



UFAM

MISTURAS DE MODELOS DE REGRESSÃO LINEAR COM ERROS NAS
VARIÁVEIS USANDO MISTURAS DE ESCALA DA NORMAL ASSIMÉTRICA

Renata Evangelista Monteiro

Dissertação de Mestrado apresentada ao
Programa de Pós-graduação em Matemática,
da Universidade Federal do Amazonas, como
parte dos requisitos necessários à obtenção do
título de Mestre em Matemática

Orientador: Dr. Celso Rômulo Barbosa Cabral

Manaus

Março de 2018

MISTURAS DE MODELOS DE REGRESSÃO LINEAR COM ERROS NAS
VARIÁVEIS USANDO MISTURAS DE ESCALA DA NORMAL ASSIMÉTRICA

Renata Evangelista Monteiro

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
PÓS-GRADUAÇÃO EM MATEMÁTICA, DA UNIVERSIDADE FEDERAL DO
AMAZONAS, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM MATEMÁTICA.

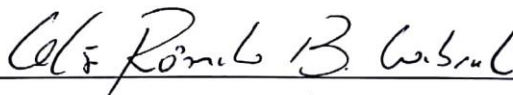
Examinada por:



Prof^a. Larissa Avila Matos, D.Sc.



Prof. José Raimundo Gomes Pereira, D.Sc.



Prof. Celso Rômulo Barbosa Cabral, D.Sc.

MANAUS, AM – BRASIL

MARÇO DE 2018

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

M775m Monteiro, Renata Evangelista
Misturas de Modelos de Regressão Linear com Erros nas Variáveis usando Misturas de Escala da Normal Assimétrica / Renata Evangelista Monteiro. 2018
64 f.: il. color; 31 cm.

Orientador: Celso Rômulo Barbosa Cabral
Dissertação (Mestrado em Matemática - Estatística) - Universidade Federal do Amazonas.

1. Distribuição normal assimétrica. 2. Mistura de escala normal assimétrica. 3. Algoritmo MCMC. 4. Mistura de modelos de regressão. 5. Modelo com erro nas covariáveis. I. Cabral, Celso Rômulo Barbosa II. Universidade Federal do Amazonas III. Título

*Esta dissertação é dedicada ao
grande amor da minha vida, meu
filho Gabriel.*

Agradecimentos

Primeiramente a Deus, pelo seu infinito amor e pelo conforto nos momentos difíceis da vida.

Ao professor Celso Rômulo por sua excelente orientação, por sua infinita paciência para comigo e por compartilhar sua sabedoria que vai além da sala de aula.

Ao meu pai, Alberto Monteiro, que sempre me incentivou a estudar.

Ao meu marido Daniel, que tem se mostrado, a cada dia, um grande companheiro e incentivador. Ao meu amado filho Gabriel que entende (na medida do possível) a ausência mesmo que presente da sua mãe.

As minhas amigas de mestrado Regina, Alice, Sarah e Milena pelo companheirismo nessa jornada.

Aos amigos, Alex, Érico, Jhonata, Ciço, Thiago, Vinícius e Natan que deram um toque de humor aos dias acadêmicos.

Aos amigos (as) do tempo da graduação, Carla Zeline, Nelson, Camila, Márcia, Vanessa e Carina pelas palavras de incentivo.

Ao meu amigo Guilherme Peña por ter contribuído neste trabalho, sua ajuda foi valiosíssima.

Aos professores do Departamento de Estatística pelos ensinamentos ao longo da vida acadêmica. Em especial, aos professores José Raimundo, José Cardoso, James Dean, Max pelas críticas construtivas, sempre norteando o caminho do conhecimento.

A Dona Neisa, que sempre tinha um café quentinho para oferecer.

À CAPES, (Coordenação de Aperfeiçoamento Pessoal de Nível Superior) pela assistência financeira ao PPGMAT.

Cada pessoa tem sua parcela no desenvolvimento neste trabalho. Sozinha, eu seria um nada, ninguém. Obrigada por tudo.

"O êxito da vida não se mede pelo caminho que você conquistou, mas sim pelas dificuldades que superou no caminho. "

Abraham Lincoln

Resumo da Dissertação apresentada ao Programa de Pós-Graduação em Matemática, da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do grau de Mestre em Matemática. (M.Sc.)

MISTURAS DE MODELOS DE REGRESSÃO LINEAR COM ERROS NAS
VARIÁVEIS USANDO MISTURAS DE ESCALA DA NORMAL ASSIMÉTRICA

Renata Evangelista Monteiro

Março/2018

Orientador: Dr. Celso Rômulo Barbosa Cabral

Linha de Pesquisa: Estatística

A estimação tradicional em mistura de modelos de regressão é baseada na suposição de normalidade para os erros aleatórios, sendo assim, sensível a outliers, caudas pesadas e erros assimétricos. Outra desvantagem é que, em geral, a análise é restrita a preditores que são observados diretamente.

Apresentamos uma proposta para lidar com estas questões simultaneamente no contexto de mistura de regressões estendendo o modelo normal clássico. Assumimos que, conjuntamente e em cada componente da mistura, os erros aleatórios e as covariáveis seguem uma mistura de escala da distribuição normal assimétrica. Além disso, é feita a suposição de que as covariáveis são observadas com erro aditivo.

Um algoritmo do tipo MCMC foi desenvolvido para realizar inferência Bayesiana. A eficácia do modelo proposto é verificada via análises de dados simulados e reais.

Palavras-chave: Distribuição normal assimétrica; Algoritmo MCMC; Mistura de escala normal assimétrica; Mistura de modelos de regressão; Modelo com erro nas covariáveis.

Abstract of Dissertation presented to Postgraduate in Mathematics, of the Federal University of Amazonas, as a partial fulfillment of the requirements for the degree of Master of Mathematics. (M.Sc.)

MIXTURES OF LINEAR REGRESSION MODELS WITH MEASUREMENT
ERRORS USING SCALE MIXTURES OF SKEW-NORMAL DISTRIBUTIONS

Renata Evangelista Monteiro

March/2018

Advisor: Dr. Celso Rômulo Barbosa Cabral

Research lines: Statistics

The traditional estimation of mixture regression models is based on the assumption of normality of component errors and thus is sensitive to outliers, heavy-tailed and/or asymmetric errors. Another drawback is that, in general, the analysis is restricted to directly observed predictors.

We present a proposal to deal with these issues simultaneously in the context of mixture regression by extending the classic normal model by assuming that, for each mixture component, the random errors and the covariates jointly follow a scale mixture of skew-normal distributions. It is also assumed that the covariates are observed with error.

An MCMC-type algorithm to perform Bayesian inference is developed and, in order to show the efficacy of the proposed methods, simulated and real data sets are analyzed.

Keywords: Skew-normal distribution; MCMC algorithm; Scale mixtures of skew-normal; Mixture of regression models; Measurement error model.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xii
1 Introdução	1
1.1 Organização do Trabalho	3
2 Distribuição Normal Assimétrica e Distribuições Misturas de Escala da Normal Assimétrica	4
2.1 A Família SMSN	4
2.2 Distribuição t de Student Assimétrica	7
2.3 Distribuição Slash Assimétrica	8
2.4 Distribuição Normal Contaminada Assimétrica	8
2.5 Uma Representação Hierárquica	9
3 O Modelo Proposto	10
3.1 Uma Extensão do Modelo de Regressão Linear com Erros nas Variáveis .	10
3.2 Estimação a Posteriori	15
3.2.1 Distribuição a Priori	15
3.2.2 Algoritmo Tipo Gibbs	17
3.3 Critérios de Seleção de Modelos	22
3.3.1 Critério de Informação do Desvio Observado	22
3.3.2 Critério de Informação Watanabe-Akaike	24
4 Simulação e Aplicação em Dados Reais	26
4.1 Simulação	26
4.2 Aplicação em Dados Reais	31

5	Considerações Finais	38
6	Apêndice A	39
6.1	Detalhes do Algoritmo Gibbs	39
	Referências Bibliográficas	47

Lista de Figuras

4.1	<i>Conjunto de dados simulados. (a) Diagrama de dispersão e (b) Histograma da variável resposta.</i>	27
4.2	<i>RC (%) para as estimativas de α_1, α_2, β_1, β_2, variando em Λ.</i>	28
4.3	<i>RC (%) para as estimativas de σ_1^2, σ_2^2 e para os pesos p_1, p_2, variando em Λ.</i>	29
4.4	<i>Dic observado para os ajustes FM-SMSNME, variando em Λ.</i>	30
4.5	<i>WAIC para os ajustes FM-SMSNME, variando em Λ.</i>	30
4.6	<i>Conjunto de dados percepção de tom. (a) Diagrama de dispersão e (b) Histograma.</i>	31
4.7	<i>Histograma do erro aleatório com distribuição NIG.</i>	32
4.8	<i>Traceplots das amostras MCMC no ajuste FM-STME.</i>	33
4.9	<i>Traceplots das amostras MCMC no ajuste FM-SSLME.</i>	34
4.10	<i>Traceplots das amostras MCMC no ajuste FM-SCNME.</i>	34
4.11	<i>Dados percepção de tons. Diagramas de dispersão com os modelos FM-SNME, FM-STME, FM-SSLME e FM-SCNME ajustados.</i>	36
4.12	<i>Dados percepção de tons. Diagramas de dispersão com os modelos FM-NME, FM-TME, FM-SLME e FM-SCNME ajustados.</i>	37

Lista de Tabelas

3.1	Siglas dos modelos com membros diferentes da família <i>SMSN</i>	13
4.1	Estimativas MCMC para os parâmetros nos ajustes dos modelos <i>FM-SMSNME</i> para o conjunto de dados Cohen (1984).	35
4.2	Critérios de seleção para os modelos <i>FM-SMSNME</i> para o conjunto de dados Cohen (1984). A sigla LV denota log-verossimilhança.	36
4.3	Intervalos de credibilidade (IC) nos ajustes <i>FM-SSLME</i> e <i>FM-SCNME</i> para o conjunto de dados Cohen (1984), com 95% de credibilidade. . . .	37

Capítulo 1

Introdução

A modelagem baseada em mistura finita de distribuições é um ramo atraente da pesquisa científica, com uma gama crescente de aplicações em diversas áreas do conhecimento, como biologia, biometria, ciências agrárias, marketing, medicina, economia, entre outras. Sua importância pode ser observada a partir de inúmeros livros publicados na área, como Lindsay (1995), Böhning (2000), McLachlan & Peel (2000), Frühwirth-Schnatter (2006) e Mengersen *et al.* (2011), e a edição especial do periódico *Computational Statistics and Data Analysis* Böhning *et al.* (2007, 2014). Esses modelos são extremamente flexíveis, pois permitem capturar muitas características específicas dos dados reais, como multimodalidade, assimetria, curtose e heterogeneidade não observada.

A mistura finita de modelos de regressão é definida por:

$$Y_i | \omega_i = j \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}_j, \sigma_j^2) \quad (1.1)$$

$$P(\omega_i = j) = p_j, \quad j = 1, \dots, G, \quad i = 1, \dots, n,$$

em que $N(\mathbf{x}_i^\top \boldsymbol{\beta}_j, \sigma_j^2)$ denota a distribuição normal com média $\mathbf{x}_i^\top \boldsymbol{\beta}_j$ e variância σ_j^2 , $\boldsymbol{\beta}_j$ é um vetor p -dimensional dos coeficientes de regressão, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i(p-1)})^\top$ é o vetor de variáveis explicativas de dimensão $p \times 1$ e ω_i é uma variável latente que é, de fato, uma variável de classificação, se conhecermos o seu valor, saberemos como discriminar o indivíduo i entre as G subpopulações heterogêneas. Cada subpopulação tem um vetor específico de coeficientes ($\boldsymbol{\beta}_j$) e variância (σ_j^2). Usando (1.1) e integrando em relação a ω_i , obtemos a distribuição marginal de Y_i , que é uma mistura finita de G modelos de

regressão com densidade:

$$f(y_i|\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_G, \sigma_1^2, \dots, \sigma_G^2) = \sum_{j=1}^G p_j N(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}_j, \sigma_j^2), \quad (1.2)$$

em que $N(\cdot|\mathbf{x}_i^\top \boldsymbol{\beta}_j, \sigma_j^2)$ denota a densidade da distribuição normal com média $\mathbf{x}_i^\top \boldsymbol{\beta}_j$ e variância σ_j^2 sendo denominado pelo j -ésimo componente da mistura.

Existem várias trabalhos com aplicações do modelo (1.2), incluindo pesquisas em marketing (Quandt & Ramsey, 1978; DeSarbo & Cron, 1988; DeSarbo *et al.*, 1992), economia (Cosslett & Lee, 1985; Hamilton, 1989), agricultura (Turner (2000)), nutrição (Arellano-Valle *et al.* (2008)) e em psicometria (Liu *et al.* (2011)).

Muitas extensões deste modelo clássico foram propostas para aumentar a aplicabilidade em análise de regressão linear. Há situações em que assumir normalidade para os erros é inadequado, por exemplo, para conjuntos de dados que provém de distribuições assimétricas ou com caudas pesadas. Bai *et al.* (2012) propôs a modificação do algoritmo EM para mistura de normais, através do critério de substituição dos mínimos quadrados na etapa M por um critério robusto, como a função Ψ de Huber (1981). Por meio de um estudo de simulação, eles mostraram que a estimativa proposta é robusta quando os dados tem valores atípicos ou a distribuição dos erros tem caudas pesadas. Song *et al.* (2014) propuseram um procedimento de estimação robusta para mistura de modelos de regressão linear, assumindo que o erro tem distribuição de Laplace.

Baseado em Hennig (2004), Yao *et al.* (2014) argumentam que o modelo de misturas de regressão t de Student é sensível a pontos de alta influência. Para contornar este problema, os autores propõem uma versão do modelo de misturas de regressão t, ajustando este modelo aos dados após "thinning" dos pontos de alta influência. Usando a distribuição normal assimétrica definida por Azzalini (1985), Liu & Lin (2014) propuseram uma versão em mistura de normais em modelos de regressão, que parece ser uma ferramenta de modelagem teoricamente mais atraente para aplicações práticas, pois parece acomodar de forma adequada erros assimétricos. Zeller *et al.* (2016) estenderam o modelo normal, assumindo que os erros aleatórios seguem uma mistura de escala da distribuição normal assimétrica. Estes autores desenvolveram um algoritmo do tipo EM para fazer inferência por máxima verossimilhança dos parâmetros no modelo proposto.

Os trabalhos citados anteriormente estimam os parâmetros na mistura de modelos

de regressão por máxima verossimilhança. Que seja do nosso conhecimento, pouco esforço tem sido feito para tratar o problema sob a perspectiva Bayesiana. Alguns exemplos são os trabalhos de Viele & Tong (2002) e Hurn *et al.* (2003). Em geral, os trabalhos frequentistas são restritos ao caso em que as covariáveis são observadas sem erro. Até agora, apenas o trabalho de Yao & Song (2014) lida com erros aditivos tanto na variável resposta como nos preditores.

Neste trabalho pretendemos propor uma mistura finita de modelos de regressão, onde as covariáveis são observadas com erro aditivo. Além disso, apresentaremos uma extensão do modelo clássico, supondo que a distribuição conjunta do erros observados e do vetor de covariáveis pertence à classe de distribuições das misturas de escalas da normal assimétrica, o que nos permite lidar simultaneamente com a heterogeneidade latente, assimetria e curtose.

1.1 Organização do Trabalho

Esta dissertação está dividida em seis capítulos. A revisão de literatura que versa sobre mistura de modelos de regressão é descrita neste Capítulo 1. No Capítulo 2 apresentamos as construções da distribuição normal assimétrica e das distribuições misturas de escala da normal assimétrica.

No Capítulo 3 apresentamos a proposta do modelo, que trata sobre uma extensão do modelo de mistura de regressão linear com adição de erros na variável resposta e nas variáveis regressoras. O desenvolvimento do algoritmo Gibbs via inferência Bayesiana para estimação dos parâmetros também são descritos neste capítulo.

No Capítulo 4 apresentamos os resultados obtidos por meio de estudos de simulações e aplicação em dados reais. As considerações finais são apresentadas no Capítulo 5.

No Apêndice A são apresentados os detalhes do algoritmo Gibbs.

Capítulo 2

Distribuição Normal Assimétrica e Distribuições Misturas de Escala da Normal Assimétrica

2.1 A Família SMSN

Os conceitos de distribuição normal assimétrica (*skew-normal ou SN*) e distribuições mistura de escala da normal assimétrica (*scale mixtures of skew-normal distributions ou SMSN*) são cruciais no desenvolvimento da nossa teoria. Neste capítulo apresentamos uma breve introdução destas distribuições.

A distribuição normal assimétrica é uma extensão da distribuição normal. Isso ocorre através da adição de um parâmetro, com finalidade de regular a assimetria. Algumas versões da normal assimétrica são cuidadosamente apresentadas em trabalhos como Azzalini (2005) e Arellano-Valle & Azzalini (2006), veja também o livro editado por Genton (2004) e o trabalho recente de Azzalini & Capitanio (2014).

Aqui, usamos a versão da normal assimétrica utilizada por Pyne *et al.* (2009). No que segue, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é um vetor aleatório, p -dimensional com distribuição normal com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$ (definida positiva e de dimensão $p \times p$). Vamos definir a distribuição normal assimétrica por construção.

Definição 1. *Sejam $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $X_0 \sim N(0, 1)$ independentes. Seja $\boldsymbol{\Delta}$ um vetor p -*

dimensional de constantes. Então dizemos que a distribuição de

$$\mathbf{Y} = \mathbf{X} + \Delta|X_0|, \quad (2.1)$$

é normal assimétrica com vetor de locação $\boldsymbol{\mu}$, matriz de escala $\boldsymbol{\Sigma}$ e vetor de forma Δ . Usamos a notação: $\mathbf{Y} \sim \text{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Delta)$.

