em contrapartida, de um número maior de componentes de forma a obter uma aproximação similar aos demais modelos. Ao comparar os quatro ajustes nas caudas da distribuição observamos que o modelo A, seguido do modelo C, são os que apresentam caudas mais pesadas em comparação aos modelos B e D (Figura 4.3).

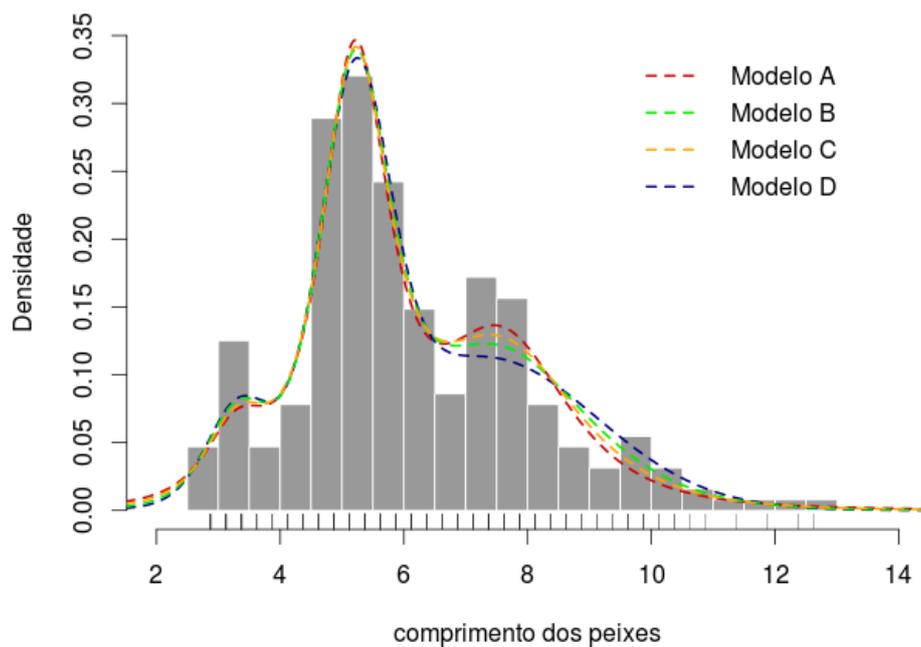


Figura 4.2: Histograma para a distribuição empírica do comprimento dos peixes com as curvas ajustadas em cada modelo.

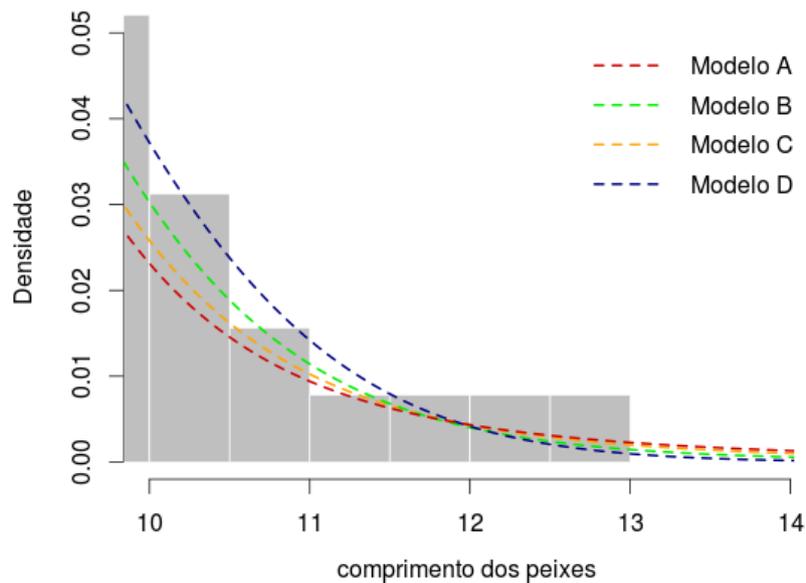


Figura 4.3: Comportamento de cauda para o comprimento dos peixes com as curvas ajustadas em cada modelo.

A partir dos resultados obtidos, observa-se que os quatro modelos apresentaram médias *a posteriori* para μ bem similares. Em relação às médias *a posteriori* para σ^2 , nota-se que há uma diferença nas estimativas referentes à terceira componente da mistura, sendo que os modelos B e D foram os que forneceram as maiores médias para esta componente. O modelo B apresentou estimativa para a média *a posteriori* de ν em torno de 15, enquanto que para o modelo C as estimativas *a posteriori* obtidas para este parâmetro ficaram entre 3.7 e 3.9 em cada componente. No modelo A nota-se que a estrutura de cauda estimada a partir do peso \hat{w} indica que $\nu = 2.1$ contribui com aproximadamente 17% da mistura em cada moda j . Já $\nu = 10$ contribui com 23% na primeira moda e respectivamente 23% e 21% na segunda e terceira modas. Estes resultados sugerem que assumir apenas um ν em cada moda j , ou ainda o mesmo ν para todo j , pode não ser suficiente para estimar a estrutura de cauda de maneira satisfatória. As probabilidades *a posteriori* para w nos modelos A, B e C indicam que a segunda moda possui o maior peso dentro destes modelos. Este resultado sugere que há uma concentração maior de peixes dentro desta componente, cuja média *a posteriori* para μ foi estimada em 5.2, resultado próximo ao comprimento mediano dos peixes que é 5.625. O modelo A foi o que apresentou a maior variância *a posteriori* para Y , no entanto, ele obteve o menor DIC dentre os quatro modelos comparados, o que sugere que este modelo forneceu um melhor ajuste dentre todos.

Tabela 4.2: Resultados *a posteriori* e intervalo HPD de 95% para o comprimento dos peixes.

Modelo	μ	HPD	σ^2	HPD	ν	HPD	w	\hat{w}	$E(\hat{Y})$	$Var(\hat{Y})$
A	3.346	[3.031, 3.759]	0.346	[0.119, 0.687]	-	-	0.091	(0.179, 0.185, 0.194, 0.201, 0.241)	6.023	7.468
	5.211	[5.065, 5.383]	0.328	[0.162, 0.503]	-	-	0.488	(0.173, 0.179, 0.202, 0.213, 0.233)		
	7.551	[7.130, 8.003]	1.503	[0.769, 2.308]	-	-	0.421	(0.173, 0.189, 0.208, 0.218, 0.212)		
B	3.349	[3.065, 3.748]	0.340	[0.121, 0.675]	15.190	[2.056, 41.681]	0.097	-	6.064	3.834
	5.208	[5.039, 5.366]	0.322	[0.175, 0.535]			0.432	-		
	7.429	[6.852, 7.966]	2.343	[1.004, 3.610]			0.471	-		
C	3.364	[3.023, 3.867]	0.382	[0.094, 0.840]	3.713	[2.010, 6.183]	0.092	-	6.034	5.132
	5.210	[5.052, 5.366]	0.340	[0.159, 0.549]	3.990	[2.017, 6.753]	0.470	-		
	7.522	[6.965, 7.928]	1.687	[0.855, 2.603]	3.988	[2.011, 6.649]	0.438	-		
D	3.382	[3.071, 3.869]	0.365	[0.121, 0.773]	-	-	0.100	-	6.102	3.676
	5.223	[5.050, 5.411]	0.333	[0.159, 0.516]	-	-	0.391	-		
	7.330	[5.050, 5.411]	3.255	[2.133, 4.284]	-	-	0.509	-		

4.2 Velocidade das galáxias

Conjunto de dados referente à velocidade de 82 galáxias (em milhares de quilômetros por segundo) situadas na constelação da Coroa Boreal. Na análise destes dados o número de componentes é em geral de interesse e pode ser interpretado por teorias astronômicas. Alguns autores, como Carlin e Chib (1995), Richardson e Green (1997), Stephens (2000b) e Lee (2010) assumiram uma mistura de distribuições normais com $J = 6$ para modelar estes dados. Stephens (2000b) e Lee (2010) também fizeram uma análise baseada em uma mistura de três distribuições t de Student, assumindo $\nu = 4$ para todas as componentes da mistura. Os dados estão disponíveis no pacote MASS (RIPLEY et al., 2013).

A análise inicial indicou que as galáxias possuem uma velocidade média de 20.83 milhares de quilômetros por segundo, com mesma variância amostral. Pela Figura 4.4 nota-se que algumas observações foram apontadas como *outliers* e que os dados possuem pelo menos quatro modas, sendo duas delas bem separadas das demais.

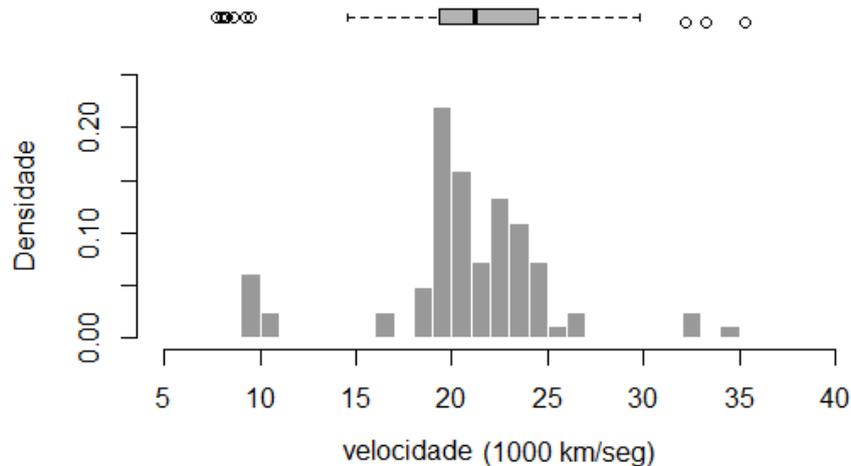


Figura 4.4: Histograma e boxplot para a velocidade de 86 galáxias.

A título de comparação, os seguintes modelos foram considerados para ajuste:

- A. Modelo Proposto com $J = 3, 4$, $K = 5$ e $\nu = (2.1, 2.3, 2.6, 3.2, 10)$ fixo;
- B. Mistura de t's com $J = 3, 4$ e estimando o mesmo $\nu \forall j$;
- C. Mistura de t's com $J = 3, 4$ e estimando ν 's diferentes $\forall j$;
- D. Mistura de Normais com $J = 3, 4$.

A Tabela 4.3 traz os resultados do DIC para os quatro modelos assumindo $J = 3$ e 4 no ajuste. O DIC sugere que $J = 4$ oferece um melhor ajuste em todos os quatro modelos considerados. Os resultados obtidos para cada modelo considerando o DIC são apresentados na Tabela 4.4.

Tabela 4.3: Critério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D referentes a velocidade das galáxias.

Modelo	DIC	
	J=3	J=4
<i>A</i>	437.9	425.4
<i>B</i>	437.9	434.8
<i>C</i>	471.3	430.0
<i>D</i>	466.8	455.0

A Figura 4.5 traz as curvas ajustadas para cada modelo. Observa-se que A, B e C possuem ajustes similares e capturaram de maneira satisfatória as quatro modas. O modelo D, no entanto, forneceu uma aproximação inferior aos demais modelos, sugerindo que provavelmente mais componentes seriam necessários na mistura de maneira a conseguir uma boa aproximação dos dados.

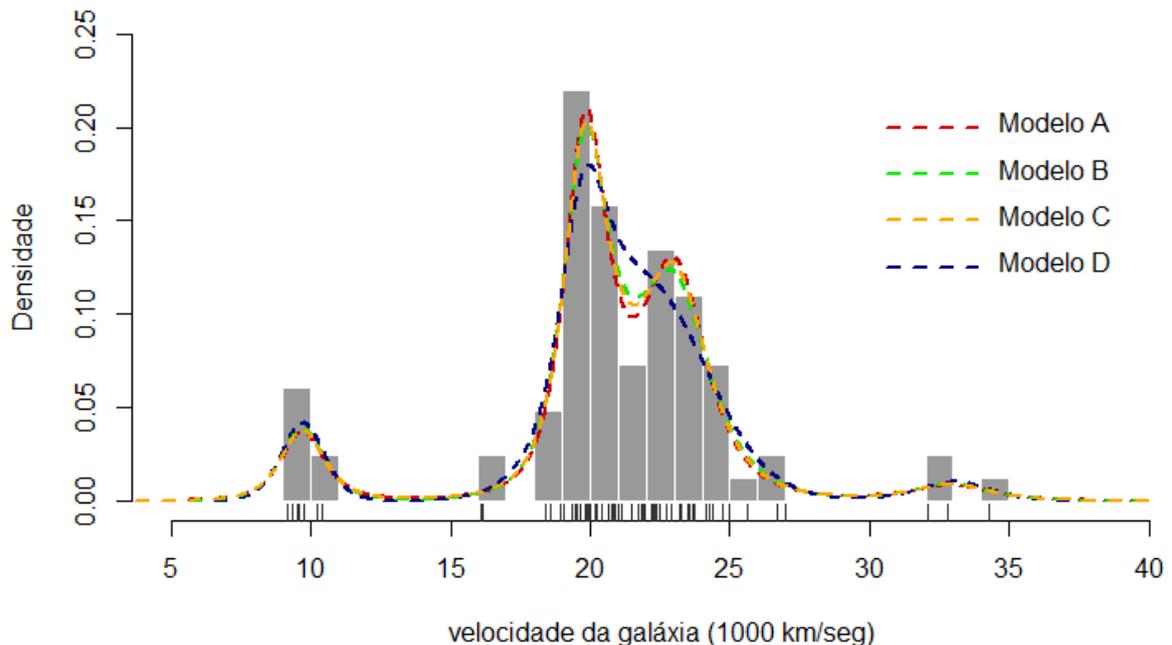


Figura 4.5: Histograma para a distribuição empírica da velocidade de 82 galáxias com as curvas ajustadas em cada modelo

Os resultados *a posteriori* para $\boldsymbol{\mu}$ foram similares em todos os modelos. O modelo B apresentou estimativa para a média *a posteriori* de ν em torno de 6.4, enquanto que para o modelo C as estimativas *a posteriori* obtidas para este parâmetro ficaram entre 3.8 e 4.6. No modelo A as probabilidades *a posteriori* para o peso \boldsymbol{w} sugerem que $\nu = 2.1$ contribui com 18% na primeira componente do modelo, enquanto que $\nu = 10$ contribui com 23%. Na segunda e terceira componentes da mistura em j , $\nu = 10$ contribui com 19% da mistura. Estes resultados sugerem que para estimar a estrutura de cauda de maneira satisfatória em cada moda j , é importante considerar pelo menos dois ν 's diferentes. Nos quatro modelos ajustados a segunda e terceira modas são as que possuem as maiores probabilidades *a posteriori* para o peso \boldsymbol{w} . Observamos que a variância *a posteriori* de \mathbf{Y} no modelo D foi a menor em comparação aos demais modelos, no entanto o DIC sugere que o modelo A forneceu um melhor ajuste aos dados (DIC=425.4).

Tabela 4.4: resultados *a posteriori* e intervalo HPD de 95% para a velocidade das galáxias

Modelo	μ	HPD	σ^2	HPD	ν	HPD	w	\hat{w}	$E(\hat{Y})$	$V\hat{ar}(\hat{Y})$
A	9.714	[9.114, 10.431]	0.769	[0.194, 1.653]	-	-	0.085	(0.184, 0.189, 0.192, 0.201, 0.234)		
	19.889	[19.507, 20.343]	0.709	[0.241, 1.438]	-	-	0.440	(0.202, 0.205, 0.205, 0.195, 0.193)		
	22.958	[22.261, 23.765]	1.668	[0.468, 3.227]	-	-	0.439	(0.195, 0.200, 0.202, 0.210, 0.193)	20.810	27.016
	32.814	[30.185, 35.624]	2.724	[0.246, 9.282]	-	-	0.036	(0.195, 0.199, 0.201, 0.201, 0.204)		
B	9.701	[9.005, 10.350]	0.724	[0.187, 1.546]			0.086	-		
	19.883	[19.429, 20.461]	0.830	[0.181, 2.039]	6.356	[2.011, 16.240]	0.411	-	20.822	23.864
	22.789	[21.688, 23.821]	2.456	[0.515, 5.066]			0.466	-		
	32.813	[30.326, 35.501]	7.057	[0.252, 7.918]			0.037	-		
C	9.722	[9.162, 10.462]	0.734	[0.146, 1.709]	3.882	[2.014, 7.008]	0.086	-		
	19.901	[19.506, 20.483]	0.798	[0.197, 1.764]	3.896	[2.014, 7.475]	0.433	-	20.815	24.917
	22.892	[22.023, 23.837]	1.947	[0.470, 4.041]	3.930	[2.011, 7.565]	0.445	-		
	32.759	[29.487, 35.194]	2.713	[0.291, 9.458]	4.644	[2.012, 10.537]	0.036	-		
D	9.715	[9.146, 10.318]	0.677	[0.186, 1.348]			0.086	-		
	19.920	[19.299, 20.996]	1.282	[0.142, 3.880]			0.301	-	20.820	20.885
	22.265	[21.008, 23.838]	4.670	[1.426, 7.372]			0.575	-		
	32.928	[30.823, 34.959]	2.784	[0.341, 8.585]			0.038	-		

4.3 Concentração de hemoglobina em atletas

Conjunto de dados de um estudo biomédico realizado pelo Instituto Australiano de esportes (AIS). Os dados estão disponíveis no pacote DAAG (MAINDONALD et al., 2015) e contêm resultados de hemograma, além de algumas medidas antropométricas de 202 atletas. Weisberg (2005) analisou estes dados no contexto de modelos de regressão linear, enquanto que Galimberti e Soffritti (2014) assumiram uma mistura finita de distribuições t de Student multivariadas para modelar os erros no modelo de regressão.

