



UFAM

MODELAGEM BAYESIANA FLEXÍVEL EM REGRESSÃO COM ERROS NAS
VARIÁVEIS

Nelson Lima de Souza Filho

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: Celso Rômulo Barbosa Cabral

Manaus
Novembro de 2012

MODELAGEM BAYESIANA FLEXÍVEL EM REGRESSÃO COM ERROS NAS
VARIÁVEIS

Nelson Lima de Souza Filho

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:

Profa. Camila Borelli Zeller, D.Sc.

Prof. James Dean Oliveira dos Santos Junior, D.Sc.

Prof. Celso Rômulo Barbosa Cabral, D.Sc.

MANAUS, AM – BRASIL
NOVEMBRO DE 2012

Lima de Souza Filho, Nelson

Modelagem Bayesiana Flexível em Regressão com Erros nas Variáveis/Nelson Lima de Souza Filho. – Manaus: UFAM/ICE, 2012.

XI, 60 p.: il.; 29,7cm.

Orientador: Celso Rômulo Barbosa Cabral

Dissertação (mestrado) – UFAM/ICE, Área de Concentração: Estatística, 2012.

Referências Bibliográficas: p. 58 – 60.

1. modelos com erros nas variáveis.
 2. misturas finitas.
 3. algoritmo tipo Gibbs.
 4. t de Student assimétrica.
 5. calibração comparativa.
- I. Rômulo Barbosa Cabral, Celso. II. Universidade Federal do Amazonas, UFAM, Área de Concentração: Estatística. III. Título.

*Aos meus pais Nelson e Sorteme
com extrema gratidão.
Ao meu irmão Djalma Araújo pelo
apoio e companheirismo.
À minha namorada Débora Pessoa
por me fazer feliz.*

Agradecimentos

Agradeço,

À Deus que me trouxe do pouco e me fez professor.

Ao Professor Celso Rômulo pela ajuda, confiança, incentivo, disponibilidade e pela excelente orientação.

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

Aos meus amigos Diego, Marcia, Carla e a minha namorada Débora pela ajuda.

Aos meus amigos David e Mauricio pela amizade e ao meu irmão Djalma pelo exemplo de bondade e caráter.

Ao meu pai e minha mãe pela educação que me foi dada e pelo exemplo de pessoas humildes e honradas.

As minhas avós Severina e Ester (in Memoriam) pela confiança que sempre depositaram em mim.

Aos professores do Departamento de Estatística da UFAM pelos ensinamentos.

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.)

MODELAGEM BAYESIANA FLEXÍVEL EM REGRESSÃO COM ERROS NAS VARIÁVEIS

Nelson Lima de Souza Filho

Novembro/2012

Orientador: Celso Rômulo Barbosa Cabral

Área de Concentração : Estatística

Em modelos de regressão, o pressuposto clássico de normalidade para a distribuição dos erros aleatórios é muitas vezes violado, mascarando algumas características importantes da variabilidade dos dados. Algumas ações práticas para resolver esse problema, como transformações nos dados, revelam-se muitas vezes ineficazes.

Neste trabalho apresentamos uma proposta para lidar com esta questão no contexto do modelo de regressão multivariada linear simples, quando a variável resposta e a variável regressora são observadas com erro aditivo – o chamado modelo de regressão linear com erros nas variáveis. Em tais modelos, o pesquisador observa uma variável substituta em vez da covariável de interesse. Nós estendemos o modelo clássico normal, modelando a distribuição conjunta da covariável e dos erros aleatórios por uma mistura finita de densidades pertencentes a uma família de distribuições bem geral, acomodando ao mesmo tempo assimetria, caudas pesadas e multimodalidade, permitindo um grau de flexibilidade que não pode ser atingido pelo modelo normal.

Para a parte de estimação desenvolvemos um algoritmo do tipo Gibbs para proceder estimação Bayesiana. Alguns modelos propostos foram comparados com modelos simétricos já existentes na literatura, utilizando um critério DIC modificado, através da análise de dados simulados e reais.

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.)

BAYESIAN FLEXIBLE MODELING IN REGRESSION MODELS WITH ERRORS
IN VARIABLES

Nelson Lima de Souza Filho

November/2012

Advisor: Celso Rômulo Barbosa Cabral

Research lines: Statistics

In regression models, the classical normal assumption for the distribution of the measurement errors is often violated, masking some important features of the variability of the data. Some practical actions to overcome this problem, like transformations of the data, sometimes are not effective.