Proposição 1. Dizemos que $\mathbf{Y} \sim \text{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Delta)$ se sua densidade é dada por

$$\text{SN}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Delta) = 2 N_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega})\Phi(\boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu})),$$

em que,

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \Delta\Delta^\top, \text{ e } \boldsymbol{\lambda} = \frac{\boldsymbol{\Omega}^{-1}\Delta}{(1 - \Delta^\top\boldsymbol{\Omega}^{-1}\Delta)^{1/2}}. \quad (2.2)$$

Demonstração. Vamos definir o vetor aleatório

$$\begin{pmatrix} T_0 \\ \mathbf{T}_1 \end{pmatrix} = \begin{pmatrix} X_0 \\ \mathbf{X} + \Delta X_0 \end{pmatrix},$$

onde X_0 e \mathbf{X} são dados na Definição 1. Pelo Teorema 3.1 em Arellano-Valle *et al.* (2002), $(\mathbf{T}_1|T_0 > 0)$ tem a mesma distribuição de $\mathbf{Y} = \mathbf{X} + \Delta|X_0|$. Assim, basta encontrar a distribuição de $(\mathbf{T}_1|T_0 > 0)$, que é $f(\mathbf{y}) = P(T_0 > 0)^{-1}P(T_0 > 0|\mathbf{T}_1 = \mathbf{y})f_{\mathbf{T}_1}(\mathbf{y})$. Observe que

$$\begin{pmatrix} T_0 \\ \mathbf{T}_1 \end{pmatrix} \sim N_{1+p} \left[\begin{pmatrix} 0 \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} 1 & \Delta^\top \\ \Delta & \boldsymbol{\Sigma} + \Delta\Delta^\top \end{pmatrix} \right],$$

que, pela teoria usual de condicionamento em distribuições normais multivariadas, implica $(T_0|\mathbf{T}_1 = \mathbf{y}) \sim N(\Delta^\top\boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}), 1 - \Delta^\top\boldsymbol{\Omega}^{-1}\Delta)$. O resultado segue imediatamente. \square

É importante notar que a SN dada na Definição 1 é a mesma utilizada em trabalhos anteriores, como Lachos *et al.* (2008), Lachos *et al.* (2010), Cabral *et al.* (2012a), Cabral *et al.* (2012b) e Cabral *et al.* (2014), definida por

$$\text{SN}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Delta) = 2 N_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega})\Phi(\boldsymbol{\lambda}^{*\top}\boldsymbol{\Omega}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))$$

em que $\boldsymbol{\lambda}^* = \boldsymbol{\Omega}^{-1/2}\Delta/(1 - \Delta^\top\boldsymbol{\Omega}^{-1}\Delta)^{1/2}$ e $\boldsymbol{\Omega}^{-1/2}$ é a inversa da raiz quadrada de $\boldsymbol{\Omega}$. Para

obter a versão usada nestes trabalhos, basta definir a parametrização $\boldsymbol{\lambda} = \boldsymbol{\Omega}^{-1/2} \boldsymbol{\lambda}^*$.

Definindo $T = |X_0|$, podemos observar em (2.1) que $\mathbf{Y} \sim \text{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ admite a seguinte representação estocástica

$$\begin{aligned} \mathbf{Y}|T = t &\sim N_p(\boldsymbol{\mu} + \boldsymbol{\Delta}t, \boldsymbol{\Sigma}); \\ T &\sim \text{TN}(0, 1, (0, \infty)), \end{aligned}$$

em que, $\text{TN}(\xi, \sigma^2, (a, b))$ denota a distribuição normal truncada no intervalo (a, b) , onde ξ e σ^2 são média e variância, respectivamente.

A próxima definição, trata sobre distribuições misturas de escala da normal assimétrica (SMSN), definida por Branco & Dey (2001).

Definição 2. Dizemos que a distribuição do vetor aleatório \mathbf{Y} de dimensão p , pertence à família de misturas de escala da normal assimétrica, quando sua representação estocástica é dada por

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2} \mathbf{X}, \quad (2.3)$$

onde $\boldsymbol{\mu}$ é um vetor p -dimensional de constantes, $\mathbf{X} \sim \text{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$, U é uma variável aleatória positiva (com probabilidade 1), independente de \mathbf{X} e com função de distribuição $H(\cdot|\boldsymbol{\nu})$.

Temos que, $\boldsymbol{\nu}$ é um parâmetro (possivelmente multivariado) indexando a distribuição de U , uma variável aleatória denominada fator de escala e a distribuição $H(\cdot|\boldsymbol{\nu})$ é conhecida por distribuição da variável aleatória de mistura. Usamos a notação $\mathbf{Y} \sim \text{SMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \boldsymbol{\nu})$.

Agora, observe que se $\mathbf{Y} \sim \text{SMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \boldsymbol{\nu})$ tem representação dada em (2.3), então substituindo a distribuição de \mathbf{X} por (2.1), temos

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\mu} + U^{-1/2} \mathbf{X} \\ &= \boldsymbol{\mu} + U^{-1/2} (\mathbf{W} + \boldsymbol{\Delta}|W_0|) \\ &= \boldsymbol{\mu} + U^{-1/2} \mathbf{W} + U^{-1/2} \boldsymbol{\Delta}|W_0|, \end{aligned}$$

onde $\mathbf{W} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, $W_0 \sim N(0, 1)$ e $U \sim H(\cdot|\boldsymbol{\nu})$ são independentes. Dado que $U = u$, então $u^{-1/2} \mathbf{W} \sim N_p(\mathbf{0}, u^{-1} \boldsymbol{\Sigma})$, isto é, quando $\mathbf{Y}|U = u$ temos exatamente a forma apre-

sentada em (2.1) da Definição 1. Assim,

$$\mathbf{Y}|U = u \sim \text{SN}_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, u^{-1/2}\boldsymbol{\Delta}), \quad (2.4)$$

o que significa dizer que a distribuição marginal de \mathbf{Y} é

$$\text{SMSN}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \boldsymbol{\nu}) = 2 \int_0^\infty \text{N}_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Omega})\Phi(u^{1/2}\boldsymbol{\lambda}'(\mathbf{y} - \boldsymbol{\mu}))dH(u|\boldsymbol{\nu}), \quad (2.5)$$

em que, $\boldsymbol{\Omega}$ e $\boldsymbol{\lambda}$ são dados em (2.2).

A família *SMSN* inclui a classe das misturas de escala da normal (*SMN*) definida por Andrews & Mallows (1974), veja também Lange & Sinsheimer (1993), onde é assumido normalidade para \mathbf{X} em (2.3) e corresponde ao caso $\boldsymbol{\Delta} = \mathbf{0}$. Neste caso usamos as notações $\mathbf{Y} \sim \text{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ e $\text{SMN}_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ para denotar a distribuição e densidade de \mathbf{Y} , respectivamente.

Dependendo da distribuição do fator de escala U , temos um membro diferente da família *SMSN*, como normal assimétrica, t de Student assimétrica, slash assimétrica a normal contaminada assimétrica e as suas respectivas versões simétricas. Tais distribuições são apresentadas a seguir.

2.2 Distribuição t de Student Assimétrica

A distribuição t de Student assimétrica (ou simplesmente t assimétrica) com $\nu > 0$ graus de liberdade, é obtida a partir de (2.5) considerando $U \sim \text{Gamma}(\nu/2, \nu/2)$ com densidade

$$\text{ST}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)\text{T} \left[\left(\frac{\nu + p}{\nu + d} \right)^{1/2} \boldsymbol{\lambda}'(\mathbf{y} - \boldsymbol{\mu})|\nu + p \right], \quad \mathbf{y} \in \mathbb{R}^p,$$

sendo $d = (\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ a distância de Mahalanobis ao quadrado, onde $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ e $\text{T}(\cdot|\nu + p)$ denotam, respectivamente, a função densidade t de Student p -variada, com vetor de locação $\boldsymbol{\mu}$, matriz de escala $\boldsymbol{\Omega}$ e ν graus de liberdade e a função de distribuição t de Student univariada padronizada com $\nu + p$ graus de liberdade, $\boldsymbol{\Omega}$ e $\boldsymbol{\lambda}$ são dados em (2.2). Para uma demonstração ver Branco & Dey (2001, sec 3.2).

Um caso particular da t assimétrica é a distribuição Cauchy assimétrica, quando

$\nu = 1$. Além disso quando $\nu \rightarrow \infty$, obtemos a distribuição normal assimétrica no limite. Aplicações da distribuição t assimétrica podem ser vistas em trabalhos como Lin *et al.* (2007) e Lee & McLachlan (2014).

2.3 Distribuição Slash Assimétrica

A distribuição slash assimétrica é obtida quando $U \sim \text{Beta}(\nu, 1)$. A sua densidade não tem forma fechada e é dada por

$$\text{SSL}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu) = 2\nu \int_0^1 u^{\nu-1} \text{N}_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Omega}) \Phi(u^{1/2}\boldsymbol{\lambda}'(\mathbf{y} - \boldsymbol{\mu})) du, \quad \mathbf{y} \in \mathbb{R}^p,$$

em que, $\boldsymbol{\Omega}$ e $\boldsymbol{\lambda}$ são dados em (2.2). Esta distribuição converge para a distribuição normal assimétrica quando $\nu \rightarrow \infty$. As suas propriedades foram exaustivamente estudadas por Wang & Genton (2006).

2.4 Distribuição Normal Contaminada Assimétrica

Esta distribuição é obtida quando U é uma variável aleatória discreta binária, com função de probabilidade

$$h(u; \boldsymbol{\nu}) = \rho \mathbb{I}_{(u=\tau)} + (1 - \rho) \mathbb{I}_{(u=1)}, \quad 0 < \nu < 1 \quad e \quad 0 < \tau < 1,$$

onde $\boldsymbol{\nu} = (\rho, \tau)'$. Segue de (2.5) que a densidade associada é dada por

$$\text{SCN}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \tau) = 2 \left\{ \rho \text{N}_p(\mathbf{y}|\boldsymbol{\mu}, \tau^{-1}\boldsymbol{\Omega}) \Phi(\tau^{1/2}\boldsymbol{\lambda}'(\mathbf{y} - \boldsymbol{\mu})) + (1 - \rho) \text{N}_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}) \Phi(\boldsymbol{\lambda}'(\mathbf{y} - \boldsymbol{\mu})) \right\}.$$

Os parâmetros ρ e τ podem ser interpretados como a proporção de outliers e fator de escala, respectivamente. Temos a normal assimétrica quando $\rho = \tau = 1$.

2.5 Uma Representação Hierárquica

Utilizando as representações (2.1) e (2.3), obtemos a seguinte representação hierárquica para $\mathbf{Y} \sim \text{SMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{v})$, dada por

$$\begin{aligned} \mathbf{Y}_i | U_i = u_i, T_i = t_i &\sim \text{N}_p(\boldsymbol{\mu} + \boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Sigma}) \\ T_i | U_i = u_i &\sim \text{TN}(0, u_i^{-1}, (0, \infty)) \\ U_i &\sim H(\cdot | \mathbf{v}). \end{aligned} \quad (2.6)$$

Essa representação é útil para gerar amostras artificiais de uma distribuição SMSN e para obter um algoritmo MCMC para fazer inferência a posteriori para os parâmetros do modelo proposto (apresentado no capítulo seguinte).

Uma aplicação importante dessa representação é que transformações afins de uma distribuição SMSN ainda estão na família SMSN, conforme indicado na seguinte proposição.

Proposição 2. *Seja $\mathbf{A} : q \times p$, $\mathbf{b} : q \times 1$ e $\mathbf{Y} \sim \text{SMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{v})$. Então $\mathbf{Z} = \mathbf{A}\mathbf{Y} + \mathbf{b} \sim \text{SMSN}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}', \mathbf{A}\boldsymbol{\Delta}, \mathbf{v})$.*

Prova. A representação hierárquica de \mathbf{Z} é obtida substituindo (2.6) com

$$\mathbf{Z} | U = u, T = t \sim \text{N}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\Delta}t + \mathbf{b}, u^{-1}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'),$$

então $\mathbf{Z} \sim \text{SMSN}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}', \mathbf{A}\boldsymbol{\Delta}, \mathbf{v})$. □

Em particular, as distribuições marginais também são SMSN. Assim, se $\mathbf{Y} \sim \text{SMSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{v})$, vamos considerar a partição

$$\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top, \text{ onde } \mathbf{Y}_1 : p_1 \times 1 \text{ e } \mathbf{Y}_2 : p_2 \times 1, \text{ com } p_1 + p_2 = p. \quad (2.7)$$

Então $\mathbf{Y}_1 = \mathbf{A}\mathbf{Y}$, onde $\mathbf{A} = (\mathbf{I}_{p_1} \ \mathbf{0}_{p_1 \times p_2})$. A matriz \mathbf{A} induz partições similares em $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ e $\boldsymbol{\Delta}$, as quais denotamos por $\boldsymbol{\mu}_1 : p_1 \times 1$, $\boldsymbol{\Sigma}_{11} : p_1 \times p_1$ e $\boldsymbol{\Delta}_1 : p_1 \times q$. Pela Proposição 2, vem que $\mathbf{Y}_1 \sim \text{SMSN}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Delta}_1, \mathbf{v})$. Um resultado análogo vale para \mathbf{Y}_2 .

Capítulo 3

O Modelo Proposto

3.1 Uma Extensão do Modelo de Regressão Linear com Erros nas Variáveis

Suponha que em um modelo de regressão ambas as variáveis, preditora e resposta, estão sujeitas a erros de mensuração. Neste caso, o(a) experimentador(a) não observa os valores da preditora diretamente. Em vez disso, observações de uma variável substituta são utilizadas para proceder inferência em conjunto com os valores observados da resposta. Neste caso, temos o chamado *modelo de regressão com erros nas variáveis*. Algumas referências são os livros de Fuller (1987), Cheng & Van Ness (1999) e Buonaccorsi (2010). O modelo é definido por

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i, \\ X_i &= x_i + \zeta_i, \quad i = 1, \dots, n, \end{aligned}$$

onde é assumido que

$$(x_i, \zeta_i, e_i)^\top \sim N_3 \left(\begin{pmatrix} \mu \\ \mathbf{0}_2 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{2 \times 1} & \mathbf{\Gamma} \end{pmatrix} \right),$$

onde α e β são parâmetros de regressão e assumimos que a matriz de covariâncias do vetor de erros de mensuração $(\zeta_i, e_i)^\top$ é diagonal, tal que $\mathbf{\Gamma} = \text{diag}(\gamma^2, \sigma^2)$.

Um fato bem conhecido em relação a este modelo é que ele é não identificável

(no sentido frequentista). Ou seja, diferentes conjuntos de parâmetros podem levar a mesma distribuição conjunta de X_i e Y_i , ver Fuller (1987, seção 1.1.3) ou Cheng & Van Ness (1999, seção 1.2.1). Algumas suposições clássicas que individualmente garantem identificabilidade são

- i. A razão entre as variâncias dos erros σ^2/γ^2 é conhecida;
- ii. A *razão de confiabilidade*

$$\kappa_x = \frac{\sigma_x^2}{(\sigma_x^2 + \gamma^2)} \quad (3.1)$$

é conhecida ;

iii. γ^2 é conhecido ;

iv. σ^2 é conhecido;

v. α é conhecido e $\mu \neq 0$.

As suposições ii e iii serão utilizadas nas aplicações com dados reais e simulados que faremos no Capítulo 4.2.

Considerando o contexto de vários regressores, Yao & Song (2014) propuseram uma mistura finita de modelos de regressão linear em que, em (1.1), o vetor de covariáveis \mathbf{x}_i ou partes de \mathbf{x}_i não podem ser observados diretamente e, em vez disso, um vetor aleatório substituto denotado por \mathbf{X}_i , é observado. Ou seja,

$$Y_i | \omega_i = j \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}_j, \sigma_j^2),$$

$$\mathbf{X}_i = \mathbf{x}_i + \boldsymbol{\zeta}_i,$$

$$P(\omega_i = j) = p_j, \quad j = 1, \dots, G, \quad i = 1, \dots, n.$$

Nesse caso, não é assumido uma distribuição para \mathbf{x}_i . Em vez disso, a densidade de \mathbf{x}_i é estimada não parametricamente. Além disso, eles consideram os casos em que a distribuição de $\boldsymbol{\zeta}_i$ é completamente conhecida (normal, exponencial dupla ou normal com matriz de covariância desconhecida). Em seguida apresentam um algoritmo EM para estimação de máxima verossimilhança.

Apresentaremos uma abordagem diferente da de Yao & Song (2014), tratando a questão de estimação pelo ponto de vista Bayesiano, adotando uma abordagem paramétrica. Além disso, estendemos os trabalhos citados no Capítulo 1, assumindo para os erros e as covariáveis uma classe de distribuições mais flexível do que a distribuição normal. Nossa proposta pode ser vista como uma extensão do trabalho de Lachos *et al.* (2009),

onde é adotada uma abordagem semelhante, mas apenas no caso de uma componente. Especificamente, assumimos que dado $\omega_i = j$,

$$Y_i = \alpha_j + \mathbf{x}_i^\top \boldsymbol{\beta}_j + e_{ij} \quad (3.2)$$

$$\mathbf{X}_i = \mathbf{x}_i + \boldsymbol{\zeta}_i, \quad (3.3)$$

com $P(\omega_i = j) = p_j$, $j = 1, \dots, G$, $i = 1, \dots, n$, tal que \mathbf{x}_i é um vetor $r \times 1$ com variáveis de regressão latentes para o indivíduo i , α_j e $\boldsymbol{\beta}_j : r \times 1$ são parâmetros específicos de grupo na estrutura do modelo de misturas de regressões, e_{ij} e $\boldsymbol{\zeta}_i : r \times 1$ são erros aleatórios. Seja $\boldsymbol{\varepsilon}_{ij} = (\boldsymbol{\zeta}_i^\top, e_{ij})^\top$ o vetor de erros de medição para o indivíduo i e o componente j na mistura e $\mathbf{Z}_i = (\mathbf{X}_i^\top, Y_i)^\top$ o vetor com as observações para o indivíduo i .

Podemos reescrever as expressões (3.2) e (3.3) na forma

$$\mathbf{Z}_i = \mathbf{a}_j + \mathbf{B}_j \mathbf{x}_i + \boldsymbol{\varepsilon}_{ij} \quad (3.4)$$

$$= \mathbf{a}_j + \mathcal{B}_j \mathbf{r}_{ij}, \quad (3.5)$$

em que,

$$\mathbf{a}_j = (\mathbf{0}_{1 \times r}, \alpha_j)^\top, \quad \mathbf{B}_j = \begin{pmatrix} \mathbf{I}_r \\ \boldsymbol{\beta}_j^\top \end{pmatrix}, \quad \mathcal{B}_j = (\mathbf{B}_j \quad \mathbf{I}_{p \times p}), \quad (3.6)$$

$\mathbf{0}_{1 \times r}$ denota uma matriz de zeros $1 \times r$, $p = r + 1$, \mathbf{I}_r denota a matriz identidade $r \times r$ e $\mathbf{r}_{ij} = (\mathbf{x}_i^\top, \boldsymbol{\zeta}_i^\top, e_{ij})^\top = (\mathbf{x}_i^\top, \boldsymbol{\varepsilon}_{ij}^\top)^\top$. Assumimos que, dado $\omega_i = j$,

$$\mathbf{r}_{ij} \sim \text{SMSN}_{r+p} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0}_p \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0}_{r \times p} \\ \mathbf{0}_{p \times r} & \boldsymbol{\Gamma}_j \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta} \\ \mathbf{0}_p \end{pmatrix}, \mathbf{v} \right), \quad (3.7)$$

em que,

$$\boldsymbol{\Gamma}_j = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{r \times 1} \\ \mathbf{0}_{1 \times r} & \sigma_j^2 \end{pmatrix}, \quad \text{com } \boldsymbol{\Gamma} = \text{diag}(\gamma_1^2, \dots, \gamma_r^2). \quad (3.8)$$

O modelo definido em (3.2) - (3.7) será denotado pela sigla *FM-SMSNME* (Finite Mixture of *SMSN* Measurement Errors Model), a sigla varia para cada membro diferente da família *SMSN*. A Tabela 3.1 apresenta as notações de acordo com as distribuições.