Uma análise inicial mostrou que 50.5% dos atletas são do sexo masculino. A concentração média de hemoglobina é de 14.57g/dl. Entre os homens esta concentração é em média 15.55 g/dl, enquanto que para as mulheres é em média 13.56 g/dl. Em relação à contagem de células brancas, os atletas possuem em média 7.11×10^{12} células por litro. A Figura 4.6(a) representa a distribuição para a concentração de hemoglobina nos atletas. A partir dela observamos que uma assimetria à esquerda e uma possível bimodalidade. Um modelo de regressão linear foi ajustado aos dados considerando a concentração de hemoglobina como variável resposta e a contagem e célula brancas como variável explicativa. Na Figura 4.6(b) apresentamos os resíduos deste ajuste. Observa-se que os resíduos possuem um comportamento bimodal e isto pode ser decorrente do efeito de algum regressor relevante que não foi considerado no modelo. Neste caso o modelo usual de regressão pode não ser suficiente para explicar a heterogeneidade apresentada.

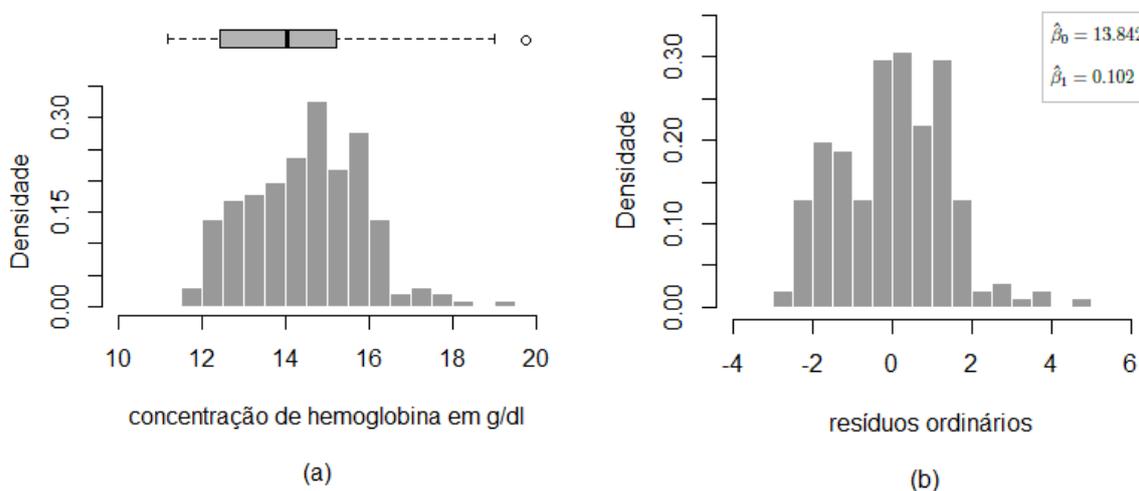


Figura 4.6: (a) Histograma e boxplot para a concentração de hemoglobina g/dl e (b) Histograma dos resíduos ordinários para o modelo de regressão linear.

Para esta aplicação assumimos que os erros do modelo de regressão linear são modelados a partir da mistura proposta na Seção 2.2. Consideramos novamente a concentração de hemoglobina em gramas por decalitro (g/dl) como variável resposta e a contagem de células brancas em 10^{12} por litro como variável explicativa. A título de

comparação foram assumidos os seguintes modelos para os erros:

- A. Modelo Proposto com $J = 2, 3$, $K = 5$ com $\nu = (2.1, 2.3, 2.6, 3.2, 10)$ fixo;
- B. Mistura de t's com $J = 2, 3$ e estimando o mesmo $\nu \forall j$;
- C. Mistura de t's com $J = 2, 3$ e estimando ν 's diferentes $\forall j$;
- D. Mistura de Normais com $J = 2, 3$;
- E. Mistura de Skew-t's com $J = 2, 3$.

A Tabela 4.5 traz os resultados do DIC para os modelos A, B, C e D e do BIC para o modelo E, assumindo $J = 2$ e 3 no ajuste. Para os modelos A, B e C o DIC sugere que $J = 3$ oferece um melhor ajuste, enquanto que para o modelos D e E o melhor ajuste seria considerando $J = 2$ componentes. Os resultados obtidos para cada modelo considerando os resultados do DIC e BIC são apresentados na Tabela 4.6. É importante ressaltar que o modelo E foi ajustado através do pacote `FMsmnReg` que realiza inferência por máxima verossimilhança, via algoritmo EM, e considera que a distribuição dos erros pertence à classe SMSN (BENITES et al., 2016).

Tabela 4.5: Critério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D e critério BIC para diferentes valores de J no modelo E

Modelo	DIC	
	J=2	J=3
<i>A</i>	799.5	674.3
<i>B</i>	809.0	787.6
<i>C</i>	798.1	688.5
<i>D</i>	845.3	875.4
<i>E</i>	BIC	
	721.895	741.102

A Figura 4.7 apresenta os resíduos do modelo com as curvas ajustadas para os modelos A, B, C e D. Observa-se que os modelos B e C forneceram ajustes similares para as curvas e dos quatro ajustes, o modelo D foi o que apresentou pior desempenho. Ao comparar os modelos nas caudas da distribuição, nota-se que os modelos A e C apresentam caudas mais pesadas em relação aos demais modelos (Figura 4.8). Não foi possível comparar o ajuste da curva para o modelo E que ajusta uma mistura com duas componentes skew-t.

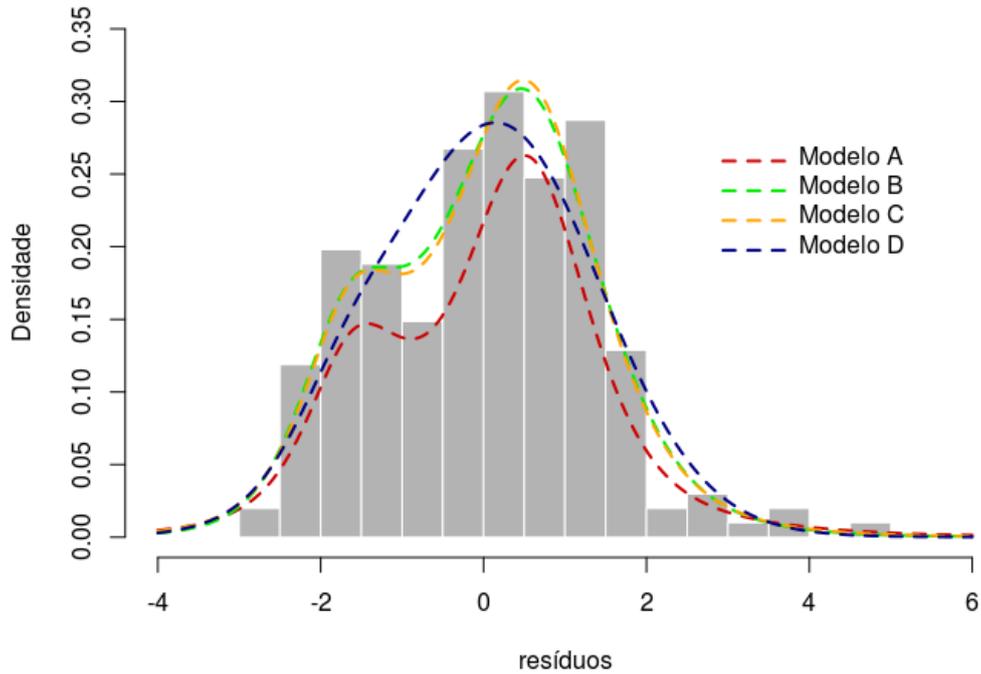


Figura 4.7: Resíduos do modelo de regressão com respectivas curvas ajustadas para os dados AIS.

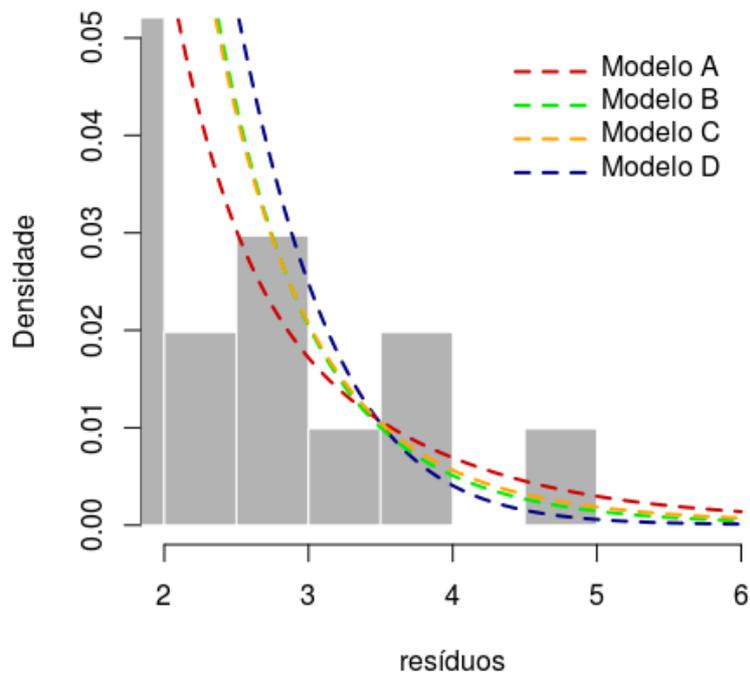


Figura 4.8: Resíduos do modelo nas caudas da distribuição com respectivas curvas ajustadas para os dados AIS.

Os resultados para a média *a posteriori* de $\boldsymbol{\mu}$ foram bem diferentes entre os modelos A, B e C, principalmente na segunda e terceira modas. Nos modelos D e E as médias *a posteriori* para $\boldsymbol{\mu}$ são similares apenas na primeira componente. Em relação à média *a posteriori* de $\boldsymbol{\sigma}^2$ observa-se que os modelos C e D apresentaram as maiores médias em comparação aos demais modelos ajustados. Observa-se, ainda, que o intervalo HPD para β_1 inclui o zero em todos os modelos, sugerindo que o efeito de β_1 não foi estatisticamente significativo. Para os modelos B e E o valor de ν estimado foi aproximadamente 13.8 e 12.3, respectivamente, o que sugere uma estrutura de cauda mais leve. No modelo C cada componente da mistura apresentou média *a posteriori* para este parâmetro em torno de 6.5. No modelo A a estrutura de cauda foi estimada a partir das probabilidades *a posteriori* para o peso \boldsymbol{w} . Os resultados obtidos atribuem as maiores probabilidades *a posteriori* para $\nu = 10$ em todas as componentes, no entanto vale ressaltar que as probabilidades *a posteriori* para valores de ν inferiores a 2.6 sugerem que é importante considerar graus de liberdade pequenos na estimação da estrutura de cauda. Dos cinco modelos ajustados o que possuiu um melhor ajuste segundo os critérios DIC e BIC apresentados na Tabela 4.5 foi o modelo A.

Tabela 4.6: Resultados *a posteriori** para o modelo de regressão com distribuição de mistura nos erros para os dados do estudo AIS.

Modelo	μ	HPD	σ^2	HPD	β_0	HPD	β_1	HPD	$\hat{\nu}$	HPD	w	ψ
A	-1.522	[-1.923, -1.036]	0.427	[0.177, 0.792]	14.228	[13.492, 14.947]	0.045	[-0.059, 0.143]	-	-	0.276	(0.155, 0.160, 0.180, 0.201, 0.304)
	0.529	[0.293, 0.785]	0.678	[0.403, 0.998]					-	-	0.710	(0.116, 0.137, 0.162, 0.201, 0.384)
	3.261	[2.416, 4.300]	3.213	[0.154, 7.597]					-	-	0.014	(0.192, 0.195, 0.197, 0.205, 0.211)
B	-2.698	[-4.747, -0.003]	0.845	[0.139, 1.559]	14.075	[13.433, 14.826]	0.067	[-0.029, 0.164]	13.828	[2.452, 33.348]	0.243	-
	0.118	[-1.594, 0.786]	0.965	[0.248, 1.995]							0.467	-
	2.761	[0.088, 7.623]	2.122	[0.203, 3.946]							0.290	-
C	-2.846	[-8.201, -0.114]	1.645	[0.151, 1.528]	14.179	[13.432, 14.976]	0.051	[-0.057, 0.159]	6.579	[2.036, 13.056]	0.228	-
	0.049	[-1.603, 0.764]	0.880	[0.220, 1.757]					6.665	[2.025, 14.188]	0.451	-
	4.421	[0.014, 14.871]	7.484	[0.257, 4.247]					6.303	[2.010, 13.601]	0.321	-
D	-1.965	[-2.073, 0.000]	1.640	[0.187, 2.279]	13.857	[13.083, 14.552]	0.099	[-0.003, 0.197]	-	-	0.457	-
	1.853	[0.000, 4.603]	1.832	[0.246, 3.353]							0.543	-
E	-1.835	-	0.252	-	14.042	-	0.102	-	12.283	-	0.268	-
	0.671	-	1.492	-							0.732	-

* exceto para o modelo E.

4.4 Pesquisa nacional de exames de saúde e nutrição - EUA

Conjunto de dados da pesquisa nacional de exames de saúde e nutrição - NHANES conduzida anualmente desde 1999 pelo Centro Nacional de Estatísticas de Saúde (NCHS) dos EUA. Os dados, disponíveis no pacote NHANES (PRUIM, 2015), referem-se às pesquisas conduzidas nos anos de 2009/2010 e 2011/2012 e contêm informações de 76 variáveis que descrevem características demográficas, físicas, de saúde e estilo de vida de 10000 participantes (5000 para cada ano).

Lin et al. (2007a) e Cabral et al. (2008) analisaram anteriormente os dados desta pesquisa referentes aos anos de 1999/2000 e 2001/2002, restringindo a amostra apenas para participantes do sexo masculino. Além disso, participantes com pesos variando entre 70.1–95 kg foram retirados da análise como forma de induzir uma heterogeneidade decorrente de dois grupos intrínsecos de pesos corporais. Lin et al. (2007a) assumiram uma mistura de distribuições skew-t para estimar a densidade referente ao índice de massa corpórea dos participantes, enquanto que Cabral et al. (2008) utilizaram uma mistura de skew-t-normal para estimação da densidade.

Para esta aplicação serão exploradas as informações dos participantes envolvidos na pesquisa realizada entre 2011/2012. Como a base possui muitas variáveis, restringimos a análise apenas para as variáveis peso (medida em kilogramas), idade, sexo e diabetes (0-Não; 1-Sim). Os participantes que não possuíam informação de pelo menos uma das variáveis consideradas foram previamente removidos da base de dados. Com isso a amostra final considerada contém 4905 participantes.

Uma análise inicial dos dados mostrou que 50.2% dos participantes são do sexo masculino e que apenas 7.5% dos participantes são diabéticos. Em média as mulheres pesam 65.3 kg, enquanto que os homens possuem um peso médio de 75.3 kg. Além disso, a idade média dos participantes é de 37 anos. A partir do boxplot apresentado na Figura 4.9(a) é possível observar que o peso é relativamente maior entre os diabéticos, em ambos os sexos, e que há uma relativa diferença no peso para diferentes faixas de idade, como mostra a Figura 4.9(b). Uma hipótese a ser levantada é que o sexo, a informação sobre diabetes ou a idade do participante podem ser uma fonte de heterogeneidade que explicaria o comportamento apresentado na distribuição do peso (Figura 4.10(a)). Um modelo de regressão linear foi ajustado aos dados considerando o peso do participante como variável resposta e a idade em anos, o sexo e a informação sobre diabetes como variáveis explicativas, e a Figura 4.10(b) representa os resíduos ordinários para este ajuste. A fonte de heterogeneidade não-observada apresentada na Figura 4.10(a) foi refletida nos resíduos do modelo, e assumir o modelo usual de regressão pode não ser suficiente para explicar o efeito de certas variáveis explicativas sobre o peso do indivíduo.

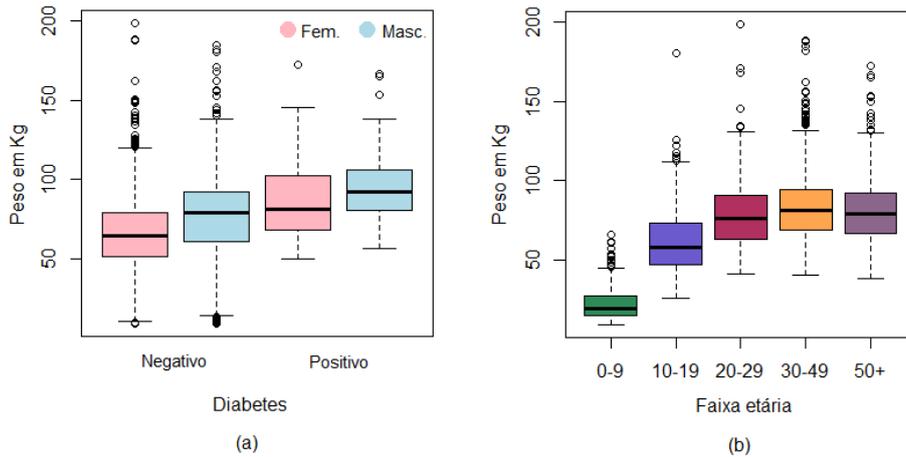


Figura 4.9: Boxplot para o peso em kg em função das variáveis (a) sexo e diabetes e (b) faixa etária.