In this work we propose a methodology to overcome this problem, in the context of multivariate linear regression with measurement errors. In these models, the covariate is unobservable and the researcher observes a surrogate variable. These measurements are made with an additive error. We extend the classical normal model, by modeling jointly the covariate and the measurement errors by a finite mixture of densities which are in a general family, accommodating skewness, heavy tails and multi-modality at the same time, allowing a degree of flexibility that can not be met by the normal model.

We proceed Bayesian inference through a Gibbs-type algorithm. Some proposed models are compared with existing symmetrical models, using a modified DIC criterion, through the analysis of simulated and real data.

Sumário

Lista de Figuras	x
Lista de Tabelas	xi
1 Introdução	1
2 Distribuição Normal assimétrica multivariada e família SNI	4
2.1 Distribuição t de Student assimétrica multivariada (ST)	7
2.2 Distribuição Slash assimétrica multivariada (SSL)	8
2.3 Representação estocástica no modelo de mistura	9
3 Modelando a covariável por misturas de SNI	12
4 Estimação via MCMC	16
4.1 Algoritmo tipo Gibbs	18
5 Comparação de Modelos	22
5.1 DIC	22
6 Aplicação com dados simulados e reais	25
6.1 Dados simulados	25
6.1.1 Simulação 1	25
6.1.2 Simulação 2	27
6.2 Aplicação em dados reais	35
7 Conclusão	38
8 Apêndice A	39
8.1 Programas em R para geração da estrutura de dados	39
8.1.1 Modelo FMST-ME	39
8.1.2 Modelo NIGM-ME	40
8.2 Programas em WinBugs para estimação dos modelos	42
8.2.1 Modelo FMN-ME	42

8.2.2	Modelo FMSN-ME	43
8.2.3	Modelo FMSSL-ME	44
8.2.4	Modelo FMST-ME	46
8.2.5	Modelo FMT-ME	47
8.3	Valores iniciais e hiperparâmetros	49
8.4	Dados Chipkevitch (1996)	49
9	Apêndice B	51
9.1	Detalhes do algoritmo	51
	Referências Bibliográficas	58

Lista de Figuras

2.1	Densidade da distribuição SN com $\xi = (0, 1)'$, $\Omega = \text{diag}(2, 1.5)'$ e $\lambda = (5, 10)'$	6
2.2	Densidade da distribuição ST com $\xi = (0, 1)'$, $\Omega = \text{diag}(2, 1.5)'$, $\lambda = (5, 10)'$ e $\nu = 2$	7
2.3	Modelo FM-SN com $\xi_1 = (3, 20)'$, $\xi_2 = (7, 20)'$, $\Omega_1 = (\Omega_{11}, \Omega_{12})$, $\Omega_{11} = (20, 6)'$, $\Omega_{12} = (6, 20)$, $\Omega_2 = (\Omega_{21}, \Omega_{22})$, $\Omega_{21} = (20, -6)$, $\Omega_{22} = (-6, 20)$, $\lambda_1 = (-2, 4)$, $\lambda_2 = (2, -4)$ e $p_1 = 0.6$	11
6.1	Simulação 1 modelo FMST-ME (A) boxplot de X , (B) boxplot da variável substituta X pela classificadora w , (C) histograma de X	26
6.2	Traceplots de alguns parâmetros do modelo FMST-ME. A linha pontilhada corresponde a estimativa do parâmetro.	27
6.3	Estimador de Kernel de alguns parâmetros do modelo FMST-ME. Para mais detalhes veja a função <code>density</code> no pacote <code>stats</code> do programa R ou a forma teórica em Scott [26].	28
6.4	Preditiva dos modelos com erros nas variáveis com dados FMST-ME. A linha contínua de cor preta corresponde à densidade real de x	29
6.5	Variável X do Modelo FMNIG-ME, (A) boxplot de X , (B) boxplot de X condicionada a variável classificadora w , (C) histograma de X	32
6.6	Traceplots de alguns parâmetros do modelo FMST-ME, para dados gerados do modelo FMNIG-ME. A linha pontilhada corresponde a estimativa do parâmetro.	33
6.7	Traceplots de alguns parâmetros do modelo FMSSL-ME, para dados gerados do modelo FMNIG-ME. A linha pontilhada corresponde a estimativa do parâmetro.	34
6.8	Preditiva da variável x nos modelos FMST-ME, FMSSL-ME, FMN-ME, FMT-ME e FMSN-ME para dados gerados do modelo FMNIG-ME.	35
6.9	Preditiva da variável x nos modelos FMST-ME, FMSSL-ME, FMN-ME, FMT-ME e FMSN-ME para o conjunto de dados Chipkevitch (1996).	36