Tabela 3.1: Siglas dos modelos com membros diferentes da família *SMSN*.

Distribuição	Notação
Normal assimétrica (<i>SN</i>)	<i>FM-SNME</i>
t de Student Assimétrica (<i>ST</i>)	<i>FM-STME</i>
Slash assimétrica (<i>SSL</i>)	<i>FM-SSLME</i>
Normal contaminada assimétrica (<i>SCN</i>)	<i>FM-SCNME</i>

Dado que $\omega_i = j$, então \mathbf{x}_i e $\boldsymbol{\varepsilon}_{ij}$ são independentes com

$$\mathbf{x}_i \sim \text{SMSN}_r(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{v}) \text{ e } \boldsymbol{\varepsilon}_{ij} \sim \text{SMN}_p(\mathbf{0}, \boldsymbol{\Gamma}_j, \mathbf{v}).$$

O que implica que $\boldsymbol{\zeta}_i \sim \text{SMN}_r(\mathbf{0}, \boldsymbol{\Gamma}, \mathbf{v})$ e $e_{ij} \sim \text{SMN}(0, \sigma_j^2, \mathbf{v})$.

A distribuição marginal de \mathbf{Z}_i pode ser obtida combinando (3.5), (3.7) e a Proposição 2, resultando em

$$\mathbf{Z}_i \sim \text{SMSN}_p(\mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu}, \mathbf{B}_j \boldsymbol{\Omega} \mathbf{B}_j^\top + \boldsymbol{\Gamma}_j, \mathbf{B}_j \boldsymbol{\Delta}, \mathbf{v}), \quad i = 1, \dots, n. \quad (3.9)$$

Então, a distribuição de \mathbf{Z}_i é uma mistura finita com j -ésimo componente dado em (3.9) e proporção de mistura p_j . Em outras palavras, para uma amostra $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$, e denotando o conjunto com todos os parâmetros por $\boldsymbol{\Theta}$, a função de verossimilhança é dada por:

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^n L_i(\boldsymbol{\Theta}), \quad (3.10)$$

tal que, (3.11)

$$L_i(\boldsymbol{\Theta}) = \sum_{j=1}^G p_j \text{SMSN}_p(\mathbf{z}_i | \mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu}, \mathbf{B}_j \boldsymbol{\Omega} \mathbf{B}_j^\top + \boldsymbol{\Gamma}_j, \mathbf{B}_j \boldsymbol{\Delta}, \mathbf{v}), \quad i = 1, \dots, n. \quad (3.12)$$

Pela representação hierárquica de um vetor aleatório com distribuição *SMSN* dada em (2.6) temos que, dado $\omega_i = j$,

$$\mathbf{r}_{ij} | U_i = u_i, T_i = t_i \sim \text{N}_{r+p} \left(\left(\begin{array}{c} \boldsymbol{\mu} + \boldsymbol{\Delta} t_i \\ \mathbf{0}_{p+1} \end{array} \right), \left(\begin{array}{cc} u_i^{-1} \boldsymbol{\Omega} & \mathbf{0}_{r \times p} \\ \mathbf{0}_{p \times r} & u_i^{-1} \boldsymbol{\Gamma}_j \end{array} \right) \right);$$

$$T_i | U_i = u_i \sim \text{TN}(0, u_i^{-1}, (0, \infty));$$

$$U_i \sim H(\cdot | \mathbf{v}).$$

Então, usando a expressão (3.4), obtemos a seguinte representação hierárquica para o modelo *FM-SMSNME*:

$$\mathbf{Z}_i | \mathbf{x}_i, U_i = u_i, \omega_i = j \sim N_p(\mathbf{a}_j + \mathbf{B}_j \mathbf{x}_i, u_i^{-1} \mathbf{\Gamma}_j); \quad (3.13)$$

$$\mathbf{x}_i | U_i = u_i, T_i = t_i \sim N_r(\boldsymbol{\mu} + \Delta t_i, u_i^{-1} \mathbf{\Omega});$$

$$T_i | U_i = u_i \sim \text{TN}(0, u_i^{-1}, (0, \infty));$$

$$U_i \sim H(\cdot | \mathbf{v});$$

$$P(\omega_i = j) = p_j. \quad (3.14)$$

Uma representação alternativa é obtida particionando o vetor \mathbf{Z}_i , substituindo a equação (3.13) por

$$Y_i | \mathbf{x}_i, U_i = u_i, \omega_i = j \sim N_p(\alpha_j + \mathbf{x}_i^\top \boldsymbol{\beta}_j, u_i^{-1} \sigma_j^2);$$

$$\mathbf{X}_i | \mathbf{x}_i, U_i = u_i \sim N_r(\mathbf{x}_i, u_i^{-1} \mathbf{\Gamma}),$$

e mantendo as outras.

É possível obter outra representação hierárquica para o modelo *FM-SMSNME*, integrando a variável latente \mathbf{x}_i . Para isso, é suficiente considerar a distribuição marginal (3.9) e a representação hierárquica do vetor aleatório com distribuição *SMSN* dada em (2.6), resultando em

$$\mathbf{Z}_i | U_i = u_i, T_i = t_i, \omega_i = j \sim N_p(\mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu} + \mathbf{B}_j \Delta t_i, u_i^{-1} (\mathbf{B}_j \mathbf{\Omega} \mathbf{B}_j^\top + \mathbf{\Gamma}_j));$$

$$T_i | U_i = u_i \sim \text{TN}(0, u_i^{-1}, (0, \infty));$$

$$U_i \sim H(\cdot | \mathbf{v}).$$

$$P(\omega_i = j) = p_j. \quad (3.15)$$

Estas representações serão importantes para desenvolver um algoritmo do tipo Gibbs para obter estimativas a posteriori para os parâmetros do modelo *FM-SMSNME*, como veremos na próxima seção.

3.2 Estimação a Posteriori

3.2.1 Distribuição a Priori

No contexto Bayesiano, especificações para distribuição a priori são necessárias para inferência. Seja Θ o vetor de parâmetros a ser estimado, isto é, o vetor contendo os parâmetros $\gamma_1^2, \dots, \gamma_p^2, \sigma_1^2, \dots, \sigma_G^2, \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_G)^\top, \boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_G^\top)^\top, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \mathbf{p} = (p_1, \dots, p_G)'$ e \mathbf{v} . Fixamos a seguinte especificação a priori:

$$\boldsymbol{\phi} = (\boldsymbol{\mu}^\top, \boldsymbol{\Delta}^\top)^\top \sim N_{2p}(\mathbf{a}_\phi, \mathbf{A}_\phi), \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \boldsymbol{\beta}_j \sim N_p(\mathbf{c}_\beta, \mathbf{C}_\beta), \quad j = 1, \dots, G,$$

onde supomos que todos os hiperparâmetros envolvidos são conhecidos, sendo que as matrizes de covariância \mathbf{A}_ϕ e \mathbf{C}_β são positivas definidas e diagonais. A razão para concatenar os vetores $\boldsymbol{\mu}$ e $\boldsymbol{\Delta}$ é a obtenção de um algoritmo eficiente MCMC para extrair amostras a partir da distribuição a posteriori, como ficará mais claro a seguir.

Em relação aos parâmetros de escala, adotamos a estrutura a priori hierárquica dada por

$$\begin{aligned} \sigma_j^{-2} | f &\sim \text{Gamma}(e, f) \quad j = 1, \dots, G \\ f &\sim \text{Gamma}(g, h). \end{aligned}$$

Esta configuração foi sugerida por Richardson & Green (1997) para o caso de misturas de normais univariadas. Além disso fixamos

$$\gamma_j^{-2} \sim \text{Gamma}(l, m), \quad j = 1, \dots, p,$$

onde e, g, h, l e m são conhecidos e positivos. No que diz respeito a inversa da matriz escala $\boldsymbol{\Omega}$, escolhemos a priori

$$\boldsymbol{\Omega}^{-1} \sim \text{Wish}_p(r, \mathbf{S}),$$

que é a distribuição Wishart com densidade

$$\pi(\mathbf{U} | r, \mathbf{S}) \propto \det(\mathbf{U})^{r-(p+1)/2} \exp\{-\text{tr}(\mathbf{S}\mathbf{U})\},$$

em que $\boldsymbol{\Omega}^{-1}, \mathbf{S}$ é positiva definida, $r > (p-1)/2$, $\det(\cdot)$ e $\text{tr}(\cdot)$, denotam o determinante

e o traço respectivamente. Para o vetor de pesos, aplicamos a suposição habitual

$$\mathbf{p} \sim \text{Dir}(\kappa_1, \dots, \kappa_G),$$

que é a distribuição Dirichlet com hiperparâmetros conhecidos e positivos.

Em relação ao parâmetro fator de escala \mathbf{v} , cada modelo específico na família SMSN é tratado de maneira diferente. Por exemplo, existem várias sugestões para estimar graus de liberdade desconhecidos do modelo t de Student, veja as discussões em Fonseca *et al.* (2008) e Garay *et al.* (2015). Considerando o caso da t assimétrica adotamos a estrutura

$$\begin{aligned} v &\sim \exp(\lambda) \\ \lambda &\sim \text{Uniforme}(\lambda_0, \lambda_1), \end{aligned}$$

onde $0 < \lambda_0 < \lambda_1$ são conhecidos.

A configuração a priori para os modelos slash assimétrica e normal contaminada assimétrica são dados por

Distribuição	Priori
Slash assimétrica	$v \sim \text{Gama}(v_0, v_1)$
Normal contaminada assimétrica	$\rho \sim \text{Beta}(\rho_0, \rho_1)$ $\tau \sim \text{Beta}(\tau_0, \tau_1)$

onde $v_0, v_1, \rho_0, \rho_1, \tau_0$ e τ_1 são conhecidos e positivos. Estas configurações a priori foram usadas com êxito em outros trabalhos, veja Cabral *et al.* (2012a).

Assumindo independência entre os parâmetros, a especificação a priori completa é dada por

$$\pi(\Theta) = \pi(\mathbf{p})\pi(\phi)\pi(\mathbf{v}|\lambda)\pi(\lambda)\pi(\mathbf{\Omega}^{-1}) \prod_{j=1}^G \pi(\boldsymbol{\beta}_j) \prod_{j=1}^G \pi(\sigma_j^{-2}|f)\pi(j), \prod_{j=1}^P \pi(\gamma_j^2),$$

excluindo λ quando necessário, onde $\pi(\cdot)$ denota uma densidade genérica.

3.2.2 Algoritmo Tipo Gibbs

Nesta seção propomos um algoritmo eficiente do tipo Gibbs para estimação Bayesiana. O modelo *FM-SMSNME* permite um desenvolvimento direto de vários amostradores de Gibbs, através das representações hierárquicas apresentadas anteriormente. Temos, por exemplo, o algoritmo padrão que consiste em amostrar os parâmetros marginalmente a partir das distribuições condicionais completas, ou seja, os parâmetros são atualizados um de cada vez. Infelizmente, sabemos que esse método apresenta convergência lenta e resulta em amostras altamente correlacionadas. Um procedimento mais eficiente pode ser obtido atualizando os parâmetros simultaneamente, um método usualmente chamado blocagem, ou quando as amostras são obtidas a partir das distribuições parcialmente marginalizadas, integrando algumas variáveis latentes que estão presentes no esquema de dados aumentados, um método denominado colapsagem, veja Liu (1994) para mais detalhes.

Antes de apresentar o algoritmo, vamos introduzir algumas novas notações. Para um vetor fixo $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$, definimos

$$\mathcal{A}_j = \{i \in \{1, \dots, n\}; \omega_i = j\} \text{ e } m_j = \text{cardinal de } \mathcal{A}_j.$$

Além disso, seja $\boldsymbol{\gamma} = (\gamma_1^2, \dots, \gamma_p^2)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_G^\top)^\top$ e $\boldsymbol{\sigma} = (\sigma_1^2, \dots, \sigma_G^2)^\top$. Dada uma amostra observada $\mathbf{z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$, obteremos amostras da distribuição a posteriori dos parâmetros do modelo *FMSMSN-ME* através de um algoritmo MCMC. Com este objetivo, obteremos as distribuições condicionais completas, que serão denotadas por $\pi(\cdot | \mathbf{z}, \dots)$. Detalhes do algoritmo podem ser encontrados no Apêndice A.

Para obtermos uma amostra da distribuição a posteriori do modelo *FM-SMSNME*, os passos são os seguintes:

Passo 1. Para todo $i = 1, \dots, n$, gere ω_i independentes, com a seguinte distribuição discreta

$$P(\omega_i = j | \mathbf{z}_i, \dots) = \frac{p_j \text{SMSN}_{p+1}(\mathbf{z}_i | \boldsymbol{\theta}_j, \boldsymbol{\Psi}_j, \boldsymbol{\Lambda}_j, \mathbf{v})}{\sum_{k=1}^G p_k \text{SMSN}_{p+1}(\mathbf{z}_i | \boldsymbol{\theta}_k, \boldsymbol{\Psi}_k, \boldsymbol{\Lambda}_k, \mathbf{v})}, \quad j = 1, \dots, G,$$

tal que $\boldsymbol{\theta}_j = \mathbf{B}_j \boldsymbol{\mu}_j$, $\boldsymbol{\Lambda}_j = \mathbf{B}_j \boldsymbol{\Delta}$ e $\boldsymbol{\Psi}_j = \mathbf{B}_j \boldsymbol{\Omega} \mathbf{B}_j^\top + \boldsymbol{\Gamma}_j$, $j = 1, \dots, G$.

Passo 2. Para obter amostras de $\mathbf{p} = (p_1, \dots, p_G)^\top$ gere a partir da distribuição de $\pi(\mathbf{p} | \boldsymbol{\omega})$,

dada pela Dirichilet

$$\text{Dir}(m_1 + \kappa_1, \dots, m_G + \kappa_G),$$

onde m_j indica a quantidade de elementos alocados na classe j , para todo $j = 1, \dots, G$.

Passo 3. Para todo $i = 1, \dots, n$, gere amostras t_i independentes de $\pi(t_i | \mathbf{z}_i, u_i, w_i = j, \dots)$, que equivale a distribuição normal truncada, dada por

$$\text{TN} \left(\frac{\boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \boldsymbol{\theta}_j)}{1 + \boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j}, \frac{1}{u_i (1 + \boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j)}, (0, \infty) \right),$$

os parâmetros $\boldsymbol{\theta}_j$ e $\boldsymbol{\Psi}_j$ são dados no *Passo 1*.

Passo 4. Para todo $i = 1, \dots, n$, gere amostras u_i independentes de $\pi(u_i | \mathbf{z}_i, t_i, w_i = j, \dots)$, que é equivalente às seguintes distribuições

(a) Para o caso da distribuição t de Student,

$$\text{Gama} \left((v + p + 1)/2, (1/2)(\lambda_i^{(j)} + v) \right).$$

(b) Para o caso da distribuição Slash assimétrica,

$$\text{TG} \left((2v + p + 1)/2, \lambda_i^{(j)}/2, (0, 1) \right),$$

onde TG denota a distribuição Gama truncada no intervalo $(0, 1)$, e $\lambda_i^{(j)}$ é dado por

$$\lambda_i^{(j)} = (\mathbf{z}_i - \boldsymbol{\theta}_j - \boldsymbol{\Lambda}_j t_i)' \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \boldsymbol{\theta}_j - \boldsymbol{\Lambda}_j t_i) + t_i^2.$$

(c) Para a distribuição normal contaminada assimétrica, gere amostras u_i independentes a partir da distribuição discreta que assume o valor τ com probabilidade $\eta_i / (\eta_i + \xi_i)$ e 1 com probabilidade $\xi_i / (\eta_i + \xi_i)$, tal que

$$\begin{aligned} \eta_i &= \rho \tau^{\frac{p+1}{2}} \exp \left\{ -\frac{\tau}{2} [\mathbf{z}_i - (\boldsymbol{\theta}_j + \boldsymbol{\Lambda}_j t_i)]^\top \boldsymbol{\Psi}_j^{-1} [\mathbf{z}_i - (\boldsymbol{\theta}_j + \boldsymbol{\Lambda}_j t_i)] + t_i^2 \right\} \text{ e} \\ \xi_i &= (1 - \rho) \exp \left\{ -\frac{1}{2} [\mathbf{z}_i - (\boldsymbol{\theta}_j + \boldsymbol{\Lambda}_j t_i)]^\top \boldsymbol{\Psi}_j^{-1} [\mathbf{z}_i - (\boldsymbol{\theta}_j + \boldsymbol{\Lambda}_j t_i)] + t_i^2 \right\}. \end{aligned}$$

Passo 5. Para todo $i = 1, \dots, n$, gere \mathbf{x}_i independentes a partir de $\pi(\mathbf{x}_i | \mathbf{z}_i, t_i, u_i, w_i = j, \dots)$, com distribuição normal $N_p(\boldsymbol{\mu}_{\mathbf{x}_i}, u_i^{-1} \boldsymbol{\Sigma}_{\mathbf{x}_i})$, onde $\boldsymbol{\mu}_{\mathbf{x}_i}$ e $\boldsymbol{\Sigma}_{\mathbf{x}_i}$ são dados pelas expressões

$$\boldsymbol{\mu}_{\mathbf{x}_i} = \boldsymbol{\Sigma}_{\mathbf{x}_i} \left[\mathbf{B}_j^\top \boldsymbol{\Gamma}_j \mathbf{z}_i + \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} + \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta} t_i \right] \text{ e}$$

$$\boldsymbol{\Sigma}_{\mathbf{x}_i} = \left(\boldsymbol{\Omega}^{-1} + \mathbf{B}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j \right)^{-1}.$$

Passo 6. Gere amostras $\boldsymbol{\phi} = (\boldsymbol{\mu}^\top, \boldsymbol{\Delta}^\top)^\top$ de $\pi(\boldsymbol{\phi} | \mathbf{z}^*, \mathbf{u}, \mathbf{t}, \boldsymbol{\omega}, \dots)$, em que $\mathbf{z}^* = \{\mathbf{z}_1^*, \dots, \mathbf{z}_n^*\}$, $\mathbf{z}_i^* = u_i^{1/2} \mathbf{z}_i$, $\mathbf{u} = \{u_1, \dots, u_n\}$, $\mathbf{t} = \{t_1, \dots, t_n\}$ e $\boldsymbol{\omega} = \{\omega_1, \dots, \omega_n\}$, com distribuição $N_{2p}(\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\phi)$, tal que,

$$\boldsymbol{\Sigma}_\phi = \left(\sum_{i=1}^n (\mathbf{W}_i^*)^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{W}_i^* + \mathbf{A}_\phi^{-1} \right)^{-1} \text{ e } \boldsymbol{\mu}_\phi = \boldsymbol{\Sigma}_\phi \left(\sum_{i=1}^n (\mathbf{W}_i^*)^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{z}_i^* + \mathbf{A}_\phi^{-1} \mathbf{a}_\phi \right),$$

onde \mathbf{W}_i^* é uma matriz de dimensão $(p+1) \times 2p$, dada por

$$\mathbf{W}_i^* = \begin{bmatrix} u_i^{1/2} \mathbf{B}_{\omega_i} & u_i^{1/2} t_i \mathbf{B}_{\omega_i} \end{bmatrix}, \quad i = 1, \dots, n.$$

Passo 7. Gere amostras $\boldsymbol{\beta}_j$ de $\pi(\boldsymbol{\beta}_j | \mathbf{z}, \mathbf{u}, \mathbf{t}, \boldsymbol{\omega}, \dots)$ com distribuição $N_p(\boldsymbol{\mu}_{\beta_j}, \boldsymbol{\Sigma}_{\beta_j})$, onde

$$\boldsymbol{\mu}_{\beta_j} = \boldsymbol{\Sigma}_{\beta_j} \left(\sum_{i \in \mathcal{A}_j} \frac{1}{\sigma_j^2} u_i Y_i \mathbf{x}_i + \mathbf{C}_\beta^{-1} \mathbf{c}_\beta \right) \text{ e } \boldsymbol{\Sigma}_{\beta_j} = \left(\sum_{i \in \mathcal{A}_j} \frac{1}{\sigma_j^2} u_i \mathbf{x}_i \mathbf{x}_i^\top + \mathbf{C}_\beta^{-1} \right)^{-1}.$$

Passo 8. Gere γ_j^{-2} de $\pi(\gamma_j^{-2} | \mathbf{z}, \mathbf{u}, \dots)$, $j = 1, \dots, p$, independentes, com distribuição

$$\text{Gama} \left(\frac{n}{2} + l, \frac{1}{2} \sum_{i=1}^n u_i (X_{ij} - x_{ij})^2 + m \right),$$

onde $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^\top$ e $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$.