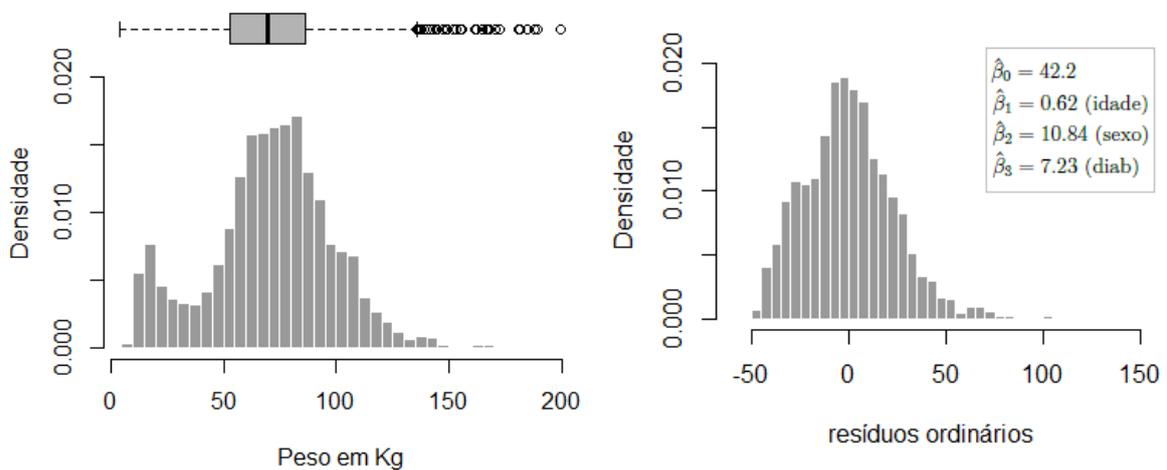


Figura 4.10: (a) Histograma e boxplot para o peso em quilogramas e (b) Histograma dos resíduos ordinários para o modelo de regressão linear.

Como forma de modelar a estrutura apresentada na Figura ??(b), que sugere a existência de pelo menos três modas, assumimos que a distribuição dos erros segue a densidade de mistura proposta na Seção 2.2. Consideramos o peso dos participantes como variável resposta e a idade em anos, o sexo e a informação sobre diabetes como variáveis explicativas. A título de comparação, os seguintes modelos foram considerados para modelar os erros:

- A. Modelo Proposto com $J = 2, 3, 4$, $K = 5$ e $\nu = (2.1, 2.3, 2.6, 3.2, 10)$ fixo;
- B. Mistura de t's com $J = 2, 3, 4$ e estimando o mesmo $\nu \forall j$;
- C. Mistura de t's com $J = 2, 3, 4$ e estimando ν 's diferentes $\forall j$;

D. Mistura de Normais com $J = 2, 3, 4$;

E. Mistura de Skew-t's com $J = 2, 3, 4$.

A Tabela 4.7 traz os resultados do DIC para os modelos A, B, C e D, e do BIC para o modelo E, assumindo $J = 2, 3$ e 4 no ajuste. Para os modelos A, B, C e D o DIC sugere que $J = 4$ oferece um melhor ajuste, ocorrendo o mesmo com o BIC no modelo E. Os resultados obtidos para cada modelo considerando os resultados do DIC e BIC são apresentados na Tabela 4.8, ressaltando que o modelo E foi ajustado através do pacote `FMsmnReg` que realiza inferência por máxima verossimilhança, via algoritmo EM, e considera que a distribuição dos erros pertence à classe SMSN (??).

Tabela 4.7: Critério DIC para diferentes valores de J no ajuste dos modelos A, B, C e D e critério BIC para diferentes valores de J no modelo E

Modelo	DIC		
	J=2	J=3	J=4
A	64519.5	64011.5	60138.8
B	68259.6	65789.4	58776.9
C	64822.1	64403.5	58614.6
D	68714.5	66387.7	64403.5
E	BIC		
	44814.10	44688.96	44686.42

A Figura 4.11 apresenta os resíduos do modelo com as curvas ajustadas para os modelos A, B, C e D. Observa-se que os modelos A, B e C forneceram ajustes similares para as curvas, em contrapartida, nota-se o modelo D precisaria de um número maior de componentes de maneira a capturar de maneira satisfatória o comportamento multimodal apresentado nos resíduos. Ao comparar os modelos nas caudas da distribuição, nota-se que o modelo A apresenta caudas levemente mais pesadas em relação aos demais modelos (Figura 4.12). Não foi possível comparar o ajuste da curva para o modelo E que ajusta uma mistura com duas componentes skew-t.

Os resultados obtidos indicam que as médias *a posteriori* para μ foram similares nos cinco modelos ajustados, com exceção da quarta moda no modelo A cuja média estimada foi em torno de 65, enquanto que nos demais modelos esta média ficou entre 45-49. Em relação aos resultados obtidos para σ^2 , observa-se que, exceto para a primeira moda, todas as médias *a posteriori* foram bem altas. Este resultado pode ter impacto no efeito dos regressores do modelo, especialmente no efeito de β_3 (associado à variável diabetes). No modelo A, por exemplo, o efeito deste regressor não foi estatisticamente significativo, ao contrário dos demais modelos. Os resultados *a posteriori* para

o peso \mathbf{w} indicam que a segunda e terceira modas possuem as maiores probabilidades. Em relação à estimação do parâmetro ν , observa-se que os modelos B e D apresentaram estimativas similares, enquanto que o modelo C estimou este parâmetro em torno de 8.9 para cada componente j . Em relação à estrutura de cauda estimada para o modelo A, através dos pesos $\hat{\mathbf{w}}$, observa-se que para as três primeiras modas foi atribuído maior peso para $\nu = 10$, sendo que para a segunda moda o peso estimado foi superior a 90%. Em relação à última moda os pesos atribuídos a cada ν_k ficaram em torno de 20%. Estes resultados indicam que para estes dados é importante considerar graus de liberdade menores que três para estimar de maneira satisfatória a cauda à direita. É importante ressaltar que as estimativas obtidas para o parâmetro de assimetria no modelo E sugerem uma assimetria forte na cauda à direita ($\hat{\lambda} = (-0.85, -0.99, 1.21)$). Dos cinco modelos apresentados na Tabela 4.8, o que possuiu um melhor ajuste segundo os critérios DIC e BIC apresentados na Tabela 4.7 foi o modelo E.

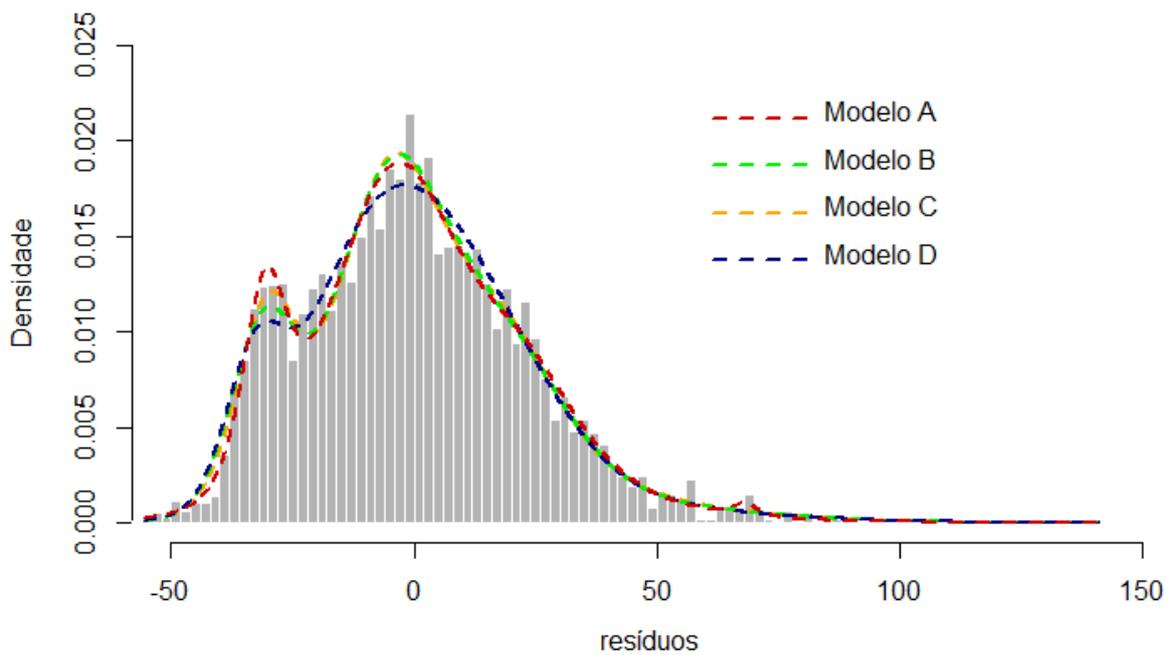


Figura 4.11: Resíduos do modelo com respectivas curvas ajustadas para os dados NHANES.

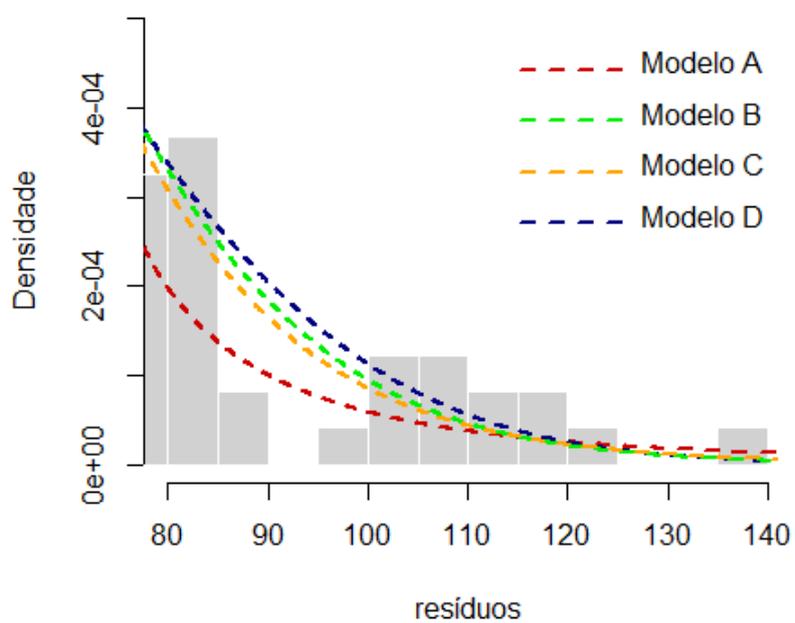


Figura 4.12: Resíduos do modelo nas caudas da distribuição com respectivas curvas ajustadas para os dados NHANES.

Tabela 4.8: Resultados *a posteriori** para o modelo de regressão com distribuição de mistura para os erros nos dados do estudo NHANES.

Modelo	μ	HPD	σ^2	HPD	β_0	HPD	β	HPD	ν	HPD	w	ψ
A	-30.569	[-31.670, -29.306]	23.165	[12.632, 34.575]	36.016	[29.824, 41.363]	0.653	[0.621, 0.684]	-	-	0.132	(0.091, 0.103, 0.136, 0.211, 0.459)
	-4.338	[-6.312, -2.379]	175.155	[129.420, 217.578]			5.393	[-3.939, 6.931]			0.587	(0.014, 0.016, 0.021, 0.033, 0.916)
	21.762	[17.292, 26.194]	200.965	[97.441, 281.056]			2.284	[-2.202, 6.856]			0.268	(0.072, 0.100, 0.156, 0.294, 0.378)
B	65.416	[48.999, 70.812]	54.159	[1.480, 283.806]	29.757	[26.757, 32.733]	0.609	[0.5799, 0.638]	21.857	[8.423, 44.422]	0.160	(0.180, 0.196, 0.210, 0.206, 0.208)
	-31.071	[-32.877, -29.304]	45.857	[33.245, 60.898]			8.148	[6.942, 9.355]			0.465	-
	-5.425	[-8.044, -2.354]	130.315	[65.758, 192.088]			6.046	[3.6315, 8.441]			0.322	-
C	16.133	[7.512, 24.328]	212.986	[92.206, 326.355]	31.279	[26.846, 36.423]	0.630	[0.594, 0.668]	8.972	[3.202, 19.714]	0.163	-
	46.671	[38.785, 58.560]	653.309	[402.611, 943.784]			7.351	[5.858, 9.001]	8.945	[2.914, 19.548]	0.473	-
	-30.400	[-32.209, -28.669]	39.334	[24.884, 54.697]			4.968	[1.210, 8.443]	8.927	[3.103, 19.672]	0.318	-
D	-5.206	[-7.532, -2.726]	125.720	[76.081, 171.323]	25.576	[22.008, 29.086]	0.608	[0.582, 0.635]	8.834	[3.221, 20.314]	0.098	-
	17.211	[11.035, 23.026]	188.025	[96.983, 287.084]			9.135	[7.824, 10.630]			0.213	-
	48.673	[38.775, 65.296]	503.485	[189.681, 937.521]			8.575	[6.173, 10.933]			0.639	-
E	-32.834	[-34.243, -31.499]	35.430	[24.921, 47.540]	24.277	-	0.641	-	20.0	-	0.256	-
	-11.681	[-12.977, -10.638]	170.459	[109.933, 233.437]			9.730	-			0.395	-
	5.595	[2.799, 8.660]	345.518	[299.081, 391.485]			7.848	-			0.262	-
	45.851	[38.639, 57.238]	819.223	[534.284, 1112.037]				-			0.087	-
	-28.973	-	102.428	-				-				-
	-5.625	-	92.080	-				-				-
	20.223	-	139.562	-				-				-
	49.796	-	693.159	-				-				-

* exceto para o modelo E.

Capítulo 5

Considerações Finais

Neste trabalho um modelo Bayesiano semi-paramétrico baseado em misturas finitas de distribuições t de Student foi introduzido. A abordagem proposta assume em sua especificação estruturas separadas para a modelagem de multimodalidade/assimetria e caudas pesadas. A mistura em dois níveis flexibiliza a modelagem de dados com diferentes estruturas, pois considera que o número necessário de componentes para acomodar multimodalidade e uma possível assimetria pode diferir do número de componentes para modelar a estrutura de cauda. Além disso, a modelagem das caudas não envolve a estimação de parâmetros de grau de liberdade. Uma extensão para modelos de regressão linear foi também apresentada onde assumimos que os erros do modelo seguem a mistura de distribuições proposta no trabalho. Como contribuição adicional, a abordagem proposta inclui os modelos univariados baseados em misturas de distribuições t de Student já existentes na literatura, estimando os parâmetros de grau de liberdade destes modelos a partir de um passo de Metropolis-Hastings incluído no MCMC proposto no trabalho. Outra contribuição da modelagem proposta é a possibilidade de se fazer classificação dos indivíduos no primeiro nível do modelo.

A performance da metodologia proposta foi avaliada através de estudos de simulação que contemplaram as várias subclasses de modelos derivadas da classe geral apresentada no Capítulo 2. Os resultados obtidos mostraram a flexibilidade do modelo em acomodar dados multimodais e com caudas pesadas, estimando a estrutura de cauda de maneira eficiente através dos pesos no segundo nível da mistura. Observamos que não houveram mudanças substanciais nos resultados *a posteriori* ao ajustar modelos com diferentes escolhas para o vetor de ν 's. Vale lembrar que neste trabalho o vetor associado ao parâmetro de grau de liberdade é previamente fixado. Em situações práticas recomendamos a estratégia apresentada no Capítulo 3, que utiliza divergência de Kulback-Liebler para definição de uma grade apropriada para ν . Ressaltamos ainda que a complexidade decorrente da estimação dos $K - 1$ pesos associados à grade assumida para ν é inferior em comparação à estimação do parâmetro de grau de liberdade no

modelo usual de mistura, cujo processo de estimação é custoso e problemático.

O modelo proposto também mostrou grande versatilidade em ajustar dados vindos de distribuições assimétricas. Uma vantagem ao aplicá-lo em dados com esta característica é que evitamos a estimação de parâmetros adicionais, como parâmetros de assimetria. Nos resultados obtidos a abordagem proposta apresentou, em geral, menores EQM's para a variância *a posteriori* de \mathbf{Y} quando comparados à mistura de t 's com parâmetro de grau de liberdade estimado, além de também apresentar as menores distâncias em relação à verdadeira distribuição geradora dos dados.

As aplicações com dados reais mostraram a flexibilidade da metodologia proposta em capturar a natureza dos diferentes dados analisados. Na análise relacionada ao comprimento dos peixes observamos uma similaridade entre os resultados obtidos para o modelo proposto em comparação com a mistura de distribuições t de Student onde diferentes graus de liberdade são estimados. E, segundo o DIC o ajuste baseado no modelo proposto forneceu melhores resultados em comparação aos demais modelos. Para os dados da galáxia foi possível mostrar a capacidade do método em capturar as eventuais modas existentes nos dados, apesar da pouca informação contida em algumas delas. Nesta aplicação o modelo proposto apresentou a maior variância *a posteriori* em comparação aos outros modelos, no entanto segundo o DIC foi o modelo que forneceu melhor ajuste aos dados. As duas aplicações no contexto de regressão também mostraram a versatilidade da abordagem proposta em capturar a multimodalidade e assimetria presentes nos resíduos de ambos os estudos, além de estimar a estrutura de cauda na mesma direção do modelo que utiliza componentes t de Student e estima o parâmetro de grau de liberdade. Observou-se, também, que para as duas aplicações os ajustes baseados na mistura de normais não forneceram boas aproximações.