Lista de Tabelas

6.1	Estimativas dos parâmetros para vários modelos com erros nas variáveis para os dados simulados FMST-ME. A sigla dp denota o desvio padrão da amostra MCMC e RV o real valor dos parâmetros	28
6.2	Intervalos de credibilidade para estimação dos modelos FMST-ME e FMSSL-ME para os dados simulados FMST-ME, 95% de credibilidade. .	29
6.3	Critério de seleção para vários modelos com erros nas variáveis para os dados simulados FMST-ME.	29
6.4	Estimativas dos parâmetros para vários modelos com erros nas variáveis para os dados simulados FMNIG-ME. A sigla dp denota o desvio padrão.	33
6.5	Intervalos de credibilidade para estimação dos modelos FMST-ME e FMSSL-ME para os dados simulados FMNIG-ME, 95% de credibilidade.	34
6.6	Critério de seleção para vários modelos com erros nas variáveis para os dados simulados FMNIG-ME.	34
6.7	Estimativas dos parâmetros para vários modelos com erros nas variáveis para o conjunto de dados Chipkevitch (1996). A sigla dp denota o desvio padrão da amostra MCMC.	36
6.8	Intervalos de credibilidade para estimação dos modelos FMST-ME e FMSSL-ME para os dados de Chipkevitch (1996), 95% de credibilidade. .	37
6.9	Critério de seleção para vários modelos com erros nas variáveis para os dados de Chipkevitch (1996).	37

Capítulo 1

Introdução

Suponha o clássico modelo de regressão multivariada simples com vetor de respostas dado por

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \beta x_i + \mathbf{e}_i, \quad (1.1)$$

onde $\mathbf{Y}_i : r \times 1$ é um vetor aleatório de medições feitas no indivíduo i , $\mathbf{e}_i : r \times 1$ é um vetor de erros de mensuração e x_i é o valor da variável regressora para o indivíduo i , para $i = 1, \dots, n$. Além disso, temos os vetores r -dimensionais de parâmetros de regressão desconhecidos $\boldsymbol{\alpha}$ e β .

Nesse texto estudaremos modelos de regressão onde se é incapaz de observar x_i diretamente. Em vez disso, observarmos a seguinte soma

$$X_i = x_i + \zeta_i. \quad (1.2)$$

Nesse modelo a variável observada X_i é mensurada com um erro aleatório ζ_i associado a observação i . Este modelo é chamado de modelo de regressão com erros nas variáveis. Algumas referências clássicas são os livros de Fuller [15] e Cheng & Van Ness [12]. Para uma mais recente podemos ver Buonaccorsi [5].

Uma utilidade particular para esse modelo é em modelos de calibração comparativa. Nesse caso o indivíduo i é submetido a uma medição (univariada no nosso caso) feita por um instrumento padrão (esta medida é de referência). Essa medição é a variável X_i e outros r instrumentos são utilizados para gerar o vetor resposta $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir})$, que é um vetor com outras r medições feitas por r outros instrumentos, isto é, Y_{ij} é a i -ésima observação feita pelo j -ésimo instrumento, $i = 1, \dots, n$ e $j = 1, \dots, G$.

A variável observada X_i é conhecida como variável substituta, enquanto a variável x_i não observada é conhecida como variável latente. Modelos com x_i fixas são chamados de modelos funcionais, enquanto que os modelos com x_i aleatórias são chamados de modelos

estruturais. No nosso caso trabalharemos apenas com os estruturais.

Para melhor exemplificar a situação vamos tomar um exemplo contido em Fuller (1987). Considere a relação entre o rendimento de milho, que será a nossa variável resposta, e a quantidade de nitrogênio no solo, que será a variável latente x_i . Assuma que (1.1) é uma aproximação adequada para a relação entre essas variáveis. Para estimar o nitrogênio disponível no solo, é necessário recolher uma amostra do solo para a parcela experimental e realizar uma análise laboratorial na amostra selecionada. Como o resultado da amostragem é obtido através de uma análise de laboratório, não observamos x_i , mas observamos uma estimativa de x_i . Portanto, nós representamos o nitrogênio observado por X_i , onde x_i satisfaz (1.2) e ζ_i é o erro de medição introduzido por intermédio de amostragem e de uma análise laboratorial.