Passo 9. Gere amostras σ_j^{-2} a partir de $\pi(\sigma_j^{-2} | \mathbf{z}, \mathbf{u}, \dots)$, $j = 1, \dots, G$, independentes, com distribuição

$$\text{Gama} \left(\frac{m_j}{2} + e, \frac{1}{2} \sum_{i \in \mathcal{A}_j} u_i (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_j)^2 + f \right).$$

Passo 9a. Gere f a partir de $\pi(f|\sigma_1^2, \dots, \sigma_G^2)$, com distribuição

$$\text{Gama}(Ge + g, \sum_{j=1}^G \sigma_j^{-2} + h).$$

Passo 10. Gere $\mathbf{U} = \mathbf{\Omega}^{-1}$ a partir $\pi(\mathbf{U}|\mathbf{x}, \mathbf{u}, \mathbf{t}, \dots)$, com distribuição

$$\text{Wish}_p \left(\frac{n}{2} + r, \frac{1}{2} \sum_{i=1}^n u_i (\mathbf{x}_i - \boldsymbol{\mu}_i - \Delta t_i) (\mathbf{x}_i - \boldsymbol{\mu}_i - \Delta t_i)^\top + \mathbf{S} \right),$$

tal que $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$.

Passo 11. Gere amostras da posteriori de \mathbf{v} , levando em consideração três casos:

- (a) *t* de Student assimétrica: Obtenha amostras de \mathbf{v} utilizando um passo de Metropolis-Hastings a partir da distribuição condicional marginalizada

$$\pi(\mathbf{v}|\mathbf{z}, \dots) \propto \exp\{-\lambda \mathbf{v}\} \prod_{i=1}^n \sum_{j=1}^G p_j \text{ST}_{p+1}(\mathbf{z}_i | \boldsymbol{\theta}_j, \boldsymbol{\Psi}_j, \boldsymbol{\Lambda}_j, \mathbf{v}).$$

Dada a observação $\mathbf{v}^{(t-1)}$, obtida na iteração $t - 1$, gere um candidato a nova observação \mathbf{v}^c a partir da distribuição lognormal $\text{LN}(\mathbf{v}^{(t-1)}, \boldsymbol{\sigma}_v^2)$. A nova observação é aceita com probabilidade

$$Q = \min \left\{ \frac{\pi(\mathbf{v}^c|\mathbf{z}, \dots) \mathbf{v}^c}{\pi(\mathbf{v}^{(t-1)}|\mathbf{z}, \dots) \mathbf{v}^{(t-1)}}, 1 \right\}. \quad (3.16)$$

- (a.1) Gere λ a partir da distribuição $\text{TGama}(2, \mathbf{v}, (\lambda_0, \lambda_1))$, que é uma distribuição Gama truncada no intervalo (λ_0, λ_1) .

- (b) *Slash* assimétrica: Gere amostras de \mathbf{v} a partir de $\pi(\mathbf{v}|\mathbf{u})$, com distribuição

$$\text{Gamma} \left(n + v_0, v_1 - \sum_{i=1}^n \log u_i \right).$$

- (c) Normal contaminada assimétrica: Gere amostras de ρ a partir $\pi(\rho|\mathbf{u}, \tau)$, com dis-

tribuição Beta ($\kappa_0 + \rho_0, \kappa_1 + \rho_1$), onde

$$\kappa_0 = (n - \sum_{i=1}^n u_i) / (1 - \tau) + \rho_0 \quad \text{e}$$

$$\kappa_1 = (\sum_{i=1}^n u_i - n\tau) / (1 - \tau) + \rho_1.$$

É possível mostrar que

$$\kappa_0 = \text{número de } u_{is} \text{ iguais a } \tau, \quad \text{e} \quad \kappa_1 = n - \kappa_0.$$

(c.1) Finalmente, usando um passo de Metropolis-Hastings, gere amostras de τ a partir da distribuição condicional marginalizada dada por

$$g(\tau | \mathbf{z}, \boldsymbol{\rho}, \dots) \propto \tau^{\tau_0 - 1} (1 - \tau)^{\tau_1 - 1} \prod_{i=1}^n \sum_{j=1}^G \text{SCN}_{p+1}(\mathbf{z}_i | \boldsymbol{\theta}_j, \boldsymbol{\Psi}_j, \boldsymbol{\Lambda}_j, \tau, \boldsymbol{\rho}).$$

Em vez de uma amostra desta distribuição, consideramos a transformação $\bar{\tau} = \tau / (1 - \tau)$, que é uma função que assume valores entre $(0, 1)$ e \mathbb{R}^+ . Então, tomamos amostras a partir da distribuição de $\bar{\tau}$, dada por

$$\pi(\bar{\tau} | \mathbf{z}, \boldsymbol{\rho}, \dots) = g(\bar{\tau} / (1 + \bar{\tau}) | \mathbf{z}, \boldsymbol{\rho}) / (1 + \bar{\tau})^2.$$

Para fazer isso, usamos um passo de Metropolis-Hastings com distribuição proposta log-normal, seguindo os mesmos passos do caso da t de Student assimétrica, veja (3.16). Então, para obter a amostra de τ desejada, revertemos a transformação, isto é, $\tau = \bar{\tau} / (1 + \bar{\tau})$.

A estratégia utilizada nos passos 11(a) e 11(c.1) é conhecida por *colapsabilidade*. Consiste em extrair amostras da distribuição condicional marginalizada, que é obtida integrando a verossimilhança aumentada tanto quanto possível. Conforme mostrado por Liu (1994), este procedimento geralmente produz convergência mais rápida e amostras menos correlacionadas do que gerar a partir de $\pi(\mathbf{v} | \mathbf{u})$.

3.3 Critérios de Seleção de Modelos

Uma das situações que nos deparamos ao modelar um conjunto de dados, seja real ou simulado, é no momento de escolher o melhor modelo entre alguns candidatos a fim explicar o fenômeno de estudo. Dentre as diversas metodologias de seleção de modelos apresentados na literatura, usaremos critérios Bayesianos, como o Critério de Informação do Desvio ou simplesmente DIC, Spiegelhalter *et al.* (2002), veja também Celeux *et al.* (2006) e como alternativa o Critério de Informação de Watanabe-Akaike, ou WAIC, proposto por Watanabe (2010). Tais critérios são baseados em medidas de ajuste e complexidade a partir da distribuição a posteriori do desvio. Dessa forma, o modelo que apresentar o menor valor de critério será o modelo mais adequado.

3.3.1 Critério de Informação do Desvio Observado

Apresentaremos uma modificação do DIC original de Spiegelhalter *et al.* (2002) que foi proposta por Celeux *et al.* (2006, Sec. 3.1). Esta modificação é elaborada de forma adequada para o contexto de dados aumentados. De fato, o DIC usual não é adequado nesta situação, porque a verossimilhança para os dados incompletos não é regular. Como consequência, os argumentos assintóticos que o validam não podem ser verificados (Li *et al.*, 2013). O critério proposto por Celeux *et al.* (2006) será chamado *DIC observado* e será denotado por DIC_{obs} .

Seja $D(\Theta) = -2\log L(\Theta)$ o *desvio*, em que $L(\Theta)$ é a função de verossimilhança definida em (3.11). Então a quantidade:

$$\overline{D(\Theta)} = E[D(\Theta)|z], \quad (3.17)$$

é denominada de desvio médio a posteriori. E a medida

$$\tau_D = \overline{D(\Theta)} - D(\tilde{\Theta}), \quad (3.18)$$

é denominada de *dimensão efetiva* (ou número efetivo de parâmetros conforme visto em Spiegelhalter *et al.* (2002)) $\tilde{\Theta}$, onde é um estimador para Θ .

Analogamente aos critérios clássicos de seleção de modelos como o AIC (Akaike,

1974), o DIC é definido como

$$\text{DIC} = D(\bar{\Theta}) + 2\tau_D,$$

em que $\bar{\Theta} = E[\Theta|\mathbf{z}]$ é a esperança a posteriori de Θ , ou seja, o primeiro termo do DIC é uma medida de ajuste e o segundo termo é uma penalidade considerando a complexidade do modelo. Se usarmos $\tilde{\Theta} = \bar{\Theta}$ em (3.18), temos:

$$\begin{aligned} \text{DIC} &= -2\log L(\bar{\Theta}) + 2E[D(\Theta)|\mathbf{z}] + 4\log L(\bar{\Theta}) \\ &= 2E[D(\Theta)|\mathbf{z}] + 2\log L(\bar{\Theta}) \\ &= \overline{D(\Theta)} + \tau_D. \end{aligned} \tag{3.19}$$

Na prática, geralmente usamos $\tilde{\Theta} = \bar{\Theta} = E[\Theta|\mathbf{z}]$. Se a distribuição de Θ é muito distante de uma normal, então $\bar{\Theta}$ não é um bom estimador. Segue que τ_D pode inclusive assumir valores negativos. Um problema em modelos de mistura é que Θ nem sempre identificável. A falta de identificabilidade resulta porque a verossimilhança do modelo de mistura finita é invariante com respeito a permutações dos índices das componentes. Se a priori também for invariante com respeito a estes índices, todas as médias a posteriori serão iguais, e a mistura plug-in $L_i(\bar{\Theta})$ terá somente uma componente. Como consequência, o estimador $D(\bar{\Theta})$ de $D(\Theta)$ não é razoável e o DIC definido em (3.19) é inútil. Para mais detalhes, veja a discussão em (Stephens, 1997, pp. 13).

Para resolver este problema observe que, como a verossimilhança $L_i(\Theta)$ é invariante a permutação dos índices das componentes, a verossimilhança associada à amostra $L(\Theta) = \prod_{i=1}^n L_i(\Theta)$ também é invariante. Uma estimativa de $L_i(\Theta)$ é a densidade preditiva a posteriori dada por $E[L_i(\Theta)|\mathbf{z}]$. Esta integral pode ser aproximada usando amostras a posteriori MCMC. Seja $\Theta^{(l)}$ a amostra MCMC gerada na l -ésima iteração do algoritmo, para $l = 1, \dots, m$. Aproximamos a densidade preditiva por

$$\hat{p}(\mathbf{z}_i) = \frac{1}{m} \sum_{l=1}^m L_i(\Theta^{(l)}).$$

Assim, um estimador de $L(\Theta)$ é:

$$\hat{p}(\mathbf{z}) = \prod_{i=1}^n \hat{p}(\mathbf{z}_i), \tag{3.20}$$

e um estimador para $D(\Theta)$ é $-2\log \hat{p}(\mathbf{z})$, que não depende de permutações dos índices das componentes.

A esperança a posteriori $E[D(\Theta)|\mathbf{z}]$ pode ser aproximada por:

$$\bar{D} = -\frac{2}{m} \sum_{l=1}^m \sum_{i=1}^n \log [L_i(\Theta^{(l)})].$$

Finalmente, a aproximação do DIC que será usada neste trabalho é

$$\text{DIC}_{\text{obs}} = \bar{D} + \tau_D, \quad \text{tal que } \tau_D = \bar{D} + 2 \sum_{i=1}^n \log \hat{p}(\mathbf{z}_i).$$

O modelo que apresentar o menor DIC_{obs} é considerado como melhor opção dentre os demais. Para um estudo detalhado, veja a seção 3.1 de Celeux *et al.* (2006).

3.3.2 Critério de Informação Watanabe-Akaike

O Critério de Informação Watanabe-Akaike foi introduzido por Watanabe (2010).

A definição de WAIC é semelhante às do AIC e DIC, ou seja,

$$\text{WAIC} = \text{medida de ajuste} + 2 \times \text{penalidade}.$$

Neste caso, usamos

$$\begin{aligned} \text{medida de ajuste} &= -2 \sum_{i=1}^n \log \pi(\mathbf{z}_i|\mathbf{z}) \\ &= -2 \sum_{i=1}^n \log \int \pi(\mathbf{z}_i|\Theta) \pi(\Theta|\mathbf{z}) d\Theta \\ &= -2 \sum_{i=1}^n \log E[L_i(\Theta)|\mathbf{z}], \end{aligned} \tag{3.21}$$

em que, $\pi(\mathbf{z}_i|\mathbf{z})$ é a densidade a posteriori de \mathbf{Z}_i . Ao contrário de $D(\bar{\Theta})$, que é a medida de ajuste usada para definir DIC, a medida de ajuste acima é invariante a reparametrizações.

As penalidades para a complexidade são definidas de duas maneiras:

$$\begin{aligned} \rho_{\text{WAIC}_1} &= \overline{D(\Theta)} + 2 \sum_{i=1}^n \log \pi(\mathbf{z}_i|\mathbf{z}) \\ \rho_{\text{WAIC}_2} &= \sum_{i=1}^n \text{Var} [\log L_i(\Theta)|\mathbf{z}], \end{aligned}$$

onde, $\overline{D(\boldsymbol{\theta})}$ é dado em (3.17). Não existe uma expressão em forma fechada para (3.21).

Uma aproximação usando amostras MCMC é dada por:

$$-2 \sum_{i=1}^n \log \left(\frac{1}{m} \sum_{l=1}^m L_i(\boldsymbol{\Theta}^{(l)}) \right) = -2 \log \hat{p}(\mathbf{z}),$$

em que $\hat{p}(\mathbf{z})$ é dado em (3.20). Logo,

$$\begin{aligned} \text{WAIC}_1 &= -2 \log \hat{p}(\mathbf{z}) + 2\rho_{\text{WAIC}_1} \\ &= -2 \log \hat{p}(\mathbf{z}) + 2\overline{D(\boldsymbol{\theta})} + 4 \log \hat{p}(\mathbf{z}) \\ &= 2 \log \hat{p}(\mathbf{z}) + 2\overline{D(\boldsymbol{\theta})} \\ &= \text{DIC}_{\text{obs}}. \end{aligned}$$

Neste trabalho, também usaremos o WAIC_2 , definido por:

$$\text{WAIC}_2 = -2 \log \hat{p}(\mathbf{z}) + 2\rho_{\text{WAIC}_2}.$$

Assim como o DIC_{obs} , o modelo que apresentar o menor valor do WAIC_2 é considerado como sendo o melhor modelo ajustado.

Capítulo 4

Simulação e Aplicação em Dados Reais

Neste capítulo vamos mostrar o desempenho dos modelos *FM-SMSNME*, comparando com os modelos que incorporam distribuições como a normal, T de student, slash e normal contaminada. Isso dar-se-á em dois processos, na seção 4.1, ilustramos através de simulação e na seção 4.2 aplicamos os modelos em dados reais. As rotinas computacionais foram implementadas nos softwares R Core Team (2017) e no *JAGS*¹, através do pacote *rjags*, para mais detalhes veja Plummer (2016). Para construção dos gráficos, utilizamos o pacote *ggplot2* (Wickham, 2009).

4.1 Simulação

O objetivo desta simulação é analisar a performance dos modelos *FM-SMSNME*. Para isso, considere o modelo visto em (3.4), com representações dadas em (2.6) e (3.14). Neste estudo, usamos duas componentes na mistura, isto é, $G = 2$.

A matriz de dados \mathbf{Z} , foi composta por duas colunas, uma com a variável resposta Y e a variável substituta X , e o número de linhas igual a 100, que é o tamanho da amostra. Os valores dos parâmetros foram fixados em: $\boldsymbol{\alpha}^\top = (\alpha_1, \alpha_2) = (0.1, 1)$, $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (0.8, 1.2)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2) = (0.05, 0.05)$, $\mathbf{p} = (p_1, p_2) = (0.6, 0.4)$, $\mu = 4$, $\Omega = 1$, $\Delta = 6$, $\nu = 3$, $\tau = 0.3$ e $\rho = 0.7$. Vamos fazer uma suposição de identificabilidade assumindo que γ^2 é conhecido, veja a seção 3.1. Aqui, vamos tomar $\gamma^2 = 0.1$.

Para estimação usamos o algoritmo de Gibbs, com 50000 iterações, descartando as 10000 primeiras observações da amostra MCMC (conhecida por burn-in). A fim de

¹Just Another Gibbs Sampler - É um programa para análise de modelos hierárquicos bayesianos usando a simulação Markov Chain Monte Carlo (MCMC).

reduzir o efeito de correlação, foram salvas as observações geradas de 10 em 10 iterações (lag), totalizando ao final uma amostra de tamanho 4 mil.

A seguir, descrevemos um breve roteiro dessa simulação.

1. Uma amostra de tamanho 100, foi gerada a partir da distribuição normal assimétrica, com valores fixos para α , β , σ , ρ , μ , Ω e Δ citados anteriormente.