Uma proposta para continuidade deste trabalho é considerar o número de componentes J desconhecido e estimá-lo utilizando alguma das abordagens existentes na literatura, como por exemplo, a metodologia proposta [Richardson e Green \(1997\)](#) que utiliza o método *reversible jump* para amostrar da distribuição conjunta de todos os parâmetros, incluindo o número de componentes, ou ainda a proposta de [Stephens \(2000a\)](#) que se baseia na construção de um processo de nascimento e morte para estimar o número de componentes. Outra proposta para continuidade do trabalho é avaliar a capacidade preditiva do modelo proposto em relação aos modelos de mistura até então existentes na literatura. Algumas extensões do modelo podem também ser consideradas, como por exemplo, assumir a modelagem proposta no contexto multivariado ou ainda no contexto de modelos mistos e modelos para dados censurados.

Apêndice A

Distribuição de Y no modelo proposto

Neste apêndice encontra-se a demonstração da distribuição marginal da variável aleatória Y no modelo proposto em 2.2.1.

$$\begin{aligned}
 f_Y(y) &= \prod_{i=1}^n \left[\int_0^\infty f(y_i|U_i, Z_{ij}, \dot{Z}_{ijk}, \mu_j, \sigma_j^2) f(U_i) du_i \right] \\
 &= \prod_{i=1}^n \left(\int_0^\infty \left[\sum_{j=1}^J w_j (2\pi\sigma_j^2 u_i^{-1})^{-\frac{1}{2}} \exp \left\{ -\frac{(y_i - \mu_j)^2}{2\sigma_j^2 u_i^{-1}} \right\} \right] \times \left[\sum_{k=1}^K \dot{w}_{jk} \frac{\left(\frac{\nu_k}{2}\right)^{\frac{\nu_k}{2}}}{\Gamma\left(\frac{\nu_k}{2}\right)} u_i^{\frac{\nu_k}{2}-1} \exp \left\{ \frac{\nu_k}{2} u_i \right\} \right] du_i \right) \\
 &= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\left(\frac{\nu_k}{2}\right)^{\frac{\nu_k}{2}}}{\Gamma\left(\frac{\nu_k}{2}\right)} \underbrace{\int_0^\infty u_i^{\left(\frac{\nu_k+1}{2}\right)-1} \exp \left\{ -\frac{(y_i - \mu_j)^2}{2\sigma_j^2} u_i - \frac{\nu_k}{2} u_i \right\} du_i}_{(*)} \right)
 \end{aligned}$$

(*) Núcleo de Gama(a,b), onde $a = \frac{\nu_k + 1}{2}$ e $b = \frac{(y_i - \mu_j)^2}{2\sigma_j^2} + \frac{\nu_k}{2}$. Logo:

$$\begin{aligned}
 &= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\left(\frac{\nu_k}{2}\right)^{\frac{\nu_k}{2}}}{\Gamma\left(\frac{\nu_k}{2}\right)} \frac{\Gamma(a)}{b^a} \underbrace{\int_0^\infty \frac{b^a}{\Gamma(a)} u_i^{a-1} \exp\{-bu_i\} du_i}_{=1} \right) \\
 &= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\left(\frac{\nu_k}{2}\right)^{\frac{\nu_k}{2}}}{\Gamma\left(\frac{\nu_k}{2}\right)} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\left[\frac{(y_i - \mu_j)^2}{2\sigma_j^2} + \frac{\nu_k}{2}\right]^{\frac{\nu_k+1}{2}}} \right) \\
 &= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\left(\frac{\nu_k}{2}\right)^{\frac{\nu_k}{2} + \frac{1}{2} - \frac{1}{2}}}{\Gamma\left(\frac{\nu_k}{2}\right)} \Gamma\left(\frac{\nu_k+1}{2}\right) \left[\frac{(y_i - \mu_j)^2}{2\sigma_j^2} + \frac{\nu_k}{2}\right]^{-\left(\frac{\nu_k+1}{2}\right)} \right)
 \end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\left(\frac{\nu_k}{2}\right)^{-\frac{1}{2}} \Gamma\left(\frac{\nu_k+1}{2}\right)}{\Gamma\left(\frac{\nu_k}{2}\right)} \left[\frac{\frac{\nu_k}{2}}{\frac{(y_i - \mu_j)^2}{2\sigma_j^2} + \frac{\nu_k}{2}} \right]^{\frac{\nu_k+1}{2}} \right) \\
&= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\left(\frac{\nu_k}{2}\right)^{\frac{1}{2}} \Gamma\left(\frac{\nu_k}{2}\right)} \left[\frac{\frac{\nu_k}{2} + \frac{(y_i - \mu_j)^2}{2\sigma_j^2}}{\frac{\nu_k}{2}} \right]^{-\left(\frac{\nu_k+1}{2}\right)} \right) \\
&= \prod_{i=1}^n \left(\sum_{j=1}^J w_j (2\pi\sigma_j^2)^{-\frac{1}{2}} \sum_{k=1}^K \dot{w}_{jk} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\left(\frac{\nu_k}{2}\right)^{\frac{1}{2}} \Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \frac{(y_i - \mu_j)^2}{\sigma_j^2 \nu_k} \right]^{-\left(\frac{\nu_k+1}{2}\right)} \right) \\
&= \prod_{i=1}^n \left(\sum_{j=1}^J w_j \left[\sum_{k=1}^K \dot{w}_{jk} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\Gamma\left(\frac{\nu_k}{2}\right) \sqrt{2\pi\sigma_j^2 \frac{\nu_k}{2}}} \left[1 + \frac{(y_i - \mu_j)^2}{\sigma_j^2 \nu_k} \right]^{-\left(\frac{\nu_k+1}{2}\right)} \right] \right) \\
&= \prod_{i=1}^n \left(\sum_{j=1}^J w_j \underbrace{\left[\sum_{k=1}^K \dot{w}_{jk} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\Gamma\left(\frac{\nu_k}{2}\right) \sqrt{\pi\sigma_j^2 \nu_k}} \left[1 + \frac{(y_i - \mu_j)^2}{\sigma_j^2 \nu_k} \right]^{-\left(\frac{\nu_k+1}{2}\right)} \right]}_{(**)} \right)
\end{aligned}$$

(**) corresponde à densidade da distribuição t de Student com parâmetros $(\mu_j, \sigma_j^2, \nu_k)$.

Logo:

$$f_Y(y) = \sum_{j=1}^J w_j \sum_{k=1}^K \dot{w}_{jk} f(y|\mu_j, \sigma_j^2, \nu_k).$$

Apêndice B

Distribuições Condicionais Completas

Neste apêndice encontram-se as expressões das condicionais completas para o modelo proposto no Capítulo 2 e sua extensão apresentada na Seção 2.4.

Para o bloco $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ a condicional completa é uma Normal-Gama-Inversa, dada por:

$$(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 | \cdot) \sim \mathcal{NGI}(\mu_0^*, \tau^*, \alpha^*, \hat{\beta}^*), \quad (\text{B.1})$$

onde

$$\begin{aligned} \mu_0^* &= \frac{\left(\sum_{i=1}^{n_1} U_i Y_i + \dots + \sum_{i=1}^{n_J} U_i Y_i \right) + \tau \mu_0}{\left(\sum_{i=1}^{n_1} U_i + \dots + \sum_{i=1}^{n_J} U_i \right) + \tau} \quad ; \\ \tau^* &= \left(\sum_{i=1}^{n_1} U_i + \dots + \sum_{i=1}^{n_J} U_i \right) + \tau \quad ; \quad \alpha^* = \hat{\alpha} + \frac{\sum_{i=1}^J n_1 + \dots + n_j}{2} ; \\ \hat{\beta}^* &= \frac{\left(\left(\sum_{i=1}^{n_1} U_i Y_i^2 + \dots + \sum_{i=1}^{n_J} U_i Y_i^2 \right) + 2\hat{\beta} + \tau \mu_0^2 \right) - \frac{\left(\left(\sum_{i=1}^{n_1} U_i Y_i + \dots + \sum_{i=1}^{n_J} U_i Y_i \right) + \tau \mu_0 \right)^2}{\left(\sum_{i=1}^{n_1} U_i + \dots + \sum_{i=1}^{n_J} U_i \right) + \tau}}{2}. \end{aligned}$$

A partir da condicional completa em (B.1) encontramos as distribuições marginais de $\boldsymbol{\sigma}^2$ e $\boldsymbol{\mu}$, dadas respectivamente por:

$$(\boldsymbol{\sigma}^2 | \cdot) \sim \mathcal{GI}(\alpha^*, \hat{\beta}^*), \quad (\text{B.2})$$

$$(\boldsymbol{\mu} | \cdot) \sim \mathcal{N}\left(\mu_0^*, \frac{\boldsymbol{\sigma}^2}{\tau^*}\right). \quad (\text{B.3})$$

Vale salientar que para gerar da condicional completa de $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ fazemos uso das distribuições marginais definidas em (B.2) e (B.3).

Para o bloco $(\boldsymbol{w}, \dot{\boldsymbol{w}})$ assumimos independência entre \boldsymbol{w} e $\dot{\boldsymbol{w}}$. Assim, suas condicionais completas possuem distribuição Dirichlet e são respectivamente dadas por:

$$(\boldsymbol{w}|\cdot) \sim \text{Dir}(\alpha_1 + n_1, \dots, \alpha_J + n_J), \quad (\text{B.4})$$

$$(\dot{\boldsymbol{w}}_j|\cdot) \sim \text{Dir}(\dot{\alpha}_{j1} + n_{j1}, \dots, \dot{\alpha}_{jK} + n_{jK}), \quad \forall j = 1, \dots, J. \quad (\text{B.5})$$

É importante ressaltar que $\dot{\boldsymbol{w}}$ é uma matriz de dimensão $J \times K$, onde cada linha é representada pelo vetor $\dot{\boldsymbol{w}}_j$, ou seja, $\dot{\boldsymbol{w}} = (\dot{w}_1, \dots, \dot{w}_J)^\top$.

Para amostrar do bloco $(\boldsymbol{U}, \boldsymbol{Z}, \dot{\boldsymbol{Z}})$, derivamos suas respectivas condicionais completas a partir da seguinte expressão:

$$\begin{aligned} \pi(\boldsymbol{U}, \boldsymbol{Z}, \dot{\boldsymbol{Z}}|\cdot) &\propto \pi(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi(\boldsymbol{U}|\boldsymbol{Z}, \dot{\boldsymbol{Z}}, \boldsymbol{\nu})\pi(\dot{\boldsymbol{Z}}|\boldsymbol{Z}, \dot{\boldsymbol{w}}_j)\pi(\boldsymbol{Z}|\boldsymbol{w}) \\ &\vdots \\ &\propto \prod_{i=1}^n \prod_{j=1}^J \left[\left(\prod_{k=1}^K \left(\pi(U_i|\cdot) \right)^{\dot{Z}_{ijk}} \right)^{Z_{ij}} \times \underbrace{\left(\prod_{k=1}^K p_{ijk} \right)^{Z_{ij}}}_{\pi(\dot{\boldsymbol{Z}}_{ij}|\cdot)} \times \underbrace{\tilde{p}_{ij}^{Z_{ij}}}_{\pi(\boldsymbol{Z}_i|\cdot)} \right]. \end{aligned}$$

Na expressão descrita acima $\boldsymbol{Z} = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n)^\top$ é uma matriz de dimensão $n \times J$ e $\dot{\boldsymbol{Z}} = (\dot{\boldsymbol{Z}}_1, \dots, \dot{\boldsymbol{Z}}_J)^\top$ uma matriz em blocos com dimensão $n \times K$, onde cada submatriz $\dot{\boldsymbol{Z}}_j$ em $\dot{\boldsymbol{Z}}$ possui dimensão $n_j \times K$, $\forall j = 1, \dots, J$.

Desta maneira, as distribuições condicionais completas para \boldsymbol{U} , $\dot{\boldsymbol{Z}}_{ij}$ e \boldsymbol{Z}_i são respectivamente dadas por:

$$\pi(U_i|\cdot) \sim \mathcal{G}\left(\frac{\nu_k + 1}{2}, \frac{\nu_k}{2} + \frac{(y_i - \mu_j)^2}{2\sigma_j^2}\right), \quad (\text{B.6})$$

$\forall i = 1, \dots, n_{jk}$, com $j = 1, \dots, J$ e $k = 1, \dots, K$, onde condicional ao j -ésimo componente da mistura, n_{jk} representa o número de indivíduos que pertencem ao componente k no segundo nível do modelo de mistura proposto.

$$\pi(\dot{\boldsymbol{Z}}_{ij}|\cdot) \sim \text{Mult}\left(1, p_{ij1}, \dots, p_{ijK}\right), \quad \forall i, j, \quad (\text{B.7})$$

$$\pi(\boldsymbol{Z}_i|\cdot) \sim \text{Mult}\left(1, \tilde{p}_{i1}, \dots, \tilde{p}_{iJ}\right), \quad \forall i, \quad (\text{B.8})$$

onde $p_{ijk} = \frac{r_{ikj}\dot{w}_{jk}}{\tilde{r}_{ij}}$, $\forall i = 1, \dots, n$; $\forall j = 1, \dots, J$; $\forall k = 1, \dots, K$; e

$$\tilde{p}_{ij} = \frac{\tilde{r}_{ij}w_j}{\tilde{p}_i}, \quad \forall i = 1, \dots, n, \quad \forall j = 1, \dots, J.$$

$$r_{ikj} = \frac{(\sigma_j^2)^{-\frac{1}{2}} \left(\frac{\nu_k}{2}\right)^{\frac{\nu_k}{2}} \Gamma\left(\frac{\nu_k+1}{2}\right)}{\Gamma\left(\frac{\nu_k}{2}\right) \left(\frac{(y_i - \mu_j)^2}{2\sigma_j^2} + \frac{\nu_k}{2}\right)^{\frac{\nu_k+1}{2}}}, \quad \tilde{r}_{ij} = \sum_{k=1}^K r_{ikj}\dot{w}_j \quad \text{e} \quad \tilde{p}_i = \sum_{j=1}^J p_{ij}w_j$$

É válido mencionar que cada vetor $\dot{\mathbf{Z}}_{ij}$ representa uma linha da submatriz $\dot{\mathbf{Z}}_j$, cuja representação para dado j é da forma:

$$\dot{\mathbf{Z}}_j = \begin{pmatrix} \dot{Z}_{1j1} & \dot{Z}_{1j2} & \cdots & \dot{Z}_{1jK} \\ \dot{Z}_{2j1} & \dot{Z}_{2j2} & \cdots & \dot{Z}_{2jK} \\ \vdots & \vdots & \vdots & \vdots \\ \dot{Z}_{nj1} & \dot{Z}_{nj2} & \cdots & \dot{Z}_{njK} \end{pmatrix}$$

Já a matriz \mathbf{Z} possui a seguinte representação:

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1J} \\ Z_{21} & Z_{22} & \cdots & Z_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nJ} \end{pmatrix}$$

As contas das condicionais completas para a extensão proposta no contexto de regressão são mantidas, acrescidas da distribuição *a priori* para $\boldsymbol{\beta} \sim \mathcal{N}_p(\theta, v^2 I_p)$, cuja distribuição condicional completa é dada por:

$$(\boldsymbol{\beta}|\cdot) \sim \mathcal{N}_p\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right),$$

$$\text{onde } \boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \left[(v^2 I_p)^{-1} \theta + \frac{(\sqrt{u_i} \odot \mathbf{X}_i)^\top (\sqrt{u_i} \odot \mathbf{Y}_i)}{\sigma^2} \right] \text{ e}$$

$$\boldsymbol{\Sigma}_\beta = \left[(v^2 I_p)^{-1} + \frac{(\sqrt{u_i} \odot \mathbf{X}_i)^\top (\sqrt{u_i} \odot \mathbf{X}_i)}{\sigma^2} \right]^{-1}.$$

Apêndice C

Outros resultados de Simulação

Neste Apêndice encontram-se resultados adicionais dos estudos de simulação apresentados no Capítulo 3. Os resultados obtidos corroboram com o que foi apresentado anteriormente. Ressalta-se, entretanto, que para tamanhos amostrais pequenos nem sempre é possível realizar uma boa inferência nas caudas da distribuição por não haver informação suficiente. À medida que o tamanho amostral cresce, as médias *a posteriori* para os parâmetros dos modelos considerados para ajuste tendem a se aproximar dos seus respectivos EMV's (exceto para a configuração 5, onde não foi possível obter as estimativas de máxima verossimilhança), calculados com base no modelo que gerou os dados.

Dados vindos de uma distribuição t de Student

Nesta Seção apresentamos resultados adicionais para o estudo apresentado no Capítulo 3. Para geração dos dados assumimos uma distribuição t de Student com $\mu = 0$, $\sigma^2 = 1$, $\nu = 4$ e ajustamos o mesmo modelo de mistura utilizado no Capítulo 3 para diferentes tamanhos de amostra. Os resultados obtidos são apresentados na Tabela C.1 e Figura C.1.

Observa-se que à medida que o tamanho da amostra cresce, as médias *a posteriori* para μ e σ^2 se aproximam dos seus respectivos EMV's (Tabela C.1). Assim como nos resultados obtidos para $\nu = 2$ e $\nu = 3$, constatamos que independente da combinação de pesos utilizada como chute inicial para a mistura, as probabilidades *a posteriori* para \hat{w} atribuem maior peso para o grau de liberdade na mistura que é mais próximo do utilizado na geração dos dados.

Na Figura C.1 nota-se que a densidade de mistura se aproxima da densidade de uma única t de Student à medida que o tamanho da amostra cresce. Os resultados obtidos nesta Seção corroboram com os que foram apresentados no Capítulo 3 e cons-

tatamos que uma mistura de distribuições t de Student com graus de liberdade fixos aproxima bem uma única t com grau de liberdade arbitrário, mas definido no intervalo de variação do vetor ν utilizado na mistura.