Para o modelo com erros nas variáveis geralmente é assumido que

$$\boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N_{1+r}(\mathbf{0}, \boldsymbol{\Omega}), \quad (1.3)$$

onde $\boldsymbol{\epsilon}_i = (\zeta_i, \mathbf{e}_i)'$, $\boldsymbol{\Omega} = \text{block diag}\{\boldsymbol{\omega}_0^2, \boldsymbol{\Omega}_e\}$ e $\mathbf{0}$ é o vetor nulo. Logo, temos

$$\zeta_i \stackrel{\text{iid}}{\sim} N(0, \boldsymbol{\omega}_0^2) \quad \text{e} \quad \mathbf{e}_i \stackrel{\text{iid}}{\sim} N_r(\mathbf{0}, \boldsymbol{\Omega}_e), \quad i = 1, \dots, n, \quad (1.4)$$

onde ζ_i é independente de \mathbf{e}_i . O termo $\stackrel{\text{iid}}{\sim}$ denota que os vetores aleatórios são independentes e identicamente distribuídos.

Na verdade temos que (1.3) é consequência da suposição de que

$$\mathbf{R}_i = \begin{bmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \stackrel{\text{iid}}{\sim} N_{1+p} [(\boldsymbol{\mu}, \mathbf{0}_p)'], \text{block diag}\{\boldsymbol{\sigma}^2, \boldsymbol{\Omega}\}, \quad (1.5)$$

onde $p = 1 + r$. Assim temos

$$x_i \stackrel{\text{iid}}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2), \quad \boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Omega}), \quad i = 1, \dots, n, \quad (1.6)$$

onde x_i é independente de $\boldsymbol{\epsilon}_i$, para $i = 1, \dots, n$.

A distribuição assumida para a variável x_i é um tema que merece uma atenção especial, pois esta variável costuma ser assumida como normal, como vimos em (1.6). Porém essa afirmação limita a modelagem dessa variável quando a mesma apresenta características não tradicionais como assimetria, outliers e heterogeneidade (o que implica em uma distribuição multimodal para o vetor de respostas). Para esses casos seria mais adequado o uso de um modelo mais robusto, no sentido de que a inferência sobre os parâmetros não seja tão sensível a ele. Alguns trabalhos neste sentido são os de Muller & Roeder [22], onde os autores supõem que a distribuição de $(x_i, X_i)'$ segue um processo de Dirichlet e Lachos *et al.* [20, 21], que estenderam o modelo (1.5) considerando que o vetor \mathbf{R}_i tem distribuição em uma família de densidades que incorpora características não existentes no

modelo normal, tornando assim o modelo bem mais flexível.

O modelo proposto por Lachos *et al.* [20] é mais flexível que (1.5), pois nos permite a incorporação de assimetria e caudas pesadas, com grande aplicabilidade onde o pressuposto de normalidade falha. Porém, esse modelo ainda não é satisfatório quando há heterogeneidade na distribuição de x_i (não observada). Geralmente o problema de heterogeneidade é resolvido com modelagem através de misturas de densidades normais como em Carroll *et al.* [8] e Carroll *et al.* [9]. Essa abordagem pode ser útil até mesmo para respostas discretas, como podemos ver em Richardson *et al.* [24], onde o modelo de regressão logística com erros nas variáveis é considerado. Também temos Roy & Banerjee [25], que estudaram o modelo de regressão generalizada com erros nas variáveis onde a covariável latente é modelada por uma mistura finita de distribuições normais e o erro aleatório \mathbf{e}_i tem uma distribuição que pertence à família Normal independente - NI que iremos definir logo abaixo.

Definição 1. *Um vetor aleatório \mathbf{Y} pertence a família NI (Normal Independente) se*

$$\mathbf{Y} = \boldsymbol{\xi} + U^{-1/2}\mathbf{X},$$

onde $\boldsymbol{\xi}$ é um vetor q -dimensional de constantes, \mathbf{X} é um vetor q -dimensional com distribuição normal com vetor de médias igual ao vetor nulo e matriz de covariâncias $\boldsymbol{\Omega}$ e U é uma variável aleatória com densidade $h(u|v)$, sendo que \mathbf{X} e U são independentes e $P(U > 0) = 1$.

É notável a aplicabilidade e a flexibilidade dos modelos citados anteriormente, porém ainda existe a ausência de flexibilidade do ponto de vista da multimodalidade e heterogeneidade (como já havíamos falado) para o modelo citado por Lachos *et al.* [20, 21] ou no caso do modelo proposto por Roy & Banerjee [25]. Esse modelo nos mostrou que mesmo usando uma modelagem com misturas de normais, podemos ter uma superestimação do número dos componentes ou um aumento no vício do parâmetro de escala de cada sub-população, a fim de capturar a natureza assimétrica que pertence naturalmente aos dados.