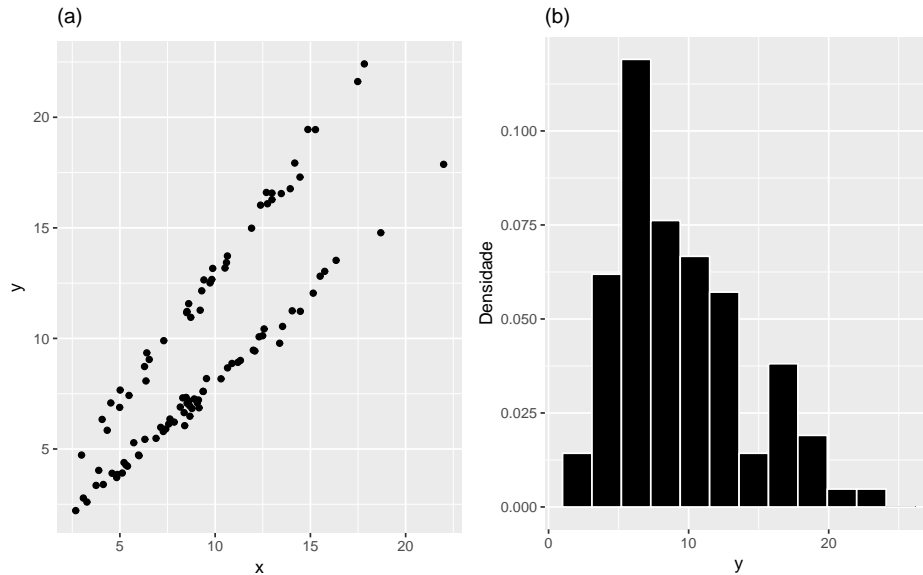


Figura 4.1: *Conjunto de dados simulados. (a) Diagrama de dispersão e (b) Histograma da variável resposta.*

A Figura 4.1 exemplifica um dos conjuntos de dados simulados com $n = 100$. Observamos na Figura 4.1 (a) duas tendências bem separadas, isto representa uma mistura com duas componentes. Já na Figura 4.1 (b) visualizamos a assimetria e caudas pesadas da distribuição da variável resposta.

2. Dessa amostra, selecionamos aleatoriamente três observações, y_3, y_{66} e y_{92} . Para deixar essa distribuição mais heterogênea, perturbamos estas observações através da seguinte expressão

$$y^* = \left(1 + \frac{\Lambda}{100}\right) \times y,$$

em que, $\Lambda = 10, 20, \dots, 150$.

3. Para os 15 padrões de perturbação, ajustamos os modelos *FM-SNME*, *FM-STME*, *FM-SSLME*, *FM-SCNME*, *FMNME*, *FM-TME*, *FM-SLME* e *FM-CNME*, com o objetivo de verificar o comportamento das estimativas à medida que Λ aumenta.

4. Para cada ajuste, foram calculados os critérios, DIC observado e WAIC.

5. A fim de revelar o impacto das contaminações sobre as estimativas dos parâme-

tros, foi calculada a mudança relativa para as estimativas obtidas, pela seguinte expressão

$$RC = \left| \frac{\hat{\theta}_j - \hat{\theta}_{j(I)}}{\hat{\theta}_j} \right| \times 100,$$

em que, $\hat{\theta}_j$ denota a estimativa sem contaminação e $\hat{\theta}_{j(I)}$ a estimativa com contaminação. O trabalho de Massuia *et al.* (2017) procedeu um estudo de simulação similar no contexto de dados censurados, com modelagem baseada em regressão linear utilizando mistura de escalas da normal assimétrica.

6. Os passos anteriores foram replicados 100 vezes.

7. Dos resultados dessas réplicas, calculamos a média dos RC 's e dos critérios de seleção.

Na sequência, mostramos o comportamento das estimativas dos parâmetros (em comum) nos ajustes $FM-SMSNME$, mediante a variação de Λ .

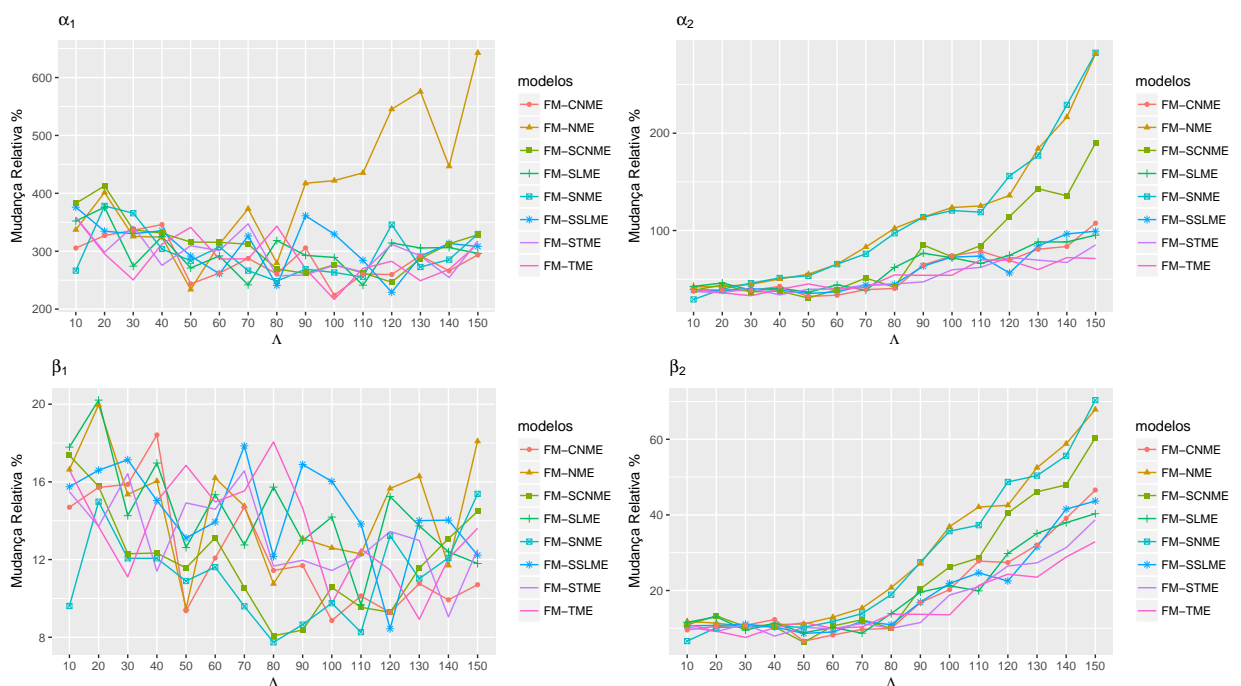


Figura 4.2: RC (%) para as estimativas de α_1 , α_2 , β_1 , β_2 , variando em Λ .

A Figura 4.2 apresenta as mudanças relativas para as estimativas dos parâmetros α_1 , α_2 , β_1 e β_2 nos ajustes $FM-SMSNME$. Observe que em α_1 os ajustes seguem bem próximos, com exceção no ajuste $FM-NME$, ultrapassando RC de 400% a partir de $\Lambda = 90$. Em β_1 o RC assume no máximo 20% de variação, aparentemente os ajustes apresentam comportamento similar para estimativa desse parâmetro. Para α_2 os ajustes $FM-SNME$ e

FM-NME começam a se distanciar dos demais a partir de $\Lambda = 40$, já para β_2 esse afastamento é evidenciado a partir de $\Lambda = 70$, isto indica que as estimativas sofrem muito mais impacto à medida que as perturbações aumentam.

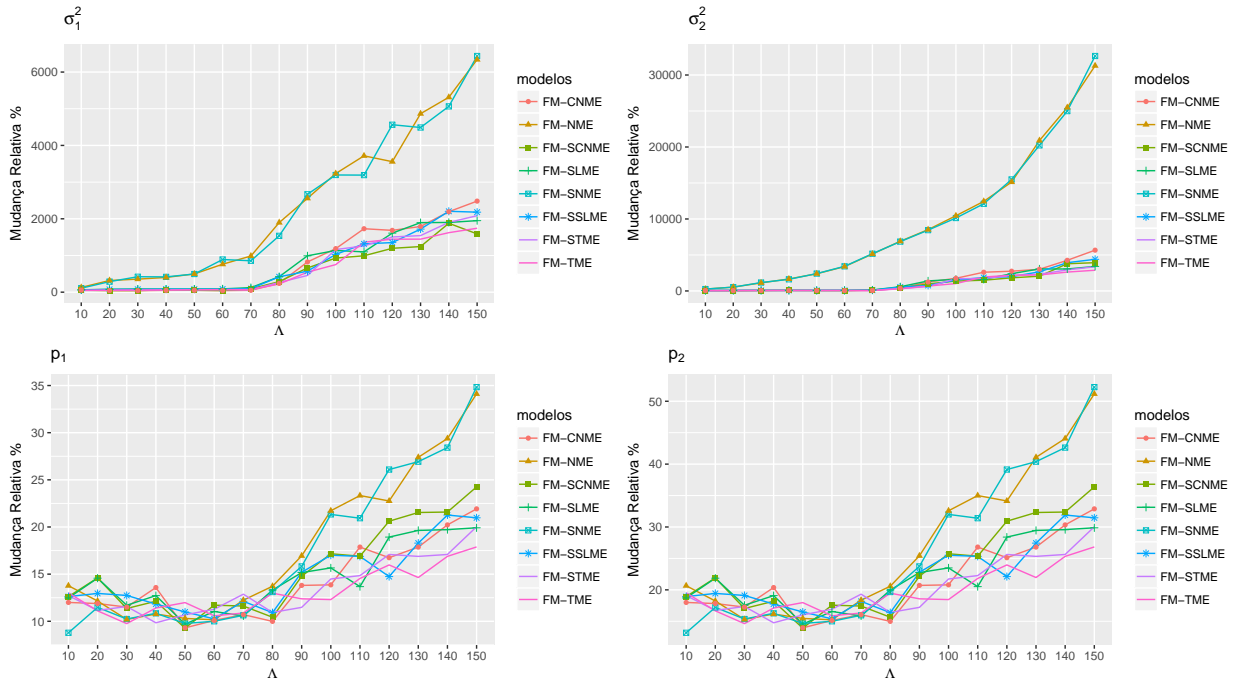


Figura 4.3: RC (%) para as estimativas de σ_1^2 , σ_2^2 e para os pesos p_1 , p_2 , variando em Λ .

Na Figura 4.3 ilustra o comportamento das estimativas para os parâmetros σ_1^2 , σ_2^2 , p_1 e p_2 . Aqui o impacto de Λ é bem mais evidente em relação as estimativas apresentadas na Figura 4.2. As mudanças relativas de σ_1^2 e σ_2^2 nos modelos de caudas pesadas, apresentam RC crescente com $\Lambda = 80$ em diante e se mantem equilibrados entre eles, enquanto que as estimativas nos ajustes *FM-SNME* e *FM-NME* são consideravelmente mais afetadas. O comportamento para σ_2^2 é similar. As estimativas dos pesos são mais afetadas nos ajustes *FM-SNME* e *FM-NME*.

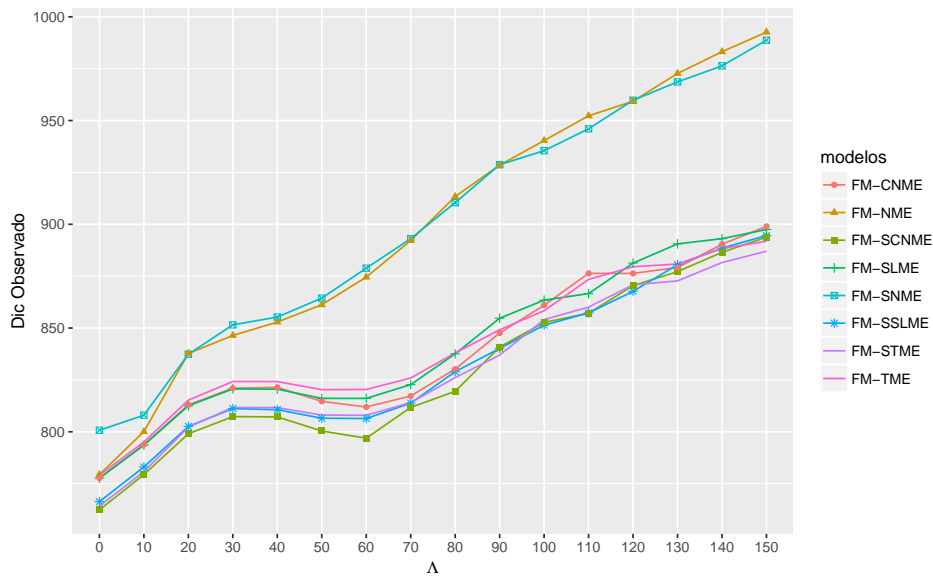


Figura 4.4: *Dic observado para os ajustes FM-SMSNME, variando em Λ .*

A Figura 4.4 apresenta a média dos critérios DIC observado para os ajustes *FM-SMSNME*. Como podemos observar, os modelos de caudas pesadas são bem mais flexíveis em comparação aos ajustes *FM-SNME* e *FM-NME*. Dentre os melhores ajustes, destacamos os modelos *FM-SCNME*, *FM-SSLME* e *FM-STME*.

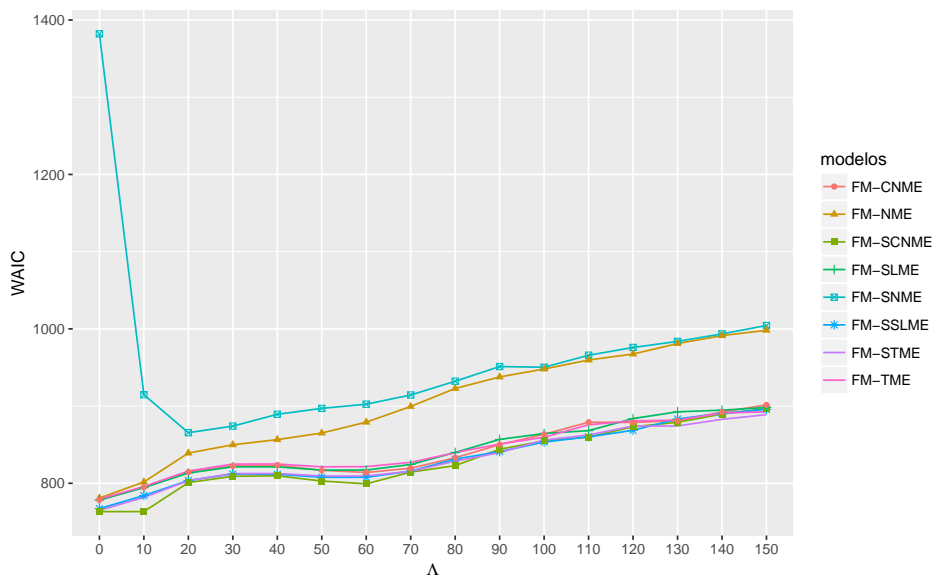


Figura 4.5: *WAIC para os ajustes FM-SMSNME, variando em Λ .*

Observamos na Figura 4.5 a média dos WAIC's nos ajustes *FM-SMSNME*. Assim como na Figura 4.4, os ajustes *FM-SCNME*, *FM-SSLME* e *FM-STME* apresentaram os melhores desempenhos.

4.2 Aplicação em Dados Reais

Para mostrar a utilidade dos métodos propostos, utilizamos o conjunto de dados de Cohen (1984), que representa a percepção de tons por um músico experiente. No experimento de percepção de tons, um som fundamental e puro² com toques gerados eletronicamente foi tocado para o músico ouvir. A missão do músico, era afinar uma nota oitava acima³ do som fundamental mediante a sua percepção do tom. O tom percebido foi gravado em relação ao tom real, em 150 tentativas do mesmo músico. Os significados foram determinados por uma relação de alongamento, que é a relação entre o tom afinado e o tom fundamental. Duas tendências separadas emergem claramente, veja a Figura 4.6 (a). Muitos artigos analisaram este conjunto de dados usando uma mistura de regressões lineares, veja DeVeaux (1989), Viele & Tong (2002) e Hunter & Young (2012). Esses dados foram analisados recentemente por Yao *et al.* (2014) e Zeller *et al.* (2016) levando-os a propor uma mistura de regressão robusta usando as distribuições t de Student e *SMSN*, respectivamente.

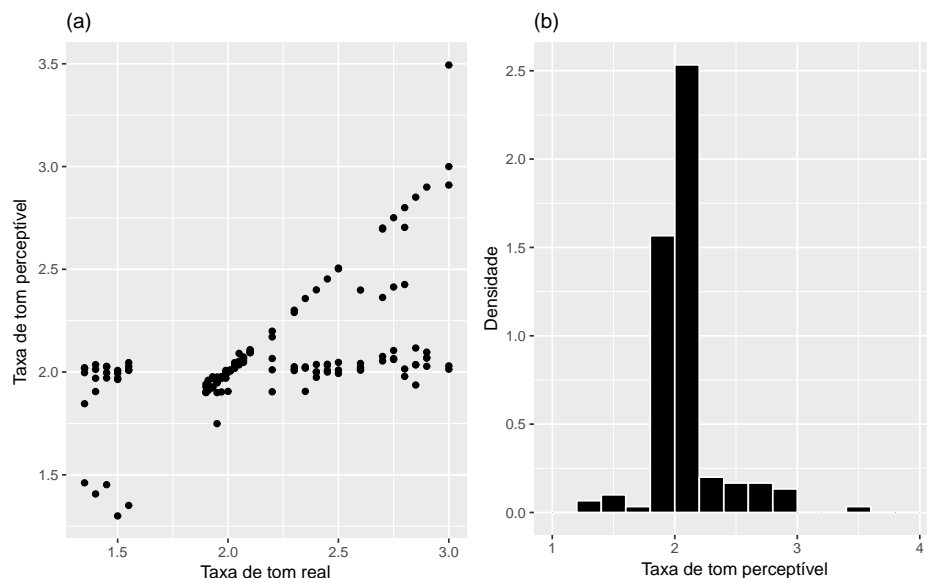


Figura 4.6: *Conjunto de dados percepção de tom. (a) Diagrama de dispersão e (b) Histograma.*

Observe que na Figura 4.6 (a) emergem duas tendências separadas no gráfico de dispersão. Para verificar a existência de assimetria e caudas pesadas nos dados, a Figura

²Também conhecido por frequência fundamental, é um som complexo composto por uma soma de sons de diversas frequências, caracterizado por um som grave de altura mais baixa.

³Uma nota está uma oitava acima significa dizer que a nota é a mesma, porém ela está em uma região mais aguda do instrumento.

4.6 (b) apresenta um histograma da variável taxa de tom perceptível e revela uma leve assimetria.

Para essa situação, ajustamos os modelos *FM-SMSNME* considerando duas componentes, isto é, $r = 1$ e $G = 2$. É bem sabido que o modelo com erro de medição não é identificável, como comentamos anteriormente, existem alguns pressupostos clássicos na literatura que o tornam identificável. Por esta razão, vamos supor que razão de confiabilidade, definida em (3.1), é conhecida. Como em Yao *et al.* (2014), nós assumiremos $\kappa_x = 0.7$.