Tabela C.1: Resultados *a posteriori* para a mistura de t's considerando diferentes tamanhos amostrais e $\nu = 4$ na geração da t de Student

n	EMV	μ	HPD	EMV	σ^2	HPD	\hat{w}
100	-0.085	-0.089	[-0.284, 0.111]	0.751	0.694	[0.402, 0.979]	(0.194, 0.376, 0.430)
500	-0.045	-0.046	[-0.150, 0.060]	1.072	1.011	[0.807, 1.215]	(0.092, 0.276, 0.632)
2500	-0.040	-0.041	[-0.086, 0.005]	0.989	0.969	[0.873, 1.070]	(0.027, 0.265, 0.708)

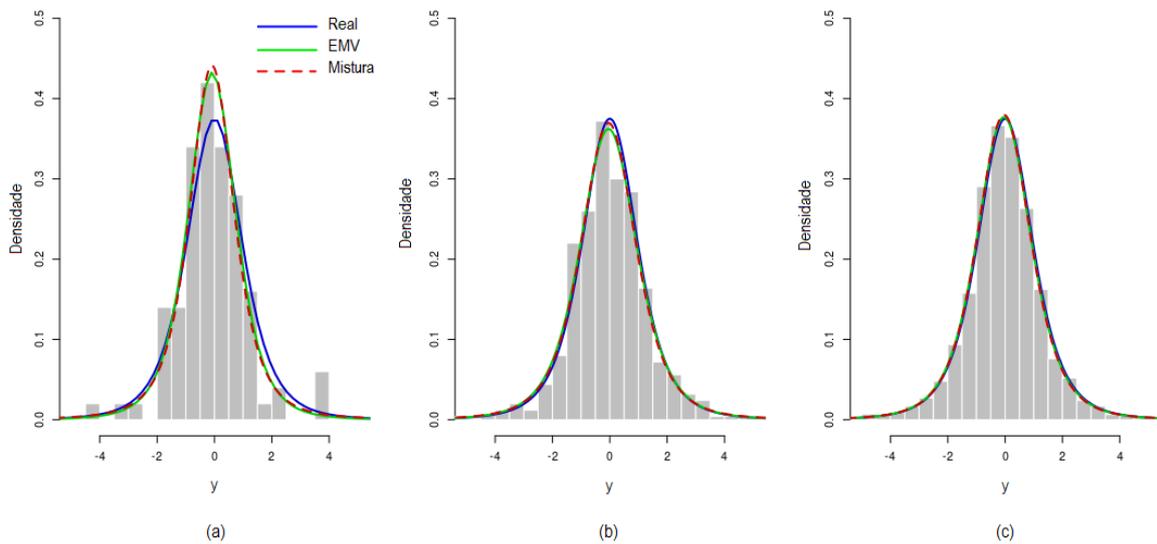


Figura C.1: Comparação entre as curvas de uma densidade t de Student com $\nu = 4$ (azul), densidade da t baseada no EMV (verde) e densidade estimada para a mistura de 3 t's com $\nu = (1.5, 2.5, 5)$ (vermelha) quando (a) $n=100$; (b) $n=500$ e (c) $n= 2500$.

Dados de uma mistura de distribuições t de Student unimodal

Os resultados apresentados nesta Seção complementam o estudo realizado no Capítulo 3, no entanto, avaliamos a influência do tamanho da amostra nos resultados *a posteriori*. Os resultados baseados nos modelos B, C e D são similares às estimativas

de máxima verossimilhança obtidas para o modelo A. É possível observar que as probabilidades *a posteriori* para \hat{w} nos modelos C e D capturaram a verdadeira estrutura de cauda usada na geração dos dados ao atribuir maiores pesos às componentes da mistura próximas do verdadeiro ν . Há uma melhor aproximação das curvas ajustadas quando o tamanho da amostra gerado é 1000. Além disso, para este tamanho de amostra o modelo B foi o que apresentou a menor DKS em relação ao valor real (verde), seguido do modelo A (Figura C.2). Em relação a DVT, nota-se que os modelos B, C e D são os que obtiveram as menores distâncias em relação à verdadeira distribuição que gerou os dados.

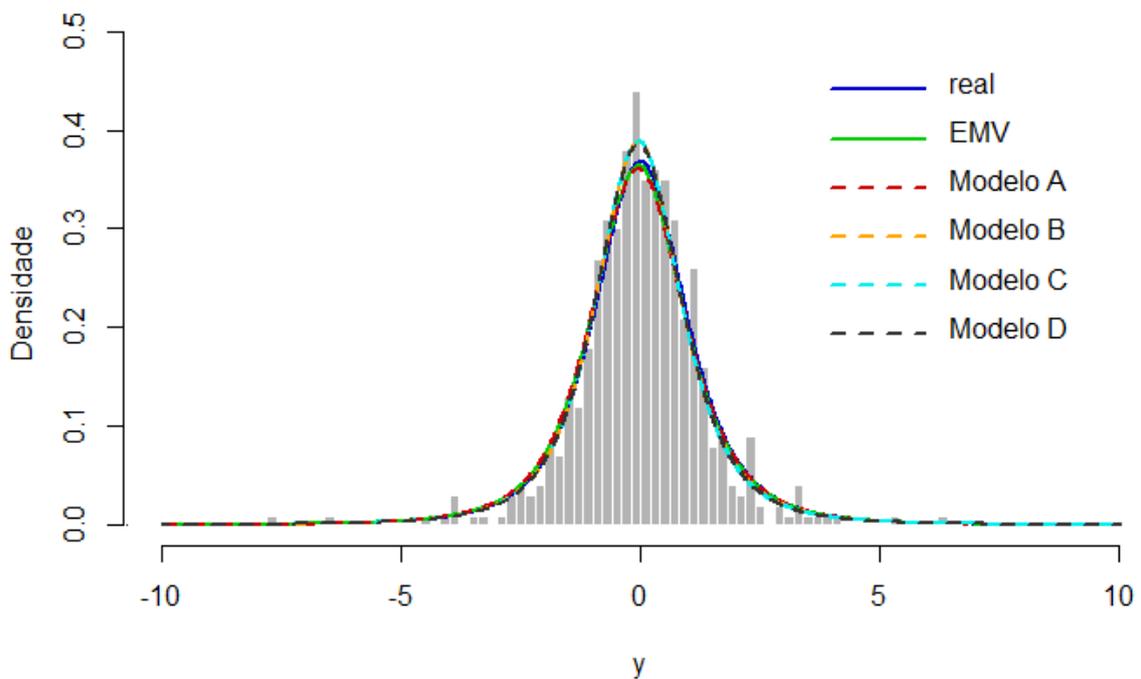


Figura C.2: Histograma para o tamanho de amostra 1000 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.

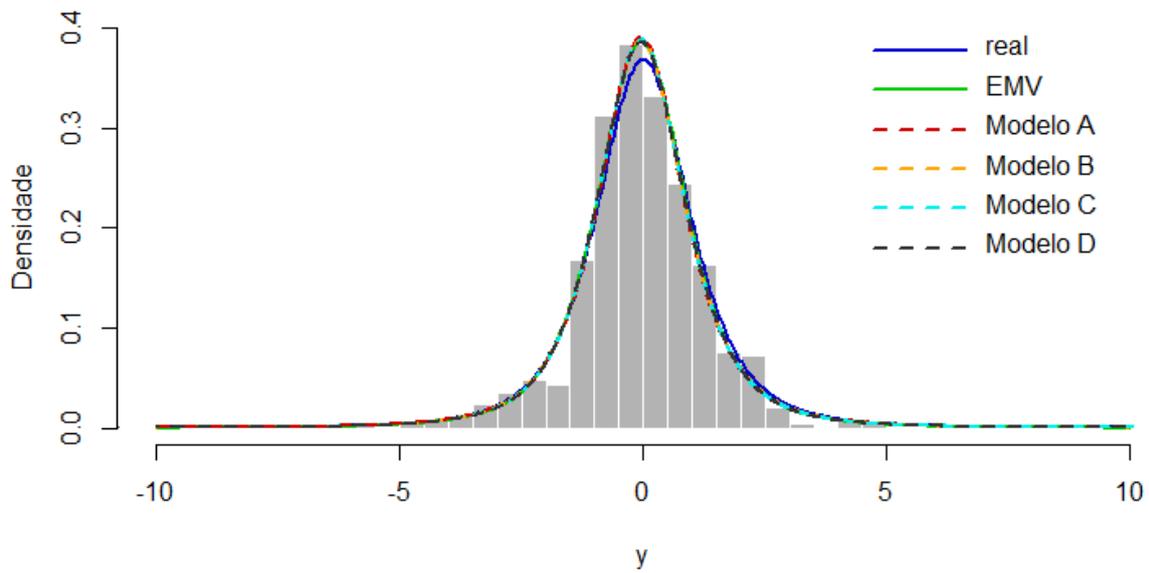


Figura C.3: Histograma para o tamanho de amostra 500 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.

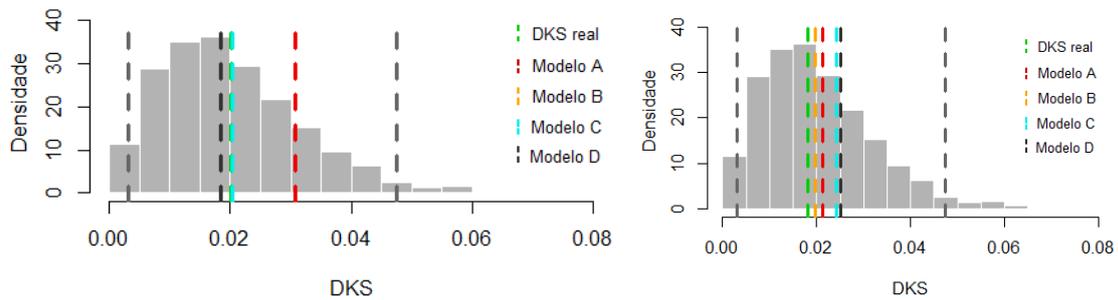


Figura C.4: Histograma para tamanhos de amostra 500 e 1000 da amostra Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.

Tabela C.2: Resultados *a posteriori** para uma mistura de distribuições t de Student com $K = 2$ e tamanho de amostra 1000.

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	\hat{w}	DVT
A	-0.045	-0.047	-	1.050	1.037	[0.895, 1.206]	-	0.0092
B	-0.045	-0.047	[-0.124, 0.025]	1.025	1.037	[0.895, 1.206]	(0.432, 0.568)	0.0039
C	-0.045	-0.045	[-0.120, 0.028]	1.074	1.074	[0.912, 1.235]	(0.341, 0.659)	0.0159
D	-0.046	-0.046	[-0.126, 0.027]	1.079	1.079	[0.931, 1.242]	(0.161, 0.517, 0.322)	0.0175

$DVT_{real} = 0.0188$ *exceto para o modelo A.

Tabela C.3: Resultados *a posteriori** para uma mistura de distribuições t de Student com $K = 2$ e tamanho de amostra 500.

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	\hat{w}	DVT
A	-0.038	-0.040	-	0.865	0.899	-	-	0.0223
B	-0.038	-0.041	[-0.145, 0.054]	0.907	0.899	[0.710, 1.067]	(0.672, 0.328)	0.0059
C	-0.039	-0.039	[-0.143, 0.059]	0.890	0.890	[0.708, 1.098]	(0.568, 0.432)	0.0067
D	-0.035	-0.035	[-0.134, 0.071]	0.908	0.908	[0.716, 1.101]	(0.355, 0.333, 0.312)	0.0068

$DVT_{real} = 0.0243$ *exceto para o modelo A.

Dados vindos de uma mistura usual de distribuições t de Student

Nesta seção apresentamos resultados adicionais para amostras de tamanho 500 e 1000 geradas a partir de uma mistura de distribuições t de Student assumindo o mesmo ν em cada componente. A metodologia proposta é comparada ao modelo que gerou os dados e avaliamos se há efeito do tamanho da amostra nos resultados *a posteriori*. A mesma configuração de parâmetros utilizada no Capítulo 3 foi assumida para geração dos dados.

Os resultados *a posteriori* para as amostras de tamanho 500 e 1000 corroboram com os resultados apresentados no Capítulo 3 para $n = 2500$. Observamos a similaridade nas médias *a posteriori* dos modelos B e C em relação às estimativas obtidas para o modelo A. Ao contrário do observado no Capítulo 3, as menores DVT's foram para os modelos B e C, onde é possível observar uma melhor aproximação das curvas ajustadas por estes modelos. Apesar de distantes da DKS real, os modelos B e C foram os que mais se aproximaram desta (Figura C.7).

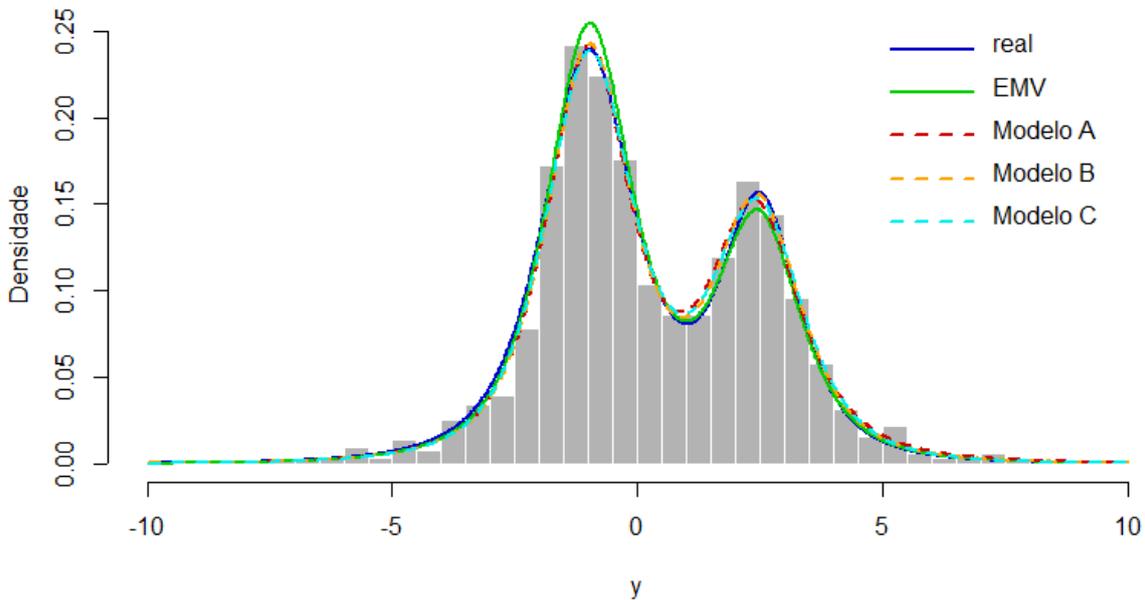


Figura C.5: Histograma para o tamanho de amostra 1000 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.

Tabela C.4: Resultados *a posteriori** para a mistura de t de Student com $J = 2$ e $K = 1$ e tamanho de amostra 1000

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	w	\dot{w}	DVT
A		-0.996 2.412	- -	0.852 0.997		- -	0.602 0.398	- -	0.0315
B	-0.978	-0.978	[-1.088, -0.864]	0.889	0.880	[0.666, 1.108]	0.618	(0.654, 0.346)	0.0289
C	2.469	2.458	[2.289, 2.599]	0.855	0.898	[0.610, 1.210]	0.382	(0.496, 0.504)	
		-0.980	[-1.099, -0.869]		0.925	[0.711, 1.155]	0.618	(0.618, 0.229, 0.153)	0.0270
		2.466	[2.316, 2.622]		0.944	[0.669, 1.244]	0.382	(0.447, 0.310, 0.243)	

$DVT_{real} = 0.0219$ *exceto para o modelo A.

Tabela C.5: Resultados *a posteriori** para a mistura de t de Student com $J = 2$ e $K = 1$ e tamanho de amostra 500

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	w	\dot{w}	DVT
A		-0.984 2.435	- -	1.033 1.253		- -	0.626 0.374	- -	0.0304
B	-0.960	-0.969	[-1.142, -0.782]	0.944	1.053	[0.677, 1.459]	0.631	(0.342, 0.658)	0.0203
C	2.485	2.435	[2.121, 2.752]	1.032	1.219	[0.698, 1.804]	0.369	(0.341, 0.659)	
		-0.972	[-1.148, -0.779]		1.094	[0.713, 1.464]	0.630	(0.279, 0.399, 0.322)	0.0262
		2.442	[2.143, 2.757]		1.272	[0.714, 1.889]	0.370	(0.285, 0.384, 0.331)	

$DVT_{real} = 0.0315$ *exceto para o modelo A.