Nossa intenção é modelar a distribuição da variável latente por uma classe mais flexível de distribuições. A extrema flexibilidade dessa classe pode tornar possível a acomodação simultânea de assimetria, outliers e multimodalidade, além de construirmos um modelo mais parcimonioso.

Capítulo 2

Distribuição Normal assimétrica multivariada e família SNI

Estender a estrutura de distribuições é um problema bastante abordado na literatura atual, principalmente quando falamos de extensões para tornar distribuições antes univariadas em multivariadas e a introdução de parâmetros reguladores de características antes inexistentes nas mesmas. A distribuição Normal Assimétrica (SN, Skew-normal, em inglês) multivariada, que podemos ver em detalhes em Azzalini [2] e Azzalini & Dalla Valle [3] é um claro exemplo de extensão da distribuição normal incorporando um vetor de parâmetros regulador de forma, controlando a assimetria da distribuição. Para definirmos formalmente a distribuição normal assimétrica, diremos que um vetor aleatório \mathbf{Z} com dimensão q tem distribuição normal assimétrica padronizada, denotado por $\mathbf{Z} \sim \text{SN}(\boldsymbol{\lambda})$, com parâmetro de forma $\boldsymbol{\lambda}$, se sua densidade é dada por

$$\text{SN}_q(\mathbf{z}|\boldsymbol{\lambda}) = 2\text{N}_q(\mathbf{z}|\mathbf{0}_q, \mathbf{I}_q)\Phi(\boldsymbol{\lambda}'\mathbf{z}); \quad \mathbf{z} \in R^q,$$

onde \mathbf{I}_q é a matriz identidade de dimensão q (vamos abstrair o índice q quando não houver nenhuma chance de confusão) e $\text{N}_q(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ representa a densidade normal q variada. Também temos que $\Phi(\cdot)$ representa a função de distribuição de uma normal padrão univariada. Com a transformação $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\Omega}^{1/2}\mathbf{Z}$ introduzimos os parâmetros de locação $\boldsymbol{\xi}$, que é um vetor $q \times 1$ de constantes e o de escala $\boldsymbol{\Omega}$, que é uma matriz $q \times q$ positiva definida, onde $\boldsymbol{\Omega}^{1/2}$ é a matriz raiz quadrada de $\boldsymbol{\Omega}$, definida tal que $\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2} = \boldsymbol{\Omega}$. Podemos ver mais detalhes em Zhang [30]. Com essas informações já podemos definir a distribuição normal assimétrica.

Definição 2. Dizemos que um vetor aleatório \mathbf{X} tem distribuição normal assimétrica com vetor de locação $\boldsymbol{\xi}$: $q \times 1$, matriz de dispersão $\boldsymbol{\Omega}$: $q \times q$ e vetor de forma (assimetria) $\boldsymbol{\lambda}$: $q \times 1$, se a sua densidade é dada por

$$\text{SN}_q(\mathbf{x}|\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}) = 2\text{N}_q(\mathbf{x}|\boldsymbol{\xi}, \boldsymbol{\Omega})\Phi(\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1/2}(\mathbf{x} - \boldsymbol{\xi})), \quad (2.1)$$

onde $\Omega^{-1/2}$ é a inversa da matriz $\Omega^{1/2}$.

A notação usada é $\mathbf{X} \sim \text{SN}_q(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda})$. É importante notar que se $\boldsymbol{\lambda}$ é o vetor nulo então $\mathbf{X} \sim N_q(\boldsymbol{\xi}, \Omega)$, ou seja, a distribuição normal q -variada usual. Quanto ao vetor de médias e a matriz de covariâncias de $\mathbf{X} \sim \text{SN}_q(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda})$, temos

$$E(\mathbf{X}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}} \Omega^{1/2} \boldsymbol{\delta}; \quad (2.2)$$

$$\text{Var}(\mathbf{X}) = \Omega - \frac{2}{\pi} \Omega^{1/2} \boldsymbol{\delta} \boldsymbol{\delta}' \Omega^{1/2}, \quad \text{onde } \boldsymbol{\delta} = \boldsymbol{\lambda} / (1 + \boldsymbol{\lambda}' \boldsymbol{\lambda})^{1/2}. \quad (2.3)$$

A seguir vamos definir um conceito extremamente importante para o nosso trabalho, que é o conceito de misturas de densidades.