Também em Yao *et al.* (2014), o erro de medição foi introduzido adicionando um erro aleatório ao preditor. Por isso, geramos uma amostra $\zeta_i \sim \text{NIG}(2.2, -2, 0.1, 0.22)$ da distribuição normal inversa Gaussiana (NIG) e adicionamos a \mathbf{x}_i . A ideia é introduzir erros aleatórios com um forte padrão de assimetria e caudas pesadas, mas com uma distribuição diferente da *SMSN*. Se uma variável aleatória $Z \sim \text{NIG}(\alpha, \beta, \delta, \mu)$, então sua densidade é dada por

$$f(z) = \exp(\delta\sqrt{\alpha^2 - \beta^2}) \frac{\alpha\delta}{\pi\sqrt{\delta^2 + (z - \mu)^2}} K_1(\alpha\sqrt{\delta^2 + (z - \mu)^2}) \exp(\beta(z - \mu)),$$

em que, $K_1(\cdot)$ é a função Bessel modificada do terceiro tipo com a ordem 1, $\alpha > 0$, $\delta > 0$ e $-\alpha < \beta < \alpha$. Então, temos $E(Z) = \mu + \beta\eta$ e $\text{Var}(Z) = \eta + \beta^2\eta^2/\omega$, onde $\eta = \delta/\sqrt{\alpha^2 - \beta^2}$ e $\omega = \delta\sqrt{\alpha^2 - \beta^2}$. A amostra foi gerada usando o software R com pacote **GeneralizedHyperbolic** (Scott, 2015). Mais detalhes sobre a distribuição NIG podem ser encontrados em Paoletta (2007). A Figura 4.7 mostra o histograma do erro aleatório gerado da distribuição NIG.

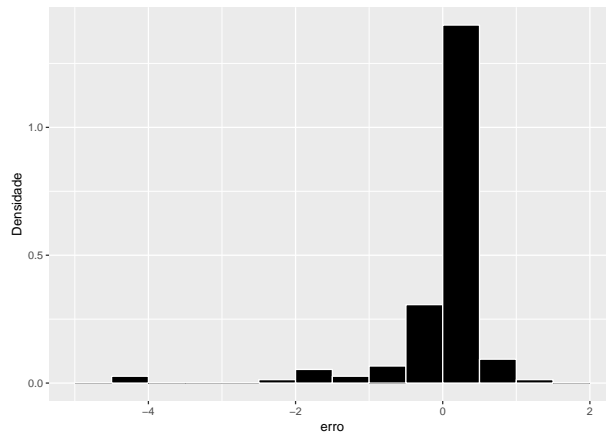


Figura 4.7: Histograma do erro aleatório com distribuição NIG.

O procedimento da modelagem é similar ao processo de simulação. Primeiramente, adicionamos um erro de medição à covariável Taxa de tom real, como citamos anteriormente. Para estimar os parâmetros nos modelos *FM-SMSNME* foi gerada uma amostra Gibbs com 50 mil iterações, e foram excluídas as 10 mil primeiras observações, sendo guardadas as observações geradas de 10 em 10 iterações, na tentativa de dirimir o efeito da autocorrelação da cadeia. Na sequência, podem ser visualizados os traceplots de alguns ajustes, como *FM-STME*, *FM-SSLME* e *FM-SCNME*.

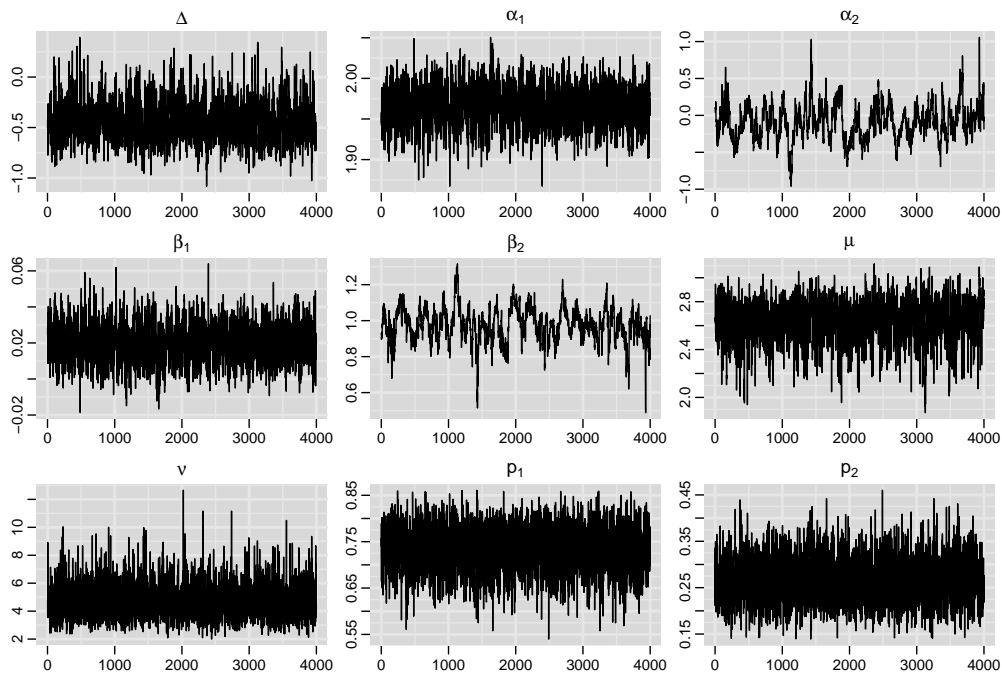


Figura 4.8: Traceplots das amostras MCMC no ajuste *FM-STME*.

A Figura 4.8 ilustra a trajetória de alguns parâmetros no ajuste *FM-STME*. Podemos visualizar uma trajetória homogênea para as estimativas do parâmetro de assimetria Δ , dos coeficientes de regressão α_1 , α_2 , β_1 e β_2 , dos pesos p_1 e p_2 .

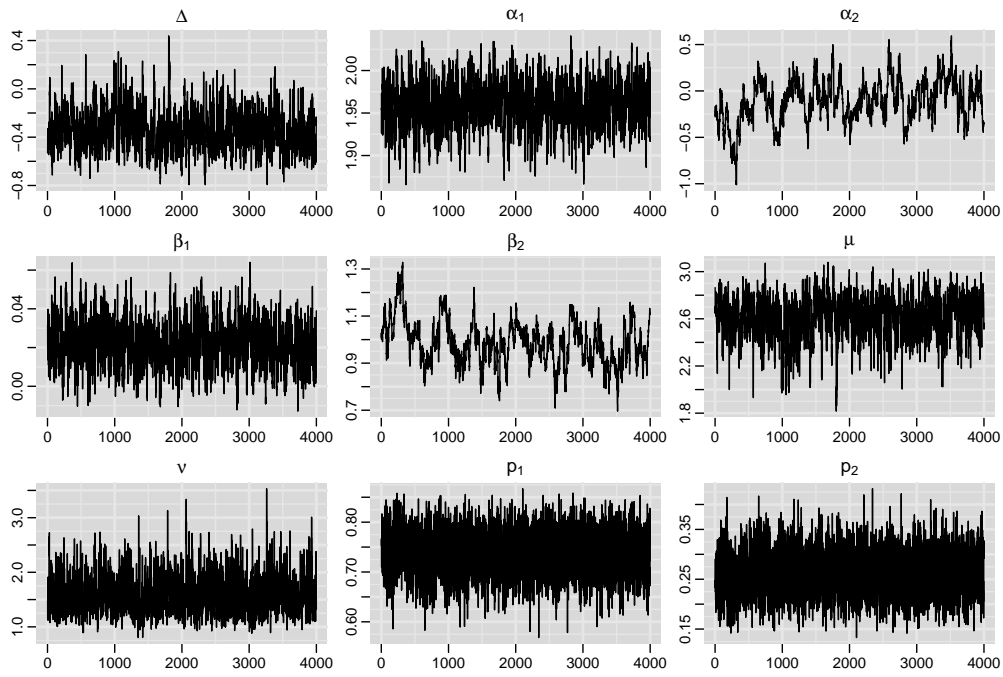


Figura 4.9: Traceplots das amostras MCMC no ajuste FM-SSLME.

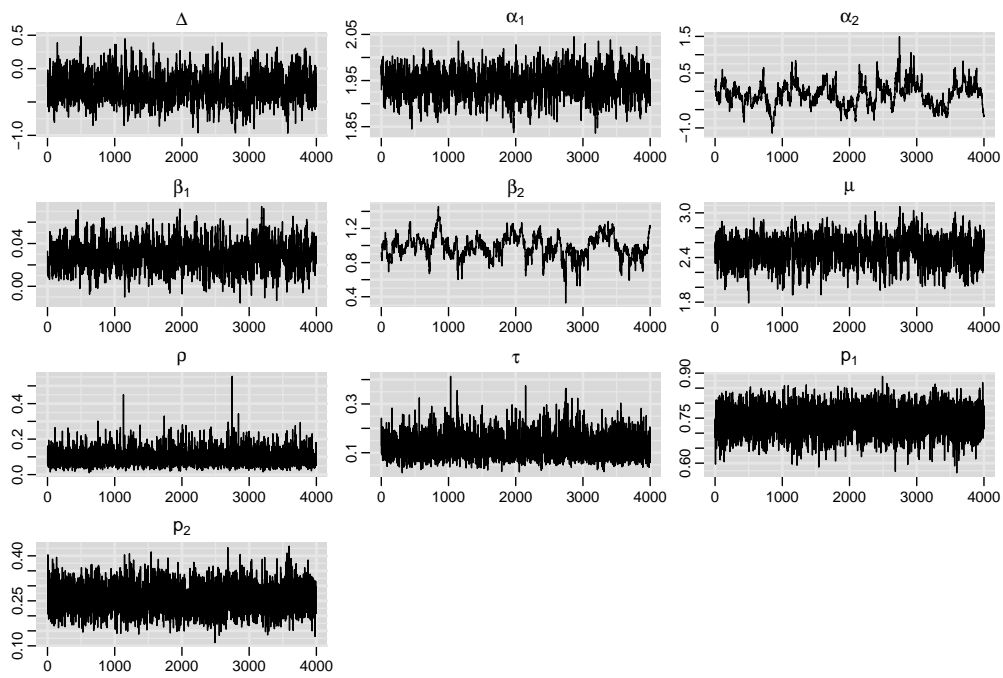


Figura 4.10: Traceplots das amostras MCMC no ajuste FM-SCNME.

Já as Figuras 4.9 e 4.10 mostram as trajetórias das estimativas nos ajustes FM-SSLME e FM-SCNME.

Tabela 4.1: Estimativas MCMC para os parâmetros nos ajustes dos modelos *FM-SMSNME* para o conjunto de dados Cohen (1984).

	Modelos							
	FM-SNME		FM-STME		FM-SSLME		FM-SCNME	
	Média	Dp	Média	Dp	Média	Dp	Média	Dp
Δ	-1,0737	0,1340	-0,4408	0,1944	-0,3652	0,1753	-0,2825	0,2044
Ω	0,0747	0,1003	0,1555	0,0431	0,1036	0,0315	0,1914	0,0350
α_1	1,9979	0,2990	1,9629	0,2531	1,9588	0,0262	1,9411	0,0283
α_2	1,1672	0,0227	-0,0619	0,0245	-0,0316	0,2463	-0,1053	0,2490
β_1	0,0043	0,0093	0,0209	0,1016	0,0222	0,0113	0,0293	0,0122
β_2	0,4749	0,1231	0,9634	0,0105	0,9522	0,0990	0,9896	0,1031
μ	3,0228	0,1093	2,6302	0,1674	2,6323	0,1789	2,4927	0,1706
σ_1^1	0,0047	0,1107	0,0045	0,0073	0,0013	0,0003	0,0025	0,0004
σ_1^2	0,1589	0,0835	0,0018	0,0004	0,0038	0,0071	0,0108	0,0170
p_1	0,7522	0,0416	0,7358	0,0453	0,7396	0,0454	0,7415	0,0437
p_2	0,2478	0,0416	0,2642	0,0453	0,2604	0,0454	0,2585	0,0437
v	-	-	4,6575	1,2259	1,6018	0,3434	-	-
ρ	-	-	-	-	-	-	0,0756	0,0304
τ	-	-	-	-	-	-	0,1222	0,0438

	FM-NME		FM-TME		FM-SLME		FM-CNME	
	Média	Dp	Média	Dp	Média	Dp	Média	Dp
	Ω	0,4058	0,09582	0,2106	0,0347	0,1375	0,0251	0,2105
α_1	1,9950	0,0263	1,9563	0,0286	-0,1450	0,2142	1,9339	0,0303
α_2	0,1247	0,38235	-0,0679	0,2178	1,9477	0,0288	-0,1223	0,3102
β_1	0,0058	0,01094	0,0238	0,0125	1,0070	0,0892	0,0321	0,0132
β_2	0,9335	0,16532	0,9739	0,0899	0,0270	0,0125	1,0017	0,1292
μ	2,1770	0,06238	2,2537	0,0519	2,2568	0,0499	2,2606	0,0484
σ_1^1	0,0373	0,1616	0,0017	0,0004	0,0021	0,0025	0,0025	0,0004
σ_1^2	0,0021	0,0025	0,0031	0,0041	0,0012	0,0003	0,0108	0,0170
p_1	0,7518	0,0409	0,7254	0,0466	0,2642	0,0446	0,7601	0,0426
p_2	0,2482	0,0409	0,2747	0,0466	0,7358	0,0446	0,2399	0,0426
v	-	-	4,7335	1,3445	1,5621	0,3274	-	-
ρ	-	-	-	-	-	-	0,0756	0,0304
τ	-	-	-	-	-	-	0,0438	0,0278

A Tabela 4.1 contém os valores para a média e desvio padrão (Dp) das estimativas MCMC, nos ajustes *FM-SNME*, *FM-STME*, *FM-SSLME*, *FM-SCNME*, *FM-NME*, *FM-*

TME, FM-SLME e FM-CNME.

Tabela 4.2: Critérios de seleção para os modelos *FM-SMSNME* para o conjunto de dados Cohen (1984). A sigla LV denota log-verossimilhança.

	FM-SNME	FM-STME	FM-SSLME	FM-SCNME	FM-NME	FM-TME	FM-SLME	FM-CNME
LV	-61,3217	-51,4535	-49,9653	-49,4891	-72,6994	-53,3180	-51,8691	-52,0098
DIC_{obs}	137,9573	113,7793	110,4559	109,5833	166,7974	115,2278	112,5698	113,2398
WAIC	146,1873	114,7382	111,4406	111,3164	166,0062	115,7825	113,0691	113,7790
$\tau_{DIC_{obs}}$	15,3139	10,8724	10,5253	10,6051	21,3987	8,5918	8,8317	9,2201
τ_{WAIC}	19,4289	11,3518	11,0176	11,4717	21,0031	8,8691	9,0813	9,4898

Podemos observar na Tabela 4.2 que os modelos *FM-STME*, *FM-SSLME* e *FM-SCNME* apresentam os menores valores para os critérios de seleção, comprovando que os modelos assimétricos e de caudas pesadas são mais eficientes. Note ainda, que os ajustes *FM-SSLME* e *FM-SCNME* que estão em negrito, destacaram-se como os melhores, apresentando os menores DIC e WAIC. Os modelos *FM-SNME* e *FM-NME* apresentaram os maiores critérios, além disso, superestimaram o número de parâmetros nos modelos *FM-SMSNME*.

As Figuras 4.11 e 4.12 mostram os gráficos de dispersão para o conjunto Percepção de tom, juntamente com as retas estimadas. Na sequência mostramos a Tabela 4.3 com os intervalos de credibilidade para os modelos *FM-SSLME* e *FM-SCNME* considerados com os melhores ajustes.

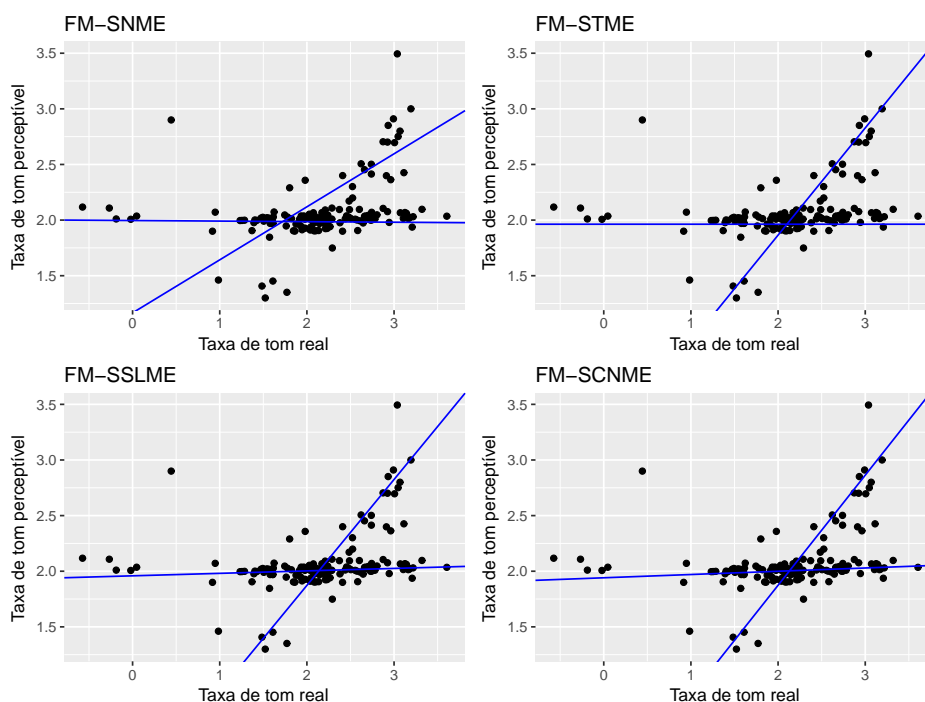


Figura 4.11: *Dados percepção de tons. Diagramas de dispersão com os modelos FM-SNME, FM-STME, FM-SSLME e FM-SCNME ajustados.*

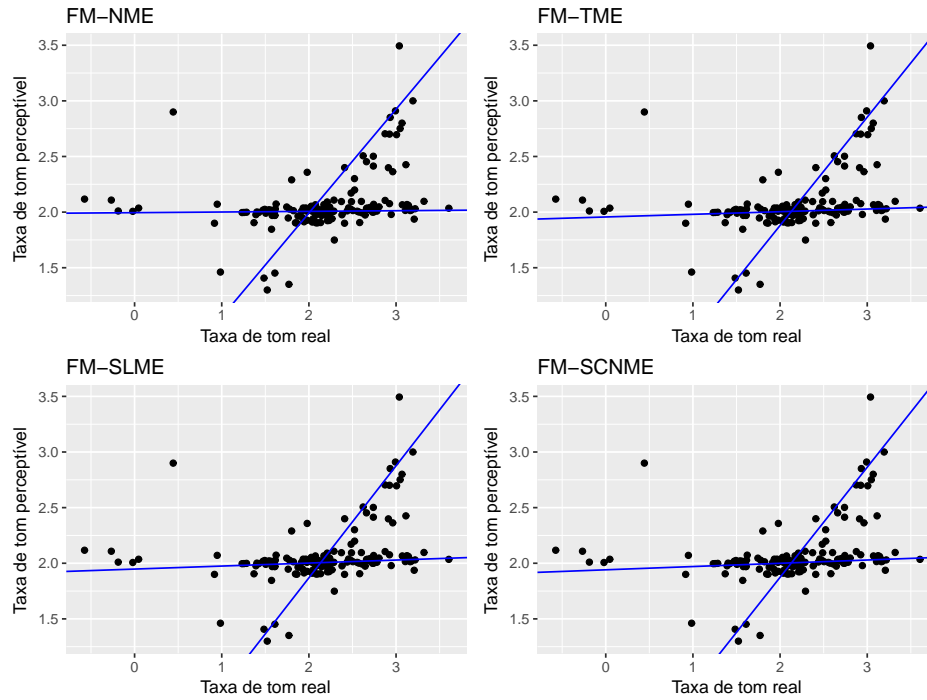


Figura 4.12: Dados percepção de tons. Diagramas de dispersão com os modelos FM-NME, FM-TME, FM-SLME e FM-SCNME ajustados.