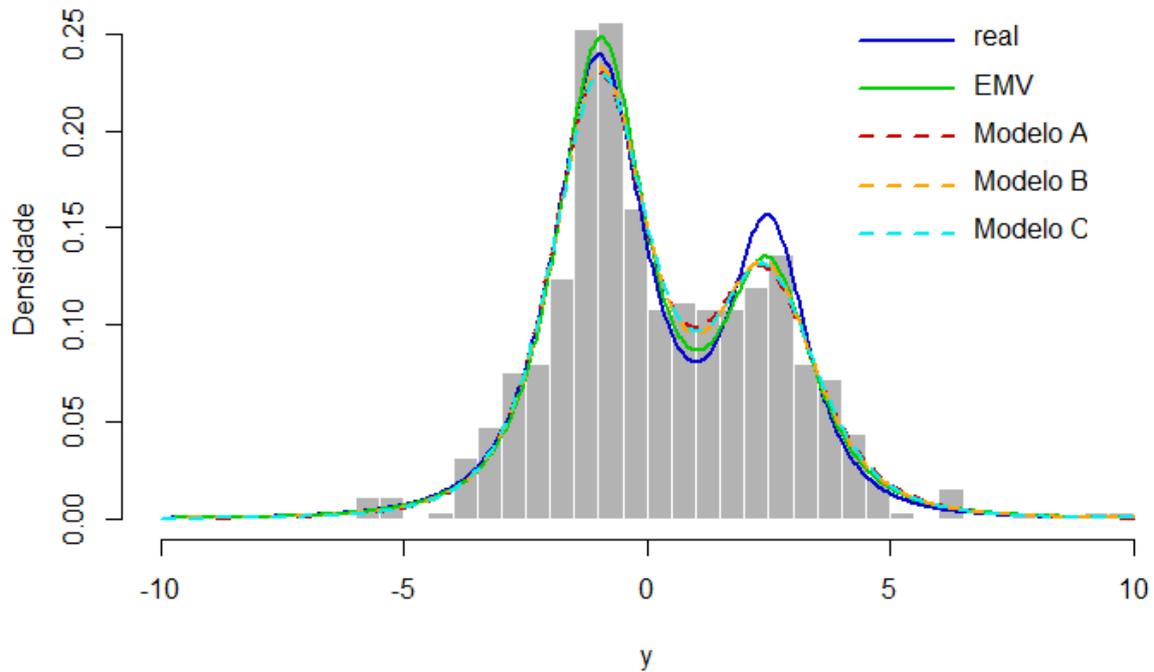


Figura C.6: Histograma para o tamanho de amostra 500 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.

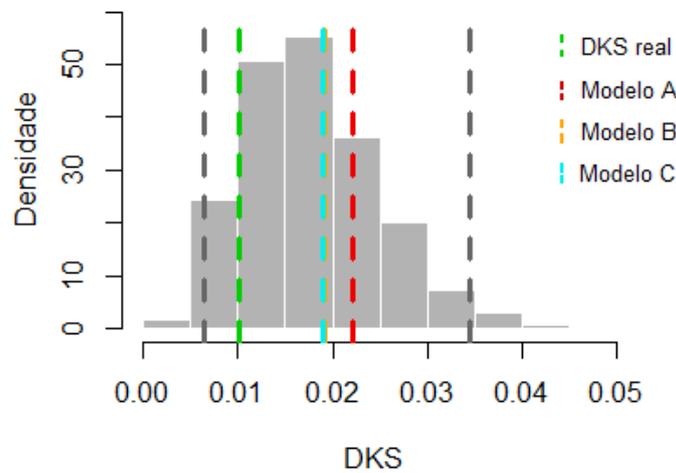


Figura C.7: Histograma para tamanho amostral 1000 da amostra Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.

Dados vindos da mistura de distribuições t de Student proposta

Nesta configuração apresentamos resultados adicionais para o estudo onde dados são gerados com base na subclasse de modelos que considera $J \geq 2$ e $K \geq 2$. Amostras de tamanho 500 e 1000 foram geradas com o objetivo de avaliar a performance da metodologia proposta para tamanhos de amostra menores. Para geração dos dados assumimos a mesma configuração de parâmetros e o mesmos modelos assumidos para ajuste no Capítulo 3.

As Tabelas C.6 e C.7 apresentam os resultados *a posteriori* obtidos. Observa-se que as médias *a posteriori* para μ e σ^2 se aproximam dos seus respectivos EMV's nos modelos B, C e D quando o tamanho da amostra cresce. O mesmo ocorre com as probabilidades *a posteriori* para o peso w . Um comportamento similar ao que foi apresentado no Capítulo 3 para amostra de tamanho 2500 é observado para as probabilidades *a posteriori* do peso w . O modelo A estimou o parâmetro ν igual a 2.29 na amostra de tamanho 500 e $\nu = 2.19$ na amostra de tamanho 1000.

As Figuras C.6 e C.7 mostram o ajuste das curvas para cada modelo. Nota-se que uma melhor aproximação em relação à curva baseada no EMV ocorre quando o tamanho da amostra cresce. Além disso, comparando o modelo A com o demais modelos ajustados, verifica-se que este possui uma DKS maior, ocorrendo o mesmo quando avaliamos as DVT's de cada modelo. O modelo D foi o que obteve a menor DKS em relação ao valor real, ocorrendo o mesmo em relação à DVT, para ambos tamanhos de amostra. Portanto, este modelo é o que fornece um melhor ajuste aos dados, quando comparado ao modelo B, que ajustou o modelo verdadeiro.

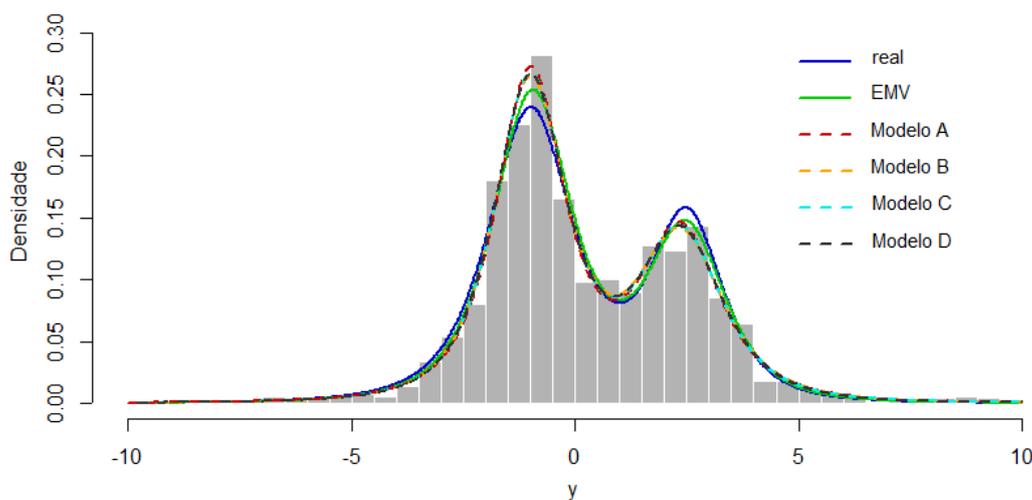


Figura C.8: Histograma para o tamanho de amostra 1000 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.

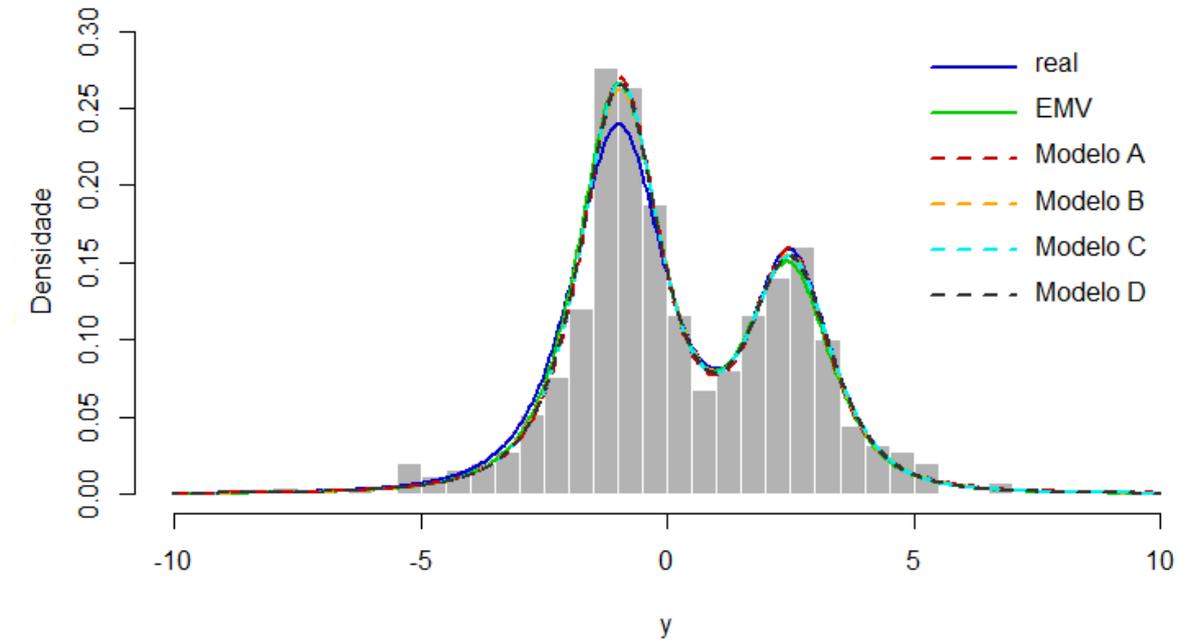


Figura C.9: Histograma para o tamanho de amostra 500 com o verdadeiro modelo usado para gerar os dados e curvas para diferentes ajustes.

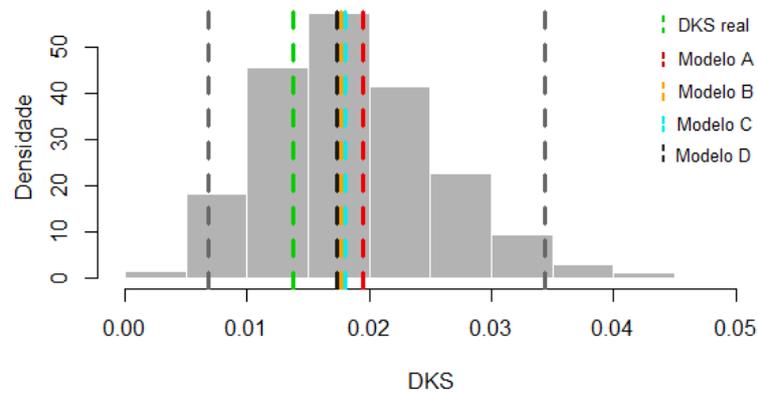


Figura C.10: Histograma para tamanho amostral 1000 da amostra Monte Carlo da distância de Kolmogorov-Smirnov entre a função de distribuição real e a baseada no EMV para o modelo verdadeiro, com bandas de 95% de confiança (cinza) e DKS's calculadas para diferentes modelos ajustados.

Tabela C.6: Resultados *a posteriori** para os parâmetros do modelo proposto considerando tamanho de amostra 1000

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	EMV	w	\hat{w}	DVT
A		-0.998 2.356	-		0.702 0.960	-		0.623 0.377	-	0.0318
B	-0.956 2.486	-1.005 [2.192, 2.554]	[-1.110, -0.901]	0.898 0.862	0.765 1.065	[0.584, 0.945] [0.728, 1.422]		0.621 0.379	(0.762, 0.238) (0.718, 0.282)	0.0254
C		-1.000 2.361	[-1.098, -0.893] [2.197, 2.541]		0.7489 1.014	[0.591, 0.944] [0.684, 1.366]		0.624 0.376	(0.781, 0.219) (0.737, 0.263)	0.0264
D		-1.005 2.357			0.757 1.029			0.623 0.377	(0.530, 0.289, 0.181) (0.492, 0.301, 0.207)	0.0263

$DVT_{real} = 0.0252$ *exceto para o modelo A.

Tabela C.7: Resultados *a posteriori** para os parâmetros do modelo proposto considerando tamanho de amostra 500

Modelos	EMV	μ	HPD	EMV	σ^2	HPD	EMV	w	\hat{w}	DVT
A		-0.953 2.487	-		0.747 0.746	-		0.638 0.362	-	0.0197
B	-0.987 2.459	-0.953 [2.301, 2.689]	[-1.084, -0.815]	0.814 0.818	0.820 0.801	[0.560, 1.075] [0.480, 1.131]		0.643 0.357	(0.706, 0.294) (0.525, 0.475)	0.0144
C		-0.952 2.504	[-1.079, -0.815] [2.311, 2.713]		0.788 0.816	[0.535, 1.047] [0.496, 1.166]		0.645 0.355	(0.824, 0.176) (0.523, 0.477)	0.0163
D		-0.951 2.501	[-1.093, -0.820] [2.289, 2.687]		0.799 0.819	[0.5404, 1.068] [0.507, 1.176]		0.644 0.356	(0.546, 0.319, 0.135) (0.331, 0.360, 0.309)	0.0160

$DVT_{real} = 0.0287$ *exceto para o modelo A.

Erros com distribuição bimodal

Os resultados apresentados nesta configuração seguem o mesmo objetivo do que foi apresentado no Capítulo 3, no entanto foram gerados dados com tamanhos amostrais 500 e 1000 como forma de avaliar a performance da extensão proposta para tamanhos de amostra menores. O *lag* considerado foi de tamanho 34 com amostra a posteriori de tamanho 2500, para os dados gerados com $n = 500$. Os modelos assumidos para ajuste foram:

- A. $J = K = 2$ com $\nu = (2.3, 3.5)$;
- B. $J = K = 2$ com $\nu = (2.05, 4.5)$;
- C. $J = K = 2$ com $\nu = (1.9, 3, 5)$.

As Tabelas C.8 e C.9 apresentam os resultados *a posteriori* obtidos para os tamanhos amostrais 1000 e 500, respectivamente. Analisando coletivamente os resultados, observamos que as médias *a posteriori* para μ foram bem similares nos dois tamanhos amostrais, no entanto os resultados são mais próximos dos valores reais quando o tamanho amostral considerado é 1000. Os intervalos HPD para todos os parâmetros incluem os valores reais. Observamos também que as probabilidades *a posteriori* para w foram bem estimadas. Além disso, as estruturas de cauda assumidas nos modelos B e C conseguiram estimar de maneira satisfatória as probabilidades *a posteriori* para \dot{w} e ambos os modelos dão maiores pesos para as componentes que mais se aproximam dos valores reais de ν .

Nas Figuras C.8 e C.9 observa-se que as curvas ajustadas forneceram melhores aproximações quando o tamanho amostral é 1000. Além disso, independente da grade assumida para ν , os modelos B e C forneceram aproximações similares ao modelo A que ajustou o modelo verdadeiro.

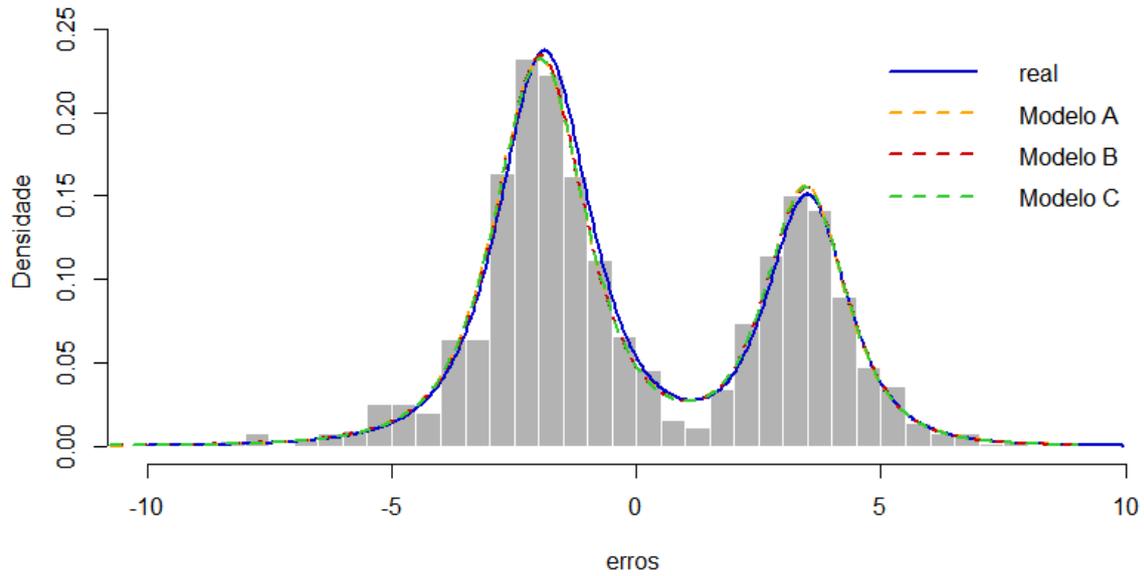


Figura C.11: Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas para a amostra de tamanho 1000.

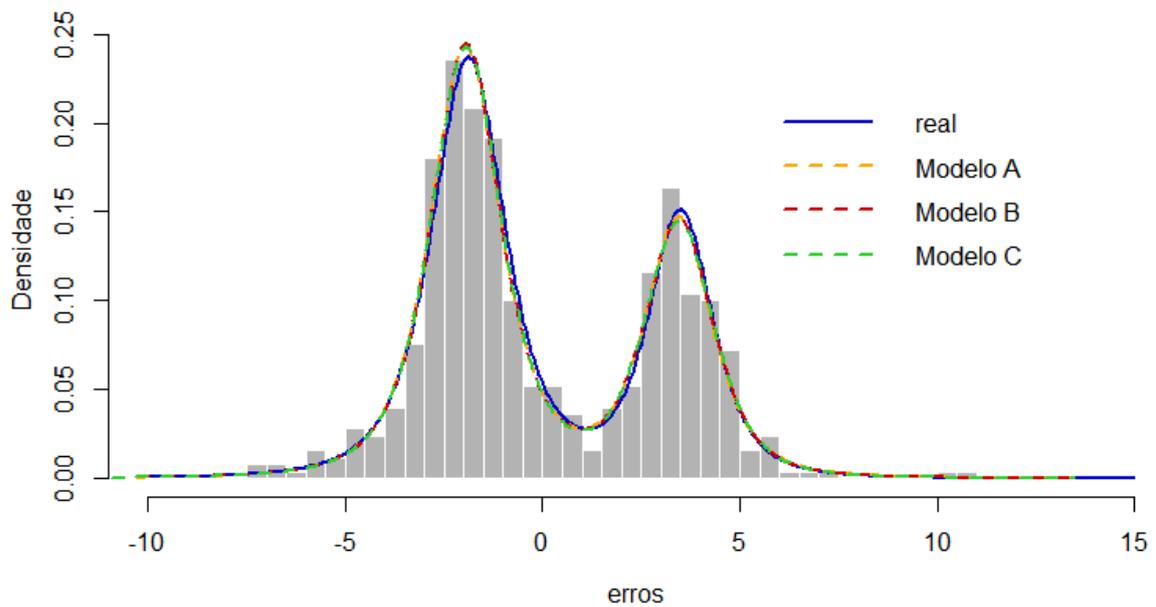


Figura C.12: Histograma dos erros baseado no modelo verdadeiro com a curva real e as respectivas curvas ajustadas para a amostra de tamanho 500.

Tabela C.8: Resultados *a posteriori* para o modelo de regressão com distribuição de mistura para os erros considerando $n=1000$

Modelos	μ^\top	HPD	$\hat{\mu}^*$	HPD	σ^2	HPD	β_0	HPD	β	HPD	w	\hat{w}
A	-1.971	[-2.069, -1.881]	-0.907	[-1.058, -0.775]	1.003	[0.799, 1.189]	1.064	[0.925, 1.182]	-1.953	[-2.047, -1.858]	0.636	(0.702, 0.298)
	3.459	[3.335, 3.579]	4.523	[4.371, 4.674]	0.752	[0.570, 0.949]			0.928	[0.752, 1.100]	0.364	(0.435, 0.565)
B	-1.961	[-2.049, -1.857]	-0.904	[-1.053, -0.760]	0.981	[0.775, 1.197]	1.056	[0.919, 1.187]	-1.952	[-2.041, -1.863]	0.639	(0.738, 0.262)
	3.469	[3.337, 3.600]	4.525	[4.380, 4.699]	0.774	[0.573, 0.970]			0.928	[0.747, 1.094]	0.361	(0.359, 0.641)
C	-1.968	[-2.054, -1.868]	-0.908	[-1.053, -0.773]	1.006	[0.794, 1.238]	1.060	[0.926, 1.186]	-1.951	[-2.041, -1.861]	0.638	(0.459, 0.347, 0.194)
	3.465	[3.345, 3.590]	4.525	[4.373, 4.672]	0.769	[0.576, 0.955]			0.929	[0.752, 1.096]	0.362	(0.234, 0.380, 0.386)

Tabela C.9: Resultados *a posteriori* para o modelo de regressão com distribuição de mistura para os erros considerando $n=500$

Modelos	μ	HPD	μ^*	HPD	σ^2	HPD	β_0	HPD	β	HPD	w	\hat{w}
A	-1.952	[-2.084, -1.821]	-0.761	[-0.953, -0.564]	0.925	[0.685, 1.180]	1.191	[1.014, 1.370]	-2.058	[-2.190, -1.944]	0.639	(0.607, 0.393)
	3.449	[3.260, 3.644]	4.640	[4.407, 4.869]	0.838	[0.563, 1.125]			0.915	[0.686, 1.176]	0.361	(0.464, 0.536)
B	-1.935	[-2.079, -1.804]	-0.752	[-0.948, -0.559]	0.912	[0.660, 1.171]	1.183	[0.997, 1.354]	-2.054	[-2.180, -1.935]	0.642	(0.692, 0.308)
	3.461	[3.267, 3.640]	4.646	[4.410, 4.879]	0.847	[0.574, 1.179]			0.911	[0.663, 1.159]	0.358	(0.401, 0.599)
C	-1.936	[-2.077, -1.805]	-0.755	[-0.953, -0.567]	0.936	[0.679, 1.217]	1.181	[0.993, 1.370]	-2.051	[-2.187, -1.933]	0.641	(0.4545, 0.360, 0.185)
	3.466	[3.268, 3.647]	4.647	[4.423, 4.873]	0.874	[0.579, 1.188]			0.918	[0.671, 1.171]	0.359	(0.276, 0.325, 0.399)

Modelando assimetria e caudas pesadas

1) Dados vindos de uma distribuição Skew-t

Para este estudo de simulação foram geradas 1000 observações vindas de uma distribuição Skew-t com $\mu = 0$, $\sigma^2 = 1$, $\nu = 2.5$ e $\lambda = 1.5$. De maneira análoga à análise apresentada no Capítulo 3, avaliamos o viés, variância e EMQ dos resultados *a posteriori* referentes a esperança e variância de \mathbf{Y} para cada modelo considerado para ajuste. A medida que avalia as densidades real e estimada também é avaliada.

Os modelos de mistura comparados neste estudo foram:

- A. Mistura de Normais com $J = 2, 3$;
- B. Mistura de t's com $J = 2, 3$ e estimando $\nu_1 = \dots = \nu_J = \nu$;
- C. Mistura de t's com $J = 2, 3$ e estimando $\nu_1 \neq \dots \neq \nu_J$;
- D1. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.3, 5)$ fixo;
- D2. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.2, 9)$ fixo.

Tabela C.10: viés, variância e erro quadrático médio para a esperança e variância de Y em diferentes distribuições de mistura quando os dados são gerados de uma distribuição skew-t

Modelos	$\widehat{E(Y)}$			$\widehat{Var(Y)}$			\bar{D}	\bar{D}_{cauda}
	vício	var	EQM	vício	var	EQM		
<u>A</u>								
$J = 2$	0.0271	0.0033	0.0041	0.7635	0.4247	1.0076	0.852	1.414
$J = 3$	0.0355	0.0032	0.0044	0.9009	0.6993	1.5109	0.437	0.640
<u>B</u>								
$J = 2$	-0.1852	0.0011	0.0353	7.1516	201.37	252.52	0.319	0.490
$J = 3$	-0.0255	0.0021	0.0028	-0.1243	0.4964	0.5118	0.402	0.659
<u>C</u>								
$J = 2$	-0.1693	0.0012	0.0299	1.5135	35.9544	38.2452	0.311	0.490
$J = 3$	-0.0127	0.0023	0.0024	-0.1538	0.4116	0.4353	0.421	0.687
<u>D1</u>								
$J = K = 2$	-0.1484	0.0011	0.0231	-0.2043	0.2108	0.2526	0.284	0.434
$J = 3, K = 2$	-0.0432	0.0021	0.0039	1.5864	2.5792	5.0961	0.400	0.669
<u>D2</u>								
$J = K = 2$	-0.1426	0.0010	0.0214	4.6005	1.4998	22.6648	0.280	0.414
$J = 3, K = 2$	-0.0535	0.0021	0.0049	6.5468	11.9294	54.7906	0.475	0.848

$E(Y) = 1.00348$ (média amostral: 1.04199); $var(Y) = 3.99303$ (variância amostral: 4.88363)

Analisando os resultados na Tabela C.10 observamos que o viés, variância e EQM para a cadeia da $E(Y)$ não apresenta diferenças substanciais entre os modelos ajustados. Já em relação à cadeia da $Var(Y)$, observa-se que o cálculo destas três medidas apresenta valores razoavelmente altos. No modelo D2 uma possível explicação para o EQM apresentar valores altos, qualquer que seja o número de componentes J

usado para aproximar a densidade que gerou os dados, é a má escolha da grade para ν . Como o cálculo da variância é função de ν , uma escolha ruim desta grade pode afetar o desempenho do modelo. O cálculo da distância de variação percentual mostra que os modelos $D1$ e $D2$ com $J = 2$ componentes apresentaram a menores distâncias globalmente e nas caudas, dentre todos os modelos ajustados. Este resultado indica que o modelo proposto fornece uma melhor aproximação, principalmente nas caudas da distribuição.

Na Figura C.13 apresentamos o histograma para a amostra gerada a partir da distribuição Skew-t e comparamos as curvas baseadas nos melhores ajustes com a curva real. Assumimos como melhor ajuste aquele que forneceu a menor distância \bar{D} globalmente e nas caudas, apesar desta escolha acarretar no maior EQM para os modelos B e C . Dentre os modelos apresentados na Figura C.13, o modelo $D1$ foi o que forneceu a menor distância em relação à curva real.

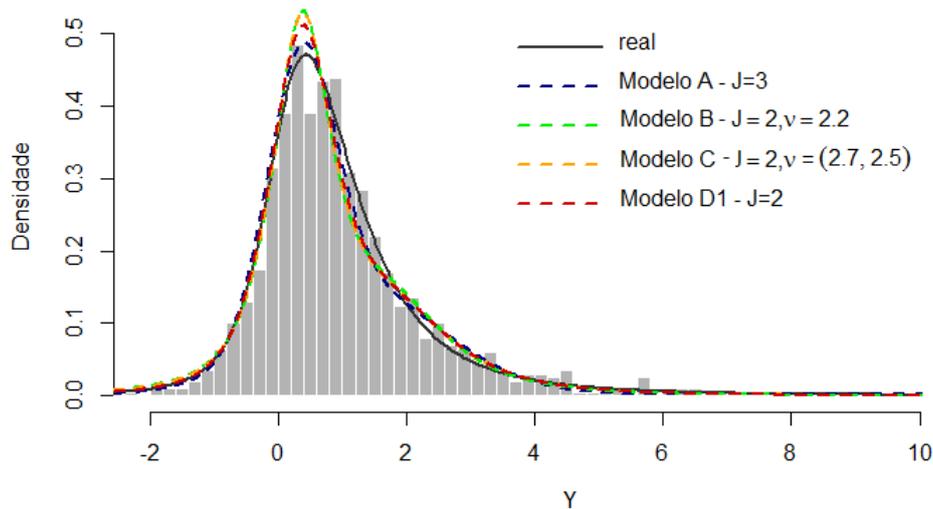


Figura C.13: Histograma para 1000 valores gerados a partir de uma Skew-t (0, 1, 1.5, 2.5) e seus respectivos ajustes em comparação à densidade real

2) Dados vindos de uma mistura de distribuições Skew-t

Foram geradas 1000 observações vindas de uma mistura de distribuições Skew-t com $\mu = (-1, 1)$, $\sigma^2 = (1, 0.81)$, $\nu = (2.8, 4)$, $\lambda = (-1.5, 0.8)$ e $w = (0.65, 0.35)$. Analogamente ao Capítulo 3, avaliamos o viés, variância e EMQ dos resultados *a posteriori* referentes a esperança e variância de \mathbf{Y} em cada um dos modelos considerado para ajuste.

Os modelos de mistura comparados neste estudo foram:

- A. Mistura de Normais com $J = 5$;
- B. Mistura de t's com $J = 2, 3$ e estimando $\nu_1 = \dots = \nu_J = \nu$;
- C. Mistura de t's com $J = 2, 3$ e estimando $\nu_1 \neq \dots \neq \nu_J$;
- D1. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.8, 4)$ fixo;
- D2. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.3, 5)$ fixo;
- D3. Modelo Proposto com $J = 2, 3$, $K = 2$ e $\nu = (2.2, 9)$ fixo.

Tabela C.11: viés, variância e erro quadrático médio para a esperança e variância de Y em diferentes distribuições de mistura quando os dados são gerados de uma mistura de skew-t.

Modelos	$\widehat{E(Y)}$			$\widehat{Var(Y)}$			\bar{D}	\bar{D}
	vício	var	EQM	vício	var	EQM		
<u>A</u> $J = 5$	0.0628	0.0019	0.0059	-0.1763	0.0825	0.1136	0.960	1.291
<u>B</u> $J = 2$	0.1299	0.0015	0.0184	9.1795	270.59	354.86	0.730	1.152
$J = 3$	0.1072	0.0020	0.0135	3.9279	135.26	150.69	1.067	1.373
<u>C</u> $J = 2$	0.1193	0.0017	0.0159	6.2979	161.79	201.46	0.596	0.921
$J = 3$	0.0879	0.0022	0.0099	-0.1939	0.9735	1.0111	0.776	0.966
<u>D1</u> $J = K = 2$	0.1242	0.0015	0.0169	-0.5207	0.0426	0.3137	0.359	0.493
$J = 3, K = 2$	0.0859	0.0017	0.0091	-0.0899	0.1553	0.1634	0.398	0.544
<u>D2</u> $J = K = 2$	0.1226	0.0016	0.0166	1.3471	0.1817	1.9965	0.537	0.812
$J = 3, K = 2$	0.0942	0.0018	0.0107	1.2610	0.4699	2.0601	0.649	0.946
<u>D3</u> $J = K = 2$	0.1206	0.0016	0.0162	2.9261	0.3528	8.9152	0.602	0.922
$J = 3, K = 2$	0.0974	0.0016	0.0112	2.7896	0.8098	8.5917	0.588	0.890

$E(Y) = -0.7181729$ (média amostral: -0.64330); $var(Y) = 4.94689$ (variância amostral: 4.68716)

O viés, variância e EQM calculados a partir dos resultados *a posteriori* para a $E(Y)$ foram similares em todos os modelos ajustados. Em relação aos resultados para a $Var(Y)$, observa-se que os menores EQM's são para os modelos A, D1 e D2, ressaltando, no entanto, que o modelo A considera $J = 5$. Em relação à medida de variação percentual, observa-se que os modelos D1 – D3 forneceram as menores distâncias em relação à curva real, sendo o modelo D1 o que obteve o menor valor dentre todos os modelos ajustados, tanto globalmente quanto nas caudas da distribuição. A Figura C.14 apresenta o histograma dos dados gerados com as curvas ajustadas baseadas no modelos que forneceram melhores ajustes.

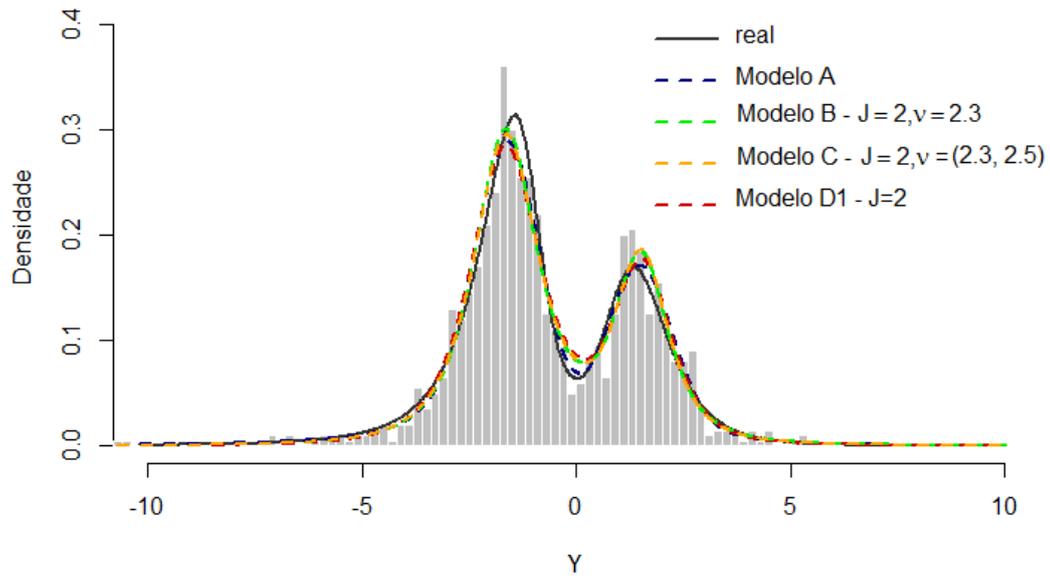
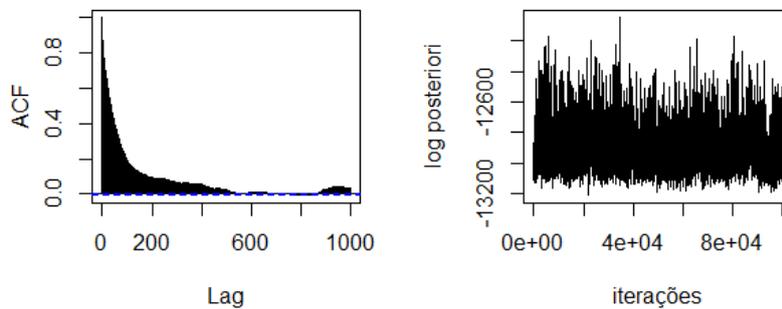


Figura C.14: Histograma para 1000 valores gerados a partir de uma mistura de Skew-t e seus respectivos ajustes em comparação à densidade real

Apêndice D

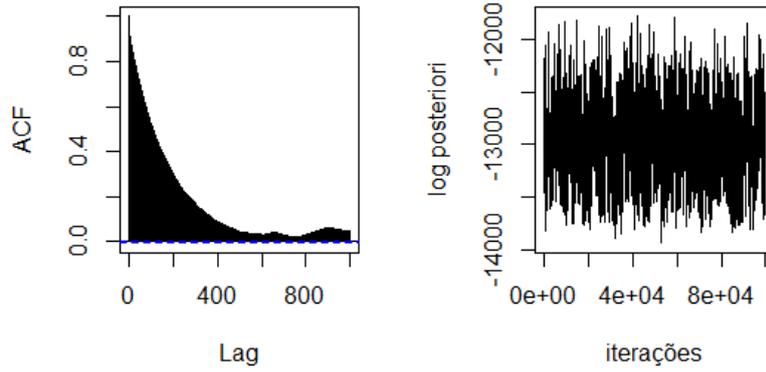
ACF e traços para as cadeias da *log* densidade *a posteriori*

Neste Apêndice encontram-se as funções de autocorrelação e traços para as cadeias da *log* densidade *a posteriori* para os modelos ajustados nos estudos simulados do Capítulo 3. Os traços para os cinco cenários apresentados sugerem que houve convergência das cadeias.



(a)

Figura D.1: ACF e traço para a cadeia da *log* densidade *a posteriori* para a amostra de tamanho 2500 gerada com $J = 1$ e $K = 2$ - modelo C.



(a)

Figura D.2: ACF e traço para a cadeia da *log* densidade *a posteriori* para a amostra de tamanho 2500 gerada com $J = 2$ e $K = 1$ - modelo B.