Tabela 4.3: Intervalos de credibilidade (IC) nos ajustes FM-SSLME e FM-SCNME para o conjunto de dados Cohen (1984), com 95% de credibilidade.

Parâmetro	FM-SSLME		FM-SCNME	
	Limite inferior	Limite superior	Limite inferior	Limite superior
Δ	-0,6792	0,0221	-0,6480	0,1730
Ω	0,0445	0,1664	0,1211	0,2583
α_1	1,9097	2,0117	1,8861	1,9966
α_2	-0,5145	0,4232	-0,5796	0,3324
β_1	-0,0005	0,0437	0,0065	0,0540
β_2	0,7714	1,1509	0,8054	1,1816
μ	2,2772	2,9881	2,1252	2,8065
σ_1^2	0,0007	0,0019	0,0013	0,0029
σ_2^2	0,00002	0,0141	0,00002	0,0204
p_1	0,6515	0,8284	0,6544	0,8205
p_2	0,1716	0,3485	0,1795	0,3456
ν	1,0030	2,2419	-	-
ρ	-	-	0,0287	0,1838
τ	-	-	0,0471	0,2083

Capítulo 5

Considerações Finais

Nesse trabalho apresentamos uma classe inédita de modelos amplamente flexível em relação ao modelo clássico de mistura de regressão linear. A nova proposta considera que, para cada componente da mistura os erros aleatórios e as covariáveis tem uma distribuição que está na classe das misturas de escala da normal assimétrica (*SMSN*). Outra característica importante, é que, a covariável foi observada com um erro aditivo.

Sob esse contexto, desenvolvemos um algoritmo tipo Gibbs para estimar os parâmetros de cada modelo. Um estudo de simulação, com 100 réplicas, foi realizado com a finalidade de avaliar a eficácia do método proposto. Os resultados mostraram que os modelos assimétricos e de caudas pesadas apresentam um desempenho superior em relação aos ajustes com os modelos que não levam estas duas características em consideração ao mesmo tempo.

A fim de tornar a análise completa, aplicamos os modelos em um conjunto de dados reais de Cohen (1984) - percepção de tons. Aqui verificamos que os modelos utilizando as distribuições slash assimétrica e normal contaminada assimétrica tem ajuste superior em relação aos demais.

Capítulo 6

Apêndice A

6.1 Detalhes do Algoritmo Gibbs

O desenvolvimento das distribuições condicionais completas são dadas a seguir.

Passo 3: A distribuição condicional completa da variável T_i , com $i = 1, \dots, n$, foi desenvolvida a partir da representação hierárquica (18). Como sendo:

$$\pi(T_i | \mathbf{Z}_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \propto \pi(\mathbf{Z}_i | T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \pi(T_i | U_i)$$

Assim,

$$\begin{aligned} &\propto \pi(\mathbf{Z}_i | T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \pi(T_i | U_i) \\ &\propto \exp \left\{ -\frac{1}{2} [\mathbf{z}_i - (\mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i)]^\top (u_i^{-1} \boldsymbol{\Psi}_j)^{-1} [\mathbf{z}_i - (\mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i)] \right\} \\ &\quad \times \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} \mathbb{I}_{(0, \infty)}(t_i) \\ &\propto \exp \left\{ -\frac{u_i}{2} [(\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu}) - \boldsymbol{\Lambda}_j t_i]^\top \boldsymbol{\Psi}_j^{-1} [(\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu}) - \boldsymbol{\Lambda}_j t_i] + t_i^2 \right\} \mathbb{I}_{(0, \infty)}(t_i) \\ &= \exp \left\{ -\frac{u_i}{2} \left[-2 \boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu}) t_i + (\boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j + 1) t_i^2 \right] \right\} \mathbb{I}_{(0, \infty)}(t_i) \end{aligned}$$

Portanto,

$$T_i | \mathbf{Z}_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j \sim \text{TN} \left(\frac{\boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \boldsymbol{\theta}_j)}{1 + \boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j}, \frac{1}{u_i (1 + \boldsymbol{\Lambda}_j^\top \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j)}, (0, \infty) \right),$$

em que, $\boldsymbol{\theta}_j = \mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu}$, $\boldsymbol{\Lambda}_j = \mathbf{B}_j \boldsymbol{\Delta}$ e $\boldsymbol{\Psi}_j = \mathbf{B}_j \boldsymbol{\Omega} \mathbf{B}_j^\top + \boldsymbol{\Gamma}_j$.

Passo 4: Apresentamos as condicionais completas quando $U_i \sim \text{Gamma}(v/2, v/2)$ ou $U_i \sim \text{Beta}(v, 1)$ com $i = 1, \dots, n$. Para o desenvolvimento das respectivas condicionais completas usamos a representação hierárquica (18).

Passo 4 (a): Quando $\pi(U_i) \sim \text{Gamma}(v/2, v/2)$, temos a condicional completa dada por

$$\begin{aligned}
\pi(U_i | \mathbf{Z}_i, T_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) &\propto \pi(\mathbf{Z}_i | T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \pi(T_i | U_i) \pi(U_i) \\
&\propto \exp \left\{ -\frac{1}{2} [\mathbf{z}_i - (\mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i)]^\top (u_i^{-1} \boldsymbol{\Psi}_j^{-1}) [\mathbf{z}_i - (\mathbf{a}_j + \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i)] \right\} \\
&\quad \times \frac{2}{u_i^{-1/2}} \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} u_i^{\frac{v}{2}-1} \exp \left\{ -\frac{v}{2} u_i \right\} \mathbb{I}_{(0, \infty)}(u_i) \\
&\propto u_i^{\frac{p}{2}} \exp \left\{ -\frac{u_i}{2} [(\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} - \boldsymbol{\Lambda}_j t_i)^\top \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} - \boldsymbol{\Lambda}_j t_i)] \right\} u_i^{\frac{1}{2}} \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} \\
&\quad \times u_i^{\frac{v}{2}-1} \exp \left\{ -\frac{v}{2} u_i \right\} \mathbb{I}_{(0, \infty)}(u_i) \\
&= u_i^{\frac{v+p+1}{2}-1} \exp \left\{ -\frac{u_i}{2} [(\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} - \boldsymbol{\Lambda}_j t_i)^\top \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} - \boldsymbol{\Lambda}_j t_i) + t_i^2 + v] \right\} \mathbb{I}_{(0, \infty)}(u_i)
\end{aligned}$$

Logo,

$$U_i | \mathbf{Z}_i, T_i, \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j \sim \text{Gama} \left((v + p + 1)/2, (1/2)(\lambda_i^{(j)} + v) \right),$$

onde

$$\lambda_i^{(j)} = (\mathbf{z}_i - \boldsymbol{\theta}_j - \boldsymbol{\Lambda}_j t_i)^\top \boldsymbol{\Psi}_j^{-1} (\mathbf{z}_i - \boldsymbol{\theta}_j - \boldsymbol{\Lambda}_j t_i) + t_i^2.$$

Passo 4 (b): Quando $\pi(U_i) \sim \text{Beta}(v, 1)$ temos a seguinte condicional completa

$$\begin{aligned}
\pi(\mathbf{Z}_i | T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \pi(T_i | U_i) \pi(U_i) & \\
&\propto \exp \left\{ -\frac{u_i}{2} [\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i]^\top \boldsymbol{\Psi}_j^{-1} [\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i] \right\} \\
&\quad \times \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} u_i^{v-1} \mathbb{I}_{(0, \infty)}(u_i) \\
&\propto u_i^{\frac{p+1}{2}} \exp \left\{ -\frac{u_i}{2} [\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i]^\top \boldsymbol{\Psi}_j^{-1} [\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \boldsymbol{\mu} + \boldsymbol{\Lambda}_j t_i] + t_i^2 \right\} u_i^{v-1} \mathbb{I}_{(0, \infty)}(u_i) \\
&\propto u_i^{\frac{p+1}{2} + v - 1} \exp \left\{ -\frac{u_i}{2} \lambda_i^{(j)} \right\} \mathbb{I}_{(0, \infty)}(u_i) \\
&= u_i^{\frac{2v+p+1}{2}-1} \exp \left\{ -\frac{u_i}{2} \lambda_i^{(j)} \right\} \mathbb{I}_{(0, \infty)}(u_i)
\end{aligned}$$

Portanto,

$$U_i | \mathbf{Z}_i, T_i, \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j \sim \text{Gamma Truncada} \left((2\nu + p + 1)/2, (\lambda_i^{(j)}/2), (0, 1) \right).$$

Passo 5: A condicional completa da variável latente \mathbf{X}_i com $i = 1, \dots, n$, foi obtida através da representação hierárquica (17). Assim, temos:

$$\pi(\mathbf{X}_i | \mathbf{Z}_i, T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \propto \pi(\mathbf{Z}_i | \mathbf{X}_i, T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j) \pi(\mathbf{X}_i | T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\Omega})$$

Assim,

$$\begin{aligned} &\propto \pi(\mathbf{Z}_i | \mathbf{X}_i, T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}_j, \boldsymbol{\Omega}, \omega_i = j) \pi(\mathbf{X}_i | T_i, U_i, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) \\ &\propto \exp \left\{ -\frac{u_i}{2} \left[(\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i) \right] \right\} \exp \left\{ -\frac{u_i}{2} \left[(\mathbf{x}_i - \boldsymbol{\mu} - \boldsymbol{\Delta} t_i)^\top \boldsymbol{\Omega}^{-1} (\mathbf{x}_i - \boldsymbol{\mu} - \boldsymbol{\Delta} t_i) \right] \right\} \\ &\propto \exp \left\{ -\frac{u_i}{2} \left[-2(\mathbf{z}_i - \mathbf{a}_j)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j \mathbf{x}_i + \mathbf{x}_i^\top \mathbf{B}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j \mathbf{x}_i \right] + \left[-2(\mathbf{x}_i \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} + \boldsymbol{\Delta}^\top \boldsymbol{\Omega}^{-1} t_i \mathbf{x}_i) + \mathbf{x}_i^\top \boldsymbol{\Omega}^{-1} \mathbf{x}_i \right] \right\} \\ &= \exp \left\{ -\frac{u_i}{2} \left[-2\mathbf{x}_i \left((\mathbf{z}_i - \mathbf{a}_j)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j + \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} + \boldsymbol{\Delta}^\top \boldsymbol{\Omega}^{-1} t_i \right) + \mathbf{x}_i^\top (\mathbf{B}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j + \boldsymbol{\Omega}^{-1}) \mathbf{x}_i \right] \right\} \end{aligned}$$

Portanto,

$$\mathbf{X}_i | \mathbf{Z}_i, T_i, \boldsymbol{\mu}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j \sim N_p(\boldsymbol{\mu}_{\mathbf{x}_i}, u_i^{-1} \boldsymbol{\Sigma}_{\mathbf{x}_i}),$$

em que,

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{x}_i} &= \boldsymbol{\Sigma}_{\mathbf{x}_i} \left((\mathbf{z}_i - \mathbf{a}_j)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j + \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} + \boldsymbol{\Delta}^\top \boldsymbol{\Omega}^{-1} t_i \right) \\ \boldsymbol{\Sigma}_{\mathbf{x}_i} &= \left(\mathbf{B}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{B}_j + \boldsymbol{\Omega}^{-1} \right)^{-1}. \end{aligned}$$

Passo 6: A condicional completa do vetor de coeficientes $\boldsymbol{\phi} = (\boldsymbol{\mu}^\top, \boldsymbol{\Delta}^\top)^\top$ foi desenvolvida usando a seguinte representação hierárquica:

$$\begin{aligned} \mathbf{Z}_i^* | u_i, t_i, \omega_i &\sim N_{p+1}(u_i^{1/2} \mathbf{B}_{\omega_i} \boldsymbol{\mu} + u_i^{1/2} t_i \mathbf{B}_{\omega_i} \boldsymbol{\Delta}, \boldsymbol{\Psi}_{\omega_i}) \\ \mathbf{Z}_i^* | \boldsymbol{\phi}, u_i, t_i, \omega_i, \dots &\sim N_p(\mathbf{W}_i^* \boldsymbol{\phi}, \boldsymbol{\Psi}_{\omega_i}), \end{aligned}$$

isto é, temos um modelo de regressão multivariado, sendo $\boldsymbol{\phi}$ um vetor comum de coefici-

entes de regressão para todo $i = 1, \dots, n$. Então, temos a seguinte condicional completa:

$$\pi(\boldsymbol{\phi} | \mathbf{Z}^*, \mathbf{U}, \mathbf{T}, \mathbf{W}^*, \mathbf{a}_\phi, \mathbf{A}_\phi, \boldsymbol{\omega}) \propto \pi(\mathbf{Z}^* | \boldsymbol{\phi}, \mathbf{U}, \mathbf{T}, \mathbf{W}^*, \mathbf{a}_\phi, \mathbf{A}_\phi, \boldsymbol{\omega}) \pi(\boldsymbol{\phi})$$

Assim temos:

$$\begin{aligned} &\propto \pi(\mathbf{Z}^* | \boldsymbol{\phi}, \mathbf{U}, \mathbf{T}, \mathbf{W}^*, \mathbf{a}_\phi, \mathbf{A}_\phi, \boldsymbol{\omega}) \pi(\boldsymbol{\phi}) \\ &\propto \prod_{i=1}^n \exp \left\{ -\frac{1}{2} (\mathbf{z}_i^* - \mathbf{W}_i^* \boldsymbol{\phi})^\top \boldsymbol{\Psi}_{\omega_i}^{-1} (\mathbf{z}_i^* - \mathbf{W}_i^* \boldsymbol{\phi}) \right\} \times \exp \left\{ -\frac{1}{2} (\boldsymbol{\phi} - \mathbf{a}_\phi)^\top \mathbf{A}_\phi^{-1} (\boldsymbol{\phi} - \mathbf{a}_\phi) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{z}_i^* - \mathbf{W}_i^* \boldsymbol{\phi})^\top \boldsymbol{\Psi}_{\omega_i}^{-1} (\mathbf{z}_i^* - \mathbf{W}_i^* \boldsymbol{\phi}) + (\boldsymbol{\phi} - \mathbf{a}_\phi)^\top \mathbf{A}_\phi^{-1} (\boldsymbol{\phi} - \mathbf{a}_\phi) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[-2 \sum_{i=1}^n \left(\mathbf{z}_i^* (\mathbf{W}_i^*)^\top \boldsymbol{\phi}^\top \boldsymbol{\Psi}_{\omega_i}^{-1} + (\mathbf{W}_i^*)^\top \boldsymbol{\phi}^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{W}_i^* \boldsymbol{\phi} \right) \right] + \left[-2 \boldsymbol{\phi}^\top \mathbf{A}_\phi^{-1} \mathbf{a}_\phi + \boldsymbol{\phi}^\top \mathbf{A}_\phi^{-1} \boldsymbol{\phi} \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[-2 \boldsymbol{\phi}^\top \left(\sum_{i=1}^n (\mathbf{W}_i^*)^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{z}_i^* + \mathbf{A}_\phi^{-1} \mathbf{a}_\phi \right) \right] + \boldsymbol{\phi}^\top \left(\sum_{i=1}^n (\mathbf{W}_i^*)^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{z}_i^* + \mathbf{A}_\phi^{-1} \right) \boldsymbol{\phi} \right\} \end{aligned}$$

Portanto,

$$\boldsymbol{\phi} | \mathbf{Z}^*, \mathbf{U}, \mathbf{T}, \mathbf{W}^*, \mathbf{a}_\phi, \mathbf{A}_\phi, \boldsymbol{\omega} \sim N_p \left(\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\phi \right)$$

em que,

$$\begin{aligned} \boldsymbol{\mu}_\phi &= \boldsymbol{\Sigma}_\phi \left(\sum_{i=1}^n (\mathbf{W}_i^*)^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{z}_i^* + \mathbf{A}_\phi^{-1} \mathbf{a}_\phi \right), \\ \boldsymbol{\Sigma}_\phi &= \left(\sum_{i=1}^n (\mathbf{W}_i^*)^\top \boldsymbol{\Psi}_{\omega_i}^{-1} \mathbf{z}_i^* + \mathbf{A}_\phi^{-1} \right)^{-1}. \end{aligned}$$

Passo 7: Nesse caso apresentamos a condicional completa para α_j com $j = 1, \dots, G$, para seu desenvolvimento usamos a representação hierárquica (17). Seja:

$$\pi(\alpha_j | \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}_j, \boldsymbol{\Gamma}_j, \dots, \boldsymbol{\omega}) \propto \pi(\mathbf{z} | \alpha_j, \mathbf{x}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}_j, \boldsymbol{\Gamma}_j, \dots, \boldsymbol{\omega}) \pi(\alpha_j)$$

Assim, temos:

$$\begin{aligned}
&\propto \pi(\mathbf{z}|\alpha_j, \mathbf{x}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}_j, \boldsymbol{\Gamma}_j, \dots, \boldsymbol{\omega}), \pi(\alpha_j) \\
&\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i) \right\} \exp \left\{ - \frac{1}{2\sigma_\alpha^2} (\alpha_j - \mu_\alpha)^2 \right\} \\
&\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} \left[-2(\mathbf{z}_i - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j + \mathbf{a}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j \right] \right\} \exp \left\{ - \frac{1}{2\sigma_\alpha^2} (\alpha_j^2 - 2\alpha_j \mu_\alpha) \right\}
\end{aligned}$$

Desenvolvendo os termos que estão em função de \mathbf{a}_j , \mathbf{B}_j , \mathbf{z}_i e $\boldsymbol{\Gamma}_j$. Primeiramente para o termo $(\mathbf{z}_i - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j$, temos:

$$\begin{aligned}
(\mathbf{z}_i - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j &= \left[\begin{pmatrix} \mathbf{X}_i \\ Y_i \end{pmatrix} - \begin{pmatrix} \mathbf{I}_r \\ \boldsymbol{\beta}_j^\top \end{pmatrix} \mathbf{x}_i \right]^\top \begin{pmatrix} \boldsymbol{\Gamma}^{-1} & \mathbf{0}_{r \times 1} \\ \mathbf{0}_{1 \times r} & \sigma_j^{-2} \end{pmatrix} \begin{pmatrix} \mathbf{0}_{r \times 1} \\ \alpha_j^\top \end{pmatrix} \\
&= \left[\begin{pmatrix} (\mathbf{X}_i - \mathbf{x}_i) \boldsymbol{\Gamma}^{-1} & (Y_i - \mathbf{B}_j \mathbf{x}_i)^\top \sigma_j^{-2} \end{pmatrix} \right] \begin{pmatrix} \mathbf{0}_{r \times 1} \\ \alpha_j^\top \end{pmatrix} \\
&= \sigma_j^{-2} \alpha_j^\top (Y_i - \mathbf{B}_j \mathbf{x}_i)
\end{aligned}$$