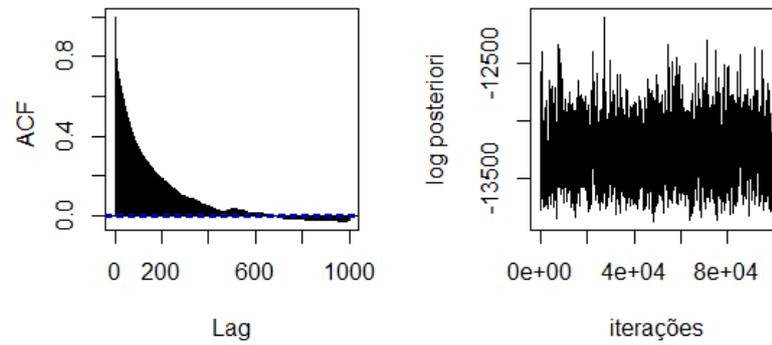


Figura D.3: ACF e traço para a cadeia da *log* densidade *a posteriori* para a amostra de tamanho 2500 gerada com $J = 2$ e $K = 2$ - modelo C.

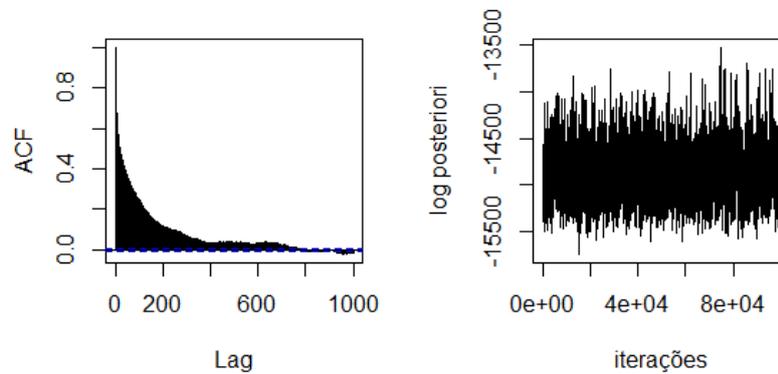


Figura D.4: ACF e traço para a cadeia da *log* densidade *a posteriori* para a amostra de tamanho 2500 gerada com $J = 2$ e $K = 2$ - modelo de regressão.

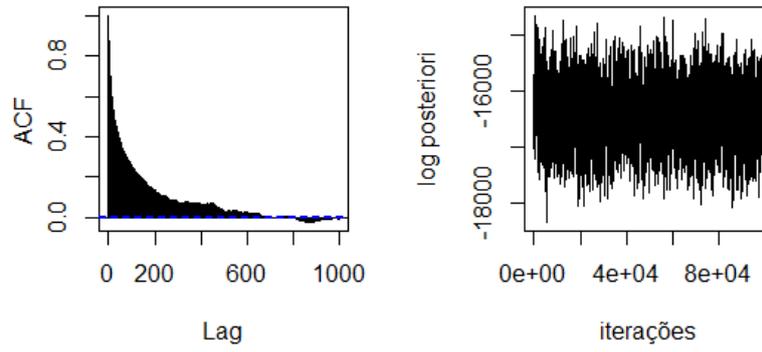


Figura D.5: ACF e traço para a cadeia da *log* densidade *a posteriori* para a amostra de tamanho 2500 gerada com $J = 3$ e $K = 2$ - modelo de regressão.

Referências Bibliográficas

ANDREWS, D. F.; MALLOWS, C. L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B - Methodological*, JSTOR, v. 36, n. 1, p. 99–102, 1974.

ARELLANO-VALLE, R. B.; AZZALINI, A. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 33, n. 3, p. 561–574, 2006.

AZZALINI, A. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 32, n. 2, p. 159–188, 2005.

AZZALINI, A.; GENTON, M. G. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, Wiley Online Library, v. 76, n. 1, p. 106–129, 2008.

BARNDORFF-NIELSEN, O. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, JSTOR, v. 5, n. 3, p. 151–157, 1978.

BARTOLUCCI, F.; SCACCIA, L. The use of mixtures for dealing with non-normal regression errors. *Computational Statistics & Data Analysis*, Elsevier, v. 48, n. 4, p. 821–834, 2005.

BASSO, R. M. et al. Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 12, p. 2926–2941, 2010.

BENITES, L.; MAEHARA, R.; LACHOS, V. H. Linear regression models with mixture of skew heavy-tailed errors. 2016. Disponível em: <<http://www.ime.unicamp.br/conteudo/linear-regression-models-finite-mixtures-skew-heavy-tailed-errors>>.

BÖHNING, D. *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*. [S.l.]: CRC press, 2000. v. 81.

BÖHNING, D. et al. The 2nd special issue on advances in mixture models. *Computational Statistics & Data Analysis*, v. 71, n. C, p. 1–2, 2014.

BÖHNING, D. et al. Advances in mixture models. *Computational Statistics & Data Analysis*, North-Holland, v. 51, n. 11, p. 5205–5210, 2007.

BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, Elsevier, v. 79, n. 1, p. 99–113, 2001.

- CABRAL, C. B.; BOLFARINE, H.; PEREIRA, J. R. G. Bayesian density estimation using skew student-t-normal mixtures. *Computational Statistics & Data Analysis*, Elsevier, v. 52, n. 12, p. 5075–5090, 2008.
- CABRAL, C. R. B.; LACHOS, V. H.; MADRUGA, M. R. Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *Journal of Statistical Planning and Inference*, Elsevier, v. 142, n. 1, p. 181–200, 2012.
- CABRAL, C. R. B.; LACHOS, V. H.; PRATES, M. O. Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 56, n. 1, p. 126–142, 2012.
- CAPPÉ, O.; ROBERT, C. P.; RYDÉN, T. Reversible jump, birth-and-death and more general continuous time markov chain monte carlo samplers. *Journal of the Royal Statistical Society: Series B - Statistical Methodology*, Wiley Online Library, v. 65, n. 3, p. 679–700, 2002.
- CARLIN, B. P.; CHIB, S. Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B - Methodological*, JSTOR, v. 57, n. 3, p. 473–484, 1995.
- CASTRO, L. et al. Partially linear censored regression models using heavy-tailed distributions: A bayesian approach. *Statistical Methodology*, Elsevier, v. 18, n. 3, p. 14–31, 2014.
- CHOY, S. B.; CHAN, J. S. Scale mixtures distributions in statistical modelling. *Australian & New Zealand Journal of Statistics*, Wiley Online Library, v. 50, n. 2, p. 135–146, 2008.
- CRAWFORD, S. An application of the laplace method to finite mixture distributions. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 89, n. 425, p. 259–267, 1994.
- CUTLER, A.; WINDHAM, M. P. Information-based validity functionals for mixture analysis. In: SPRINGER. *Proceedings of the first US/Japan Conference on the Frontiers of statistical modeling: An informational approach*. [S.l.], 1994. p. 149–170.
- DASGUPTA, A. *Asymptotic theory of statistics and probability*. [S.l.]: Springer Science & Business Media, New York, 2008.
- DEMPSTER, A.; LAIRD, N.; RUBIN, D. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B - Methodological*, v. 39, n. 1, p. 1–38, 1977.
- DIEBOLT, J.; ROBERT, C. P. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B - Methodological*, Royal Statistical Society, v. 56, n. 2, p. 363–375, 1994.
- EVERITT, B. Maximum likelihood estimation of the parameters in a mixture of two univariate normal distributions: A comparison of different algorithms. *Statistician*, **33**, n. 2, p. 205–215, 1984.

- FENG, Z. D.; MCCULLOCH, C. E. Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society. Series B - Methodological*, JSTOR, v. 58, n. 3, p. 609–617, 1996.
- FERNANDEZ, C.; STEEL, M. F. Bayesian regression analysis with scale mixtures of normals. *Econometric Theory*, Cambridge University Press, **16**, n. 1, p. 80–101, 2000.
- FERNANDEZ, C.; STEEL, M. F. J. Multivariate student-t regression models: Pitfalls and inference. *Biometrika*, Biometrika Trust, v. 86, n. 1, p. 153–167, 1999.
- FONSECA, T. C.; FERREIRA, M. A. R.; MIGON, H. S. Objective bayesian analysis for the student-t regression model. *Biometrika*, Oxford University Press, v. 95, n. 2, p. 325–333, 2008.
- FRÜHWIRTH-SCHNATTER, S. *Finite mixture and Markov switching models: Modeling and applications to random processes*. [S.l.]: Springer Science & Business Media, New York, 2006.
- FRÜHWIRTH-SCHNATTER, S.; PYNE, S. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, Biometrika Trust, v. 11, n. 2, p. 317–336, 2010.
- GALIMBERTI, G.; SOFFRITTI, G. A multivariate linear regression analysis using finite mixtures of t distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 71, n. 1, p. 138–150, 2014.
- GONÇALVES, F. B.; PRATES, M. O.; LACHOS, V. H. Robust bayesian model selection for heavy-tailed linear regression using finite mixtures. ArXiv preprint arXiv:1509.00331. 2015.
- GRÜN, B. bayesmim: Bayesian mixture models with jags. *R package version 0.7-2*, URL <http://CRAN.R-project.org/package=bayesmim>, 2011.
- HOLZMANN, H.; MUNK, A.; GNEITING, T. Identifiability of finite mixtures of elliptical distributions. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 33, n. 4, p. 753–763, 2006.
- JAMES, L. F.; PRIEBE, C. E.; MARCHETTE, D. J. Consistent estimation of mixture complexity. *The Annals of Statistics*, JSTOR, v. 29, n. 5, p. 1281–1296, 2001.
- JASRA, A.; HOLMES, C.; STEPHENS, D. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, JSTOR, v. 20, n. 1, p. 50–67, 2005.
- LANGE, K.; SINSHEIMER, J. S. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, Taylor & Francis Group, v. 2, n. 2, p. 175–198, 1993.
- LEE, J. E. *Bayesian hybrid algorithms and models: implementation and associated issues*. Tese (Doutorado) — Queensland University of Technology, 2010.
- LEE, K. et al. Bayesian inference on mixtures of distributions. ArXiv preprint arXiv:0804.2413. 2008.

- LEE, S.; MCLACHLAN, G. J. Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing*, Springer, v. 24, n. 2, p. 181–202, 2014.
- LEROUX, B. G. Consistent estimation of a mixing distribution. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 20, n. 3, p. 1350–1360, 1992.
- LIN, T.; LEE, J.; HSIEH, W. Robust mixture modeling using the skew-t distribution. *Statistics and Computing*, Springer, v. 17, n. 2, p. 81–92, 2007.
- LIN, T.; LEE, J.; YEN, S. Finite mixture modelling using the skew-normal distribution. *Statistica Sinica*, C/O DR HC HO, Statistical Science, Academia Sinica, TAIPEI 115, TAIWAN, v. 17, n. 3, p. 909–927, 2007.
- LIN, T. I.; LEE, J. C.; NI, H. F. Bayesian analysis of mixture modelling using the multivariate t distribution. *Statistics and Computing*, Springer, v. 14, n. 2, p. 119–130, 2004.
- LINDSAY, B. Mixture models: theory, geometry and applications. In: JSTOR. *NSF-CBMS regional conference series in probability and statistics*. [S.l.], 1995.
- LIU, C.; RUBIN, D. B. Ml estimation of the t distribution using em and its extensions, ecm and ecme. *Statistica Sinica*, v. 5, n. 1, p. 19–39, 1995.
- MAINDONALD, J. H.; BRAUN, W. J.; BRAUN, M. W. J. *Package DAAG*. [S.l.]: Version, 2015.
- MARIN, J.-M.; MENGENSEN, K.; ROBERT, C. P. Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics*, Elsevier-Sciences, v. 25, n. 16, p. 459–507, 2005.
- MARTINS, T. G. et al. Penalising model component complexity: A principled, practical approach to constructing priors. *arXiv preprint arXiv:1403.4630*, 2014.
- MAYRINK, V. D.; GONÇALVES, F. B. A bayesian hidden markov mixture model to detect overexpressed chromosome regions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 66, n. 2, p. 387–412, 2017.
- MCLACHLAN, G.; BASFORD, K. *Mixture models: Inference and applications to clustering*. [S.l.]: Marcel Dekker, 1988.
- MCLACHLAN, G. J. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Journal of the Royal Statistics Society, Series C - Applied Statistics*, JSTOR, v. 36, n. 3, p. 318–324, 1987.
- MCLACHLAN, G. J.; NG, S.-K.; BEAN, R. Robust cluster analysis via mixture models. *Austrian Journal of Statistics*, Osterreichische Statistische Gesellschaft, v. 35, n. 2, p. 157–174, 2006.
- MCLACHLAN, G. J.; PEEL, D. *Finite mixture models*. [S.l.]: John Wiley & Sons, New York, 2000.

- MENGERSEN, K.; ROBERT, C.; TITTERINGTON, M. *Mixtures: Estimation and applications*. [S.l.]: John Wiley & Sons, 2011. v. 896.
- MILOSLAVSKY, M.; LAAN, M. J. van der. Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational statistics & Data Analysis*, Elsevier, v. 41, n. 3, p. 413–428, 2003.
- NASCIMENTO, F. F. do; GAMERMAN, D.; LOPES, H. F. A semiparametric bayesian approach to extreme value estimation. *Statistics and Computing*, Springer, v. 22, n. 2, p. 661–675, 2012.
- NEWCOMB, S. A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, JSTOR, v. 8, n. 4, p. 343–366, 1886.
- NGUYEN, H. *Finite mixture models for regression problems*. Tese (Doutorado), 2015.
- NOBILE, A. *Bayesian analysis of finite mixture distributions*. Tese (Doutorado) — Carnegie Mellon University, 1994.
- NOBILE, A.; FEARNSIDE, A. T. Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing*, Springer, v. 17, n. 2, p. 147–162, 2007.
- PEARSON, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, v. 185, p. 71–110, 1894.
- PEEL, D.; MCLACHLAN, G. J. Robust mixture modelling using the t distribution. *Statistics and Computing*, Springer, v. 10, n. 4, p. 339–348, 2000.
- PRATES, M. O.; LACHOS, V. H.; CABRAL, C. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, Foundation for Open Access Statistics, v. 54, n. 12, p. 1–20, 2013.
- PRUIM, R. *NHANES: Data from the US National Health and Nutrition Examination Study*. [S.l.]: The R Foundation for Statistical Computing. Available at: <https://cran.r-project.org/package=NHANES>. [rDN], 2015.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: [<https://www.R-project.org/>](https://www.R-project.org/).
- REDNER, R. A.; WALKER, H. Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, SIAM, v. 26, n. 2, p. 195–239, 1984.
- RICHARDSON, S.; GREEN, P. J. On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society. Series B - Methodological*, JSTOR, v. 59, n. 4, p. 731–792, 1997.
- RIPLEY, B. et al. Package mass. 2013.

ROEDER, K.; WASSERMAN, L. Practical bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 92, n. 439, p. 894–902, 1997.

SOFFRITTI, G.; GALIMBERTI, G. Multivariate linear regression with non-normal errors: A solution based on mixture models. *Statistics and Computing*, Springer, v. 21, n. 4, p. 523–536, 2011.

SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002.

STEPHENS, M. *Bayesian methods for mixtures of normal distributions*. Tese (Doutorado) — University of Oxford, 1997.

STEPHENS, M. Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Annals of Statistics*, Institute of Mathematical Statistics, v. 28, n. 1, p. 40–74, 2000.

STEPHENS, M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B - Statistical Methodology*, Wiley Online Library, v. 62, n. 4, p. 795–809, 2000.

SUSKO, E. Weighted tests of homogeneity for testing the number of components in a mixture. *Computational Statistics & Data Analysis*, Elsevier, v. 41, n. 3, p. 367–378, 2003.

SVENSÉN, M.; BISHOP, C. M. Robust bayesian mixture modelling. *Neurocomputing*, Elsevier, v. 64, n. 1, p. 235–252, 2005.

TEICHER, H. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 34, n. 4, p. 1265–1269, 1963.

TITTERINGTON, D.; SMITH, A.; MAKOV, U. *Statistical analysis of finite mixture distributions*. [S.l.]: John Wiley, 1985.

VALLEJOS, C. A.; STEEL, M. F. On posterior propriety for the student t linear regression model under jeffreys priors. *arXiv preprint arXiv:1311.1454*, 2013.

VEAUX, R. D. D.; KRIEGER, A. M. Robust estimation of a normal mixture. *Statistics & Probability Letters*, Elsevier, v. 10, n. 1, p. 1–7, 1990.

VILLA, C.; WALKER, S. G. et al. Objective prior for the number of degrees of freedom of at distribution. *Bayesian Analysis*, International Society for Bayesian Analysis, v. 9, n. 1, p. 197–220, 2014.

WANG, B.; TITTERINGTON, D. Local convergence of variational bayes estimators for mixing coefficients. 2003.

WANG, H. X. et al. Estimation for the number of components in a mixture model using stepwise split-and-merge em algorithm. *Pattern Recognition Letters*, Elsevier, v. 25, n. 16, p. 1799–1809, 2004.

WEISBERG, S. *Applied linear regression*. 3rd. ed. [S.l.]: John Wiley & Sons, New Jersey, 2005.

WIPER, M.; INSUA, D. R.; RUGGERI, F. Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 10, n. 3, p. 440–454, 2001.

YAKOWITZ, S. J.; SPRAGINS, J. D. On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 39, n. 1, p. 209–214, 1968.

ZHANG, M.-H.; CHENG, Q.-S. Determine the number of components in a mixture model by the extended ks test. *Pattern Recognition Letters*, Elsevier, v. 25, n. 2, p. 211–216, 2004.