Agora, para o termo $\mathbf{a}_j \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j$, temos:

$$\begin{aligned}
\mathbf{a}_j \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j &= \begin{pmatrix} \mathbf{0}_{r \times 1} & \alpha_j \end{pmatrix} \begin{pmatrix} \boldsymbol{\Gamma}^{-1} & \mathbf{0}_{r \times 1} \\ \mathbf{0}_{1 \times r} & \sigma_j^{-2} \end{pmatrix} \begin{pmatrix} \mathbf{0}_{r \times 1} \\ \alpha_j^\top \end{pmatrix} \\
&= \alpha_j^\top \sigma_j^{-2} \alpha_j,
\end{aligned}$$

então,

$$\begin{aligned}
&\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} \left[-2(\mathbf{z}_i - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j + \mathbf{a}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{a}_j \right] \right\} \exp \left\{ - \frac{1}{2\sigma_\alpha^2} (\alpha_j^2 - 2\alpha_j \mu_\alpha) \right\} \\
&\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} \left[-2\sigma_j^{-2} \alpha_j^\top (Y_i - \boldsymbol{\beta}_j \mathbf{x}_i) + \alpha_j^\top \sigma_j^{-2} \alpha_j \right] \right\} \exp \left\{ - \frac{1}{2\sigma_\alpha^2} (\alpha_j^2 - 2\alpha_j \mu_\alpha) \right\} \\
&= \exp \left\{ - \frac{1}{2} \left[-2\alpha_j^\top \left(\sum_{i \in \mathcal{A}_j} u_i \sigma_j^{-2} (Y_i - \boldsymbol{\beta}_j \mathbf{x}_i) + \sigma_\alpha^{-2} \mu_\alpha \right) + \alpha_j^\top \left(\sigma_j^{-2} \sum_{i \in \mathcal{A}_j} u_i + \sigma_\alpha^{-2} \right) \alpha_j \right] \right\}
\end{aligned}$$

Portanto,

$$\alpha_j | \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}_j, \boldsymbol{\Gamma}_j, \dots, \boldsymbol{\omega} \sim N(\boldsymbol{\mu}_{\alpha_j}, \boldsymbol{\Sigma}_{\alpha_j}),$$

em que,

$$\begin{aligned}\boldsymbol{\mu}_{\alpha_j} &= \boldsymbol{\Sigma}_{\alpha_j} \left(\sum_{i=i}^n u_i \sigma_j^{-2} (Y_i - \boldsymbol{\beta}_j \mathbf{x}_i) + \sigma_{\alpha}^{-2} \boldsymbol{\mu}_{\alpha} \right) \\ \boldsymbol{\Sigma}_{\alpha_j} &= \left(\sigma_j^{-2} \sum_{i=i}^n u_i + \sigma_{\alpha}^{-2} \right)^{-1}\end{aligned}$$

Passo 8: Também obtivemos a condicional completa do vetor $\boldsymbol{\beta}_j$ com $j = 1, \dots, n$ usando a representação hierárquica (17). Seja a condicional completa:

$$\pi(\boldsymbol{\beta}_j | \mathbf{Z}_i, \mathbf{x}_i, U_i, \alpha_j, \boldsymbol{\Gamma}_j, \omega_i = j) \propto \pi(\mathbf{Z}_i | \boldsymbol{\beta}_j, \mathbf{x}_i, U_i, \alpha_j, \boldsymbol{\Gamma}_j, \omega_i = j), \pi(\boldsymbol{\beta}_j)$$

Então,

$$\begin{aligned}&\propto \pi(\mathbf{Z}_i | \mathbf{x}_i, U_i, \alpha_j, \boldsymbol{\beta}_j, \boldsymbol{\Gamma}_j, \omega_i = j) \pi(\mathbf{x}_i | U_i, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) \\ &\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i) \right\} \exp \left\{ - \frac{1}{2} (\boldsymbol{\beta}_j - \mathbf{c}_\beta)^\top \mathbf{C}_\beta^{-1} (\boldsymbol{\beta}_j - \mathbf{c}_\beta) \right\},\end{aligned}$$

desenvolvendo a condicional completa em função de \mathbf{B}_j , \mathbf{z}_i e $\boldsymbol{\Gamma}_j$, temos:

$$\begin{aligned}&\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} \left[\begin{pmatrix} \mathbf{X}_i \\ Y_i \end{pmatrix} - \mathbf{a}_j - \begin{pmatrix} \mathbf{I}_r \\ \boldsymbol{\beta}_j^\top \end{pmatrix} \mathbf{x}_i \right]^\top \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & 1/\sigma_j^2 \end{pmatrix} \left[\begin{pmatrix} \mathbf{X}_i \\ Y_i \end{pmatrix} - \mathbf{a}_j - \begin{pmatrix} \mathbf{I}_r \\ \boldsymbol{\beta}_j^\top \end{pmatrix} \mathbf{x}_i \right] \right\} \\ &\quad \times \exp \left\{ - \frac{1}{2} (\boldsymbol{\beta}_j - \mathbf{c}_\beta)^\top \mathbf{C}_\beta^{-1} (\boldsymbol{\beta}_j - \mathbf{c}_\beta) \right\} \\ &\propto \exp \left\{ - \sum_{i \in \mathcal{A}_j} \frac{u_i}{2} \left[(\mathbf{X}_i - \mathbf{a}_j - \mathbf{x}_i)^\top \boldsymbol{\Gamma}^{-1} (\mathbf{X}_i - \mathbf{a}_j - \mathbf{x}_i) + \frac{1}{\sigma_j^2} [Y_i - (\mathbf{a}_j - \mathbf{x}_i^\top \boldsymbol{\beta}_j)]^2 \right] \right\} \\ &\quad \times \exp \left\{ - \frac{1}{2} (-2\boldsymbol{\beta}_j^\top \mathbf{C}_\beta^{-1} \mathbf{c}_\beta + \boldsymbol{\beta}_j^\top \mathbf{C}_\beta^{-1} \boldsymbol{\beta}_j) \right\} \\ &= \exp \left\{ - \frac{1}{2} \left[-2\boldsymbol{\beta}_j^\top \left(\frac{1}{\sigma_j^2} \sum_{i \in \mathcal{A}_j} u_i \mathbf{x}_i (Y_i - \mathbf{a}_j) + \mathbf{C}_\beta^{-1} \mathbf{c}_\beta \right) + \boldsymbol{\beta}_j^\top \left(\frac{1}{\sigma_j^2} \sum_{i \in \mathcal{A}_j} u_i \mathbf{x}_i \mathbf{x}_i^\top + \mathbf{C}_\beta^{-1} \right) \boldsymbol{\beta}_j \right] \right\}\end{aligned}$$

Portanto,

$$\boldsymbol{\beta}_j | \mathbf{Z}_i, \mathbf{X}_i, U_i, T_i, \alpha_j, \boldsymbol{\Gamma}_j, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\Omega}, \omega_i = j \sim N_p \left(\boldsymbol{\mu}_{\boldsymbol{\beta}_j}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_j} \right)$$

em que,

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{\beta}_j} &= \boldsymbol{\Sigma}_{\boldsymbol{\beta}_j} \left(\frac{1}{\sigma_j^2} \sum_{i \in \mathcal{A}_j} u_i \mathbf{x}_i (Y_i - \mathbf{a}_j) + \mathbf{C}_{\boldsymbol{\beta}}^{-1} \mathbf{c}_{\boldsymbol{\beta}} \right) \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}_j} &= \left(\frac{1}{\sigma_j^2} \sum_{i \in \mathcal{A}_j} u_i \mathbf{x}_i \mathbf{x}_i^\top + \mathbf{C}_{\boldsymbol{\beta}}^{-1} \right)^{-1}.\end{aligned}$$

Passo 9: Como nos passos anteriores, obtivemos a condicional completa do parâmetro γ_j^{-2} com $j = 1, \dots, n$ através da representação hierárquica (17), temos que,

$$\pi(\gamma_j^{-2} | \mathbf{Z}_i, \mathbf{x}_i, U_i, \alpha_j, \boldsymbol{\Gamma}_j, \boldsymbol{\omega}_i = j) \propto \pi(\mathbf{Z}_i | \gamma_j^{-2}, \mathbf{x}_i, U_i, \alpha_j, \boldsymbol{\Gamma}_j, \boldsymbol{\omega}_i = j), \pi(\gamma_j^{-2})$$

Então,

$$\propto \exp \left\{ - \sum_{i=i}^n \frac{u_i}{2} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i)^\top \boldsymbol{\Gamma}_j^{-1} (\mathbf{z}_i - \mathbf{a}_j - \mathbf{B}_j \mathbf{x}_i) \right\} (\gamma_j^{-2})^{l-1} \exp \left\{ -m \gamma_j^{-2} \right\},$$

Usando parte do resultado do passo 7, temos,

$$\begin{aligned}& \propto \exp \left\{ - \sum_{i=i}^n \frac{u_i}{2} \left[(\mathbf{X}_i - \mathbf{a}_j - \mathbf{x}_i)^\top \boldsymbol{\Gamma}^{-1} (\mathbf{X}_i - \mathbf{a}_j - \mathbf{x}_i) + \frac{1}{\sigma_j^2} \left[Y_i - (\mathbf{a}_j - \mathbf{x}_i^\top \boldsymbol{\beta}_j) \right]^2 \right] \right\} \\ & \quad \times (\gamma_j^{-2})^{l-1} \exp \left\{ -m \gamma_j^{-2} \right\} \\ & \propto \det(\boldsymbol{\Gamma})^{-n/2} \exp \left\{ - \sum_{i=i}^n \frac{u_i}{2} \left[(\mathbf{X}_i - \mathbf{a}_j - \mathbf{x}_i)^\top \boldsymbol{\Gamma}^{-1} (\mathbf{X}_i - \mathbf{a}_j - \mathbf{x}_i) \right] \right\} (\gamma_j^{-2})^{l-1} \exp \left\{ -m \gamma_j^{-2} \right\} \\ & \propto (\gamma_j^{-2})^{-\frac{n}{2} + l - 1} \exp \left\{ - \sum_{i=1}^n \frac{u_i}{2} \sum_{k=1}^r \gamma_k^{-2} (\mathbf{X}_{ik} - \mathbf{a}_j - \mathbf{x}_{ik})^2 \right\} \exp \left\{ -m \gamma_j^{-2} \right\} \\ & \propto (\gamma_j^{-2})^{-\frac{n}{2} + l - 1} \exp \left\{ - \gamma_j^{-2} \sum_{i=1}^n \frac{u_i}{2} (\mathbf{X}_{ij} - \alpha_j - \mathbf{x}_{ij})^2 \right\} \exp \left\{ -m \gamma_j^{-2} \right\} \\ & = (\gamma_j^{-2})^{-\frac{n}{2} + l - 1} \exp \left\{ - \gamma_j^{-2} \left[m + \frac{1}{2} \sum_{i=1}^n u_i (\mathbf{X}_{ij} - \alpha_j - \mathbf{x}_{ij}) \right] \right\}\end{aligned}$$

Portanto,

$$\gamma_j^{-2} | \mathbf{Z}_i, \mathbf{x}_i, U_i, \alpha_j, \boldsymbol{\Gamma}_j, \boldsymbol{\omega}_i = j \sim \text{Gamma} \left(\frac{n}{2} + l, \frac{1}{2} \sum_{i=1}^n u_i (\mathbf{X}_{ij} - \alpha_j - \mathbf{x}_{ij})^2 + m \right)$$

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 716–723.
- Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.
- Arellano-Valle, R. B. & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, **33**, 561–574.
- Arellano-Valle, R. B., del Pino, G. & Martín, E. S. (2002). Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters*, **58**, 111 – 121.
- Arellano-Valle, R. B., Castro, L. M., Genton, M. G. & Gómez, H. W. (2008). Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis*, **3**, 513–539.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, **32**, 159–188.
- Azzalini, A. & Capitanio, A. (2014). *The Skew-Normal and Related Families*. Cambridge University Press.
- Bai, X., Yao, W. & Boyer, J. E. (2012). Robust fitting of mixture regression models. *Computational Statistics & Data Analysis*, **56**, 2347–2359.
- Böhning, D. (2000). *Computer-assisted analysis of mixtures and applications. Meta-analysis, disease mapping and others*. Chapman&Hall/CRC, Boca Raton.
- Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V. & Walther, G. (2007). Editorial: Advances in mixture models. *Computational Statistics & Data Analysis*, **51**, 5205–5210.

- Böhning, D., Hennig, C., McLachlan, G. J. & McNicholas, P. D. (2014). Editorial: The 2nd special issue on advances in mixture models. *Computational Statistics & Data Analysis*, **71**, 1–2.
- Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113.
- Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, And Applications*. Chapman & Hall/CRC, Boca Raton.
- Cabral, C. R. B., Lachos, V. H. & Madruga, M. R. (2012a). Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *Journal of Statistical Planning and Inference*, **142**, 181–200.
- Cabral, C. R. B., Lachos, V. H. & Prates, M. O. (2012b). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics and Data Analysis*, **56**, 126–142.
- Cabral, C. R. B., Lachos, V. H. & Zeller, C. B. (2014). Multivariate measurement error models using finite mixtures of skew-Student t distributions. *Journal of Multivariate Analysis*, **124**, 179–198.
- Celeux, G., Forbes, F., Robert, C. P. & Titterton, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, **1**, 651–674.
- Cheng, C. L. & Van Ness, J. W. (1999). *Statistical Regression with Measurement Error: Kendall's Library of Statistics 6*. Wiley, Chichester.
- Cohen, E. A. (1984). Some effects of inharmonic partials on interval perception. *Music Perception: An Interdisciplinary Journal*, **1**(3), 323–349.
- Cosslett, S. R. & Lee, L.-F. (1985). Serial correlation in latent discrete variable models. *Journal of Econometrics*, **27**, 79–97.
- DeSarbo, W. S. & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of classification*, **5**(2), 249–282.
- DeSarbo, W. S., Wedel, M., Vriens, M. & Ramaswamy, V. (1992). Latent class metric conjoint analysis. *Marketing Letters*, **3**, 273–288.
- DeVeaux, R. D. (1989). Mixtures of linear regressions. *Computational Statistics and Data Analysis*, **8**(3), 227–245.

- Fonseca, T. C. O., Ferreira, M. A. R. & Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika*, **95**, 325–333.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Verlag, New York.
- Fuller, W. A. (1987). *Measurement error models*. John Wiley, New York.
- Garay, A. M., Bolfarine, H., Lachos, V. H. & Cabral, C. R. B. (2015). Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. *Journal of Applied Statistics*. doi.org/10.1080/02664763.2015.1048671.
- Genton, M. G., editor (2004). *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*. Chapman & Hall.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, **57**, 357–384.
- Hennig, C. (2004). Breakdown points for maximum likelihood estimators of location-scale mixtures. *Annals of Statistics*, pages 1313–1340.
- Huber, P. (1981). J. 1981. robust statistics. *New York: John Wiley*.
- Hunter, D. R. & Young, D. S. (2012). Semiparametric mixtures of regressions. *Journal of Nonparametric Statistics*, **24**(1), 19–38.
- Hurn, M., Justel, A. & Robert, C. P. (2003). Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, **12**, 55–79.
- Lachos, V., Garibay, V., Labra, F. & Aoki, R. (2009). A robust multivariate measurement error model with skew-normal/independent distributions and bayesian mcmc implementation. *Statistical Methodology*, **6**(5), 527–541.
- Lachos, V. H., Vilca, L. F., Bolfarine, H. & Ghosh, P. (2008). Robust multivariate measurement error models with scale mixtures of skew-normal distribution. *Statistics*.
- Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**, 303–322.
- Lange, K. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.

- Lee, S. & McLachlan, G. J. (2014). Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing*, **24**(2), 181–202.
- Li, Y., Zeng, T. & Yu, J. (2013). Robust deviance information criterion for latent variable models. CAFE research paper no. 13.19, CAFE-University of Southern California.
- Lin, T. I., Lee, J. C. & Hsieh, W. J. (2007). Robust mixture modeling using the skew t distribution. *Statistics and computing*, **17**(2), 81–92.
- Lindsay, B. G. (1995). *Mixture models: theory geometry and applications*, volume 51. NSF-CBMS Regional Conference Series in Probability and Statistics, Institute of Mathematical Statistics, Hayward.
- Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**, 958–966.
- Liu, M. & Lin, T.-I. (2014). A skew-normal mixture regression model. *Educational and Psychological Measurement*, **74**, 139–162.
- Liu, M., Hancock, G. R. & Harring, J. R. (2011). Using finite mixture modeling to deal with systematic measurement error: A case study. *Journal of Modern Applied Statistical Methods*, **10**, 249–261.
- Massuia, M. B., Garay, A. M., Cabral, C. R. & Lachos, V. (2017). Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions. *Statistics and Its Interface*, **10**, 425–439.
- McLachlan, G. J. & Peel, D. (2000). *Finite mixture models*. Wiley, New York.
- Mengersen, K., Robert, C. P. & Titterton, D. M. (2011). *Mixtures: Estimation and Applications*. John Wiley and Sons.
- Paoletta, M. S. (2007). *Intermediate Probability: A Computational Approach*. Wiley, Chichester.
- Plummer, M. (2016). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-6.
- Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T. I., Maier, L. M., Baecher-Allan, C., McLachlan, G. J., Tamayo, P., Hafler, D. A., De Jager, P. L. & Mesirov, J. P. (2009). Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences*, **106**, 8519–8524.

- Quandt, R. E. & Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, **73**, 730–738.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Richardson, S. & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, **59**, 731–792.
- Scott, D. (2015). *GeneralizedHyperbolic: The Generalized Hyperbolic Distribution*. R package version 0.8-1.
- Song, W., Yao, W. & Xing, Y. (2014). Robust mixture regression model fitting by Laplace distribution. *Computational Statistics & Data Analysis*, **71**, 128–137.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Der, L. V. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical, Series B*, **64**, 583–639.
- Stephens, M. (1997). *Bayesian methods for mixtures of normal distributions*. Ph.D. thesis, Magdalen College, Oxford.
- Turner, T. R. (2000). Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **49**(3), 371–384.
- Viele, K. & Tong, B. (2002). Modeling with mixtures of linear regressions. *Statistics and Computing*, **12**, 315–330.
- Wang, J. & Genton, M. G. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**(1), 209–220.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, **11**, 3571–3594.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-0-387-98140-6.
- Yao, W. & Song, W. (2014). Mixtures of linear regression with measurement errors. *Communications in Statistics-Theory and Methods*, **44**, 1602–1614.

Yao, W., Wei, Y. & Yu, C. (2014). Robust mixture regression using the t-distribution. *Computational Statistics and Data Analysis*, **71**, 116–127.

Zeller, C. B., Cabral, C. R. B. & Lachos, V. H. (2016). Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *TEST*, **25**, 375–396.