Definição 3. *Seja \mathbf{Y} um vetor aleatório e S uma variável aleatória com densidade $h(\cdot)$, tal que $P(S \in \mathcal{S}) = 1$, ambos definidos em um mesmo espaço de probabilidade. Para $s \in \mathcal{S}$, seja $g(\cdot|s)$ a densidade da condicional $\mathbf{Y}|S = s$. Então a densidade*

$$f(\mathbf{y}) = \int_{\mathcal{S}} g(\mathbf{y}|s) h(s) ds, \quad (2.4)$$

é chamada uma *Mistura Escalonada de densidades da família* $\{g(\cdot|s); s \in \mathcal{S}\}$. S é chamado de *Fator de Escala* e $h(\cdot)$ é chamada de *Densidade da Mistura*.

Suponha agora que S é uma variável aleatória discreta, de tal modo que $\mathcal{S} = \{1, \dots, G\}$ e

$$P(S = i) = p_i, \quad \text{onde } p_i \geq 0, \quad i = 1, \dots, G, \quad \sum_{i=1}^G p_i = 1. \quad (2.5)$$

Definindo $g(\cdot|S = i) = g_i(\cdot)$ temos um caso particular e muito importante chamado de *Mistura Finita de Densidades*. Nesse caso, (2.4) é da forma

$$f(\mathbf{y}) = \sum_{i=1}^G p_i g_i(\mathbf{y}). \quad (2.6)$$

Para esse tipo de mistura, o fator de escala S pode ser interpretado como uma variável classificadora: dado que observamos $S = i$, então \mathbf{y} é oriundo da população com densidade $g_i(\cdot)$. Além disso, a probabilidade p_i e a densidade g_i são chamadas de *i -ésimo Peso da Mistura* e a *i -ésima Componente da Mistura*, respectivamente, $i = 1, \dots, G$. As misturas finitas utilizadas neste trabalho têm como componentes densidades contidas na família SNI.

Definição 4. *Um vetor aleatório \mathbf{Y} pertence a família SNI (Skew-Normal Independent)*

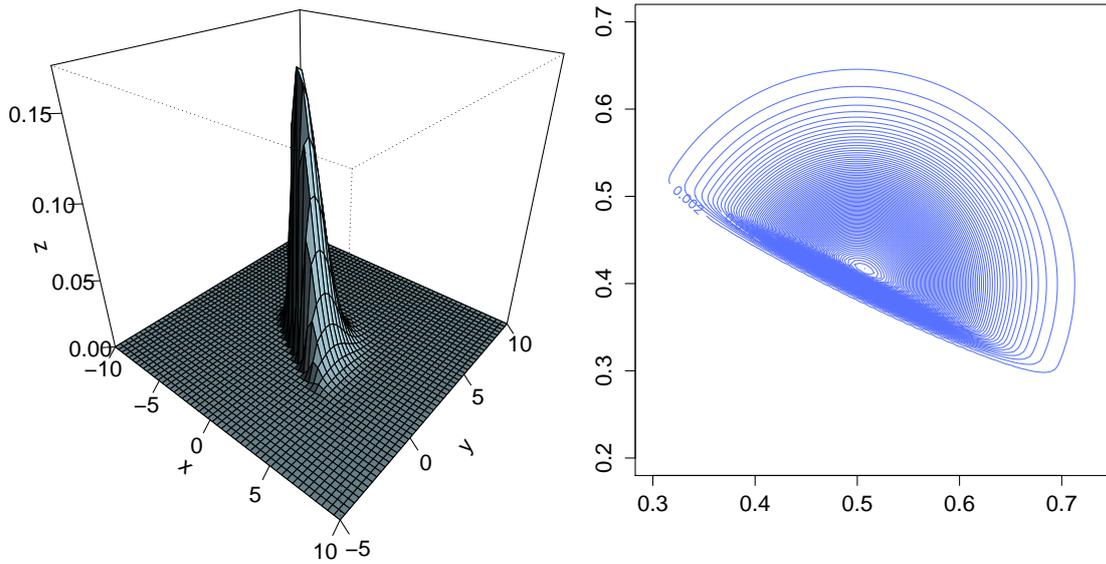


Figura 2.1: Densidade da distribuição SN com $\xi = (0, 1)'$, $\Omega = \text{diag}(2, 1.5)'$ e $\lambda = (5, 10)'$.

se

$$\mathbf{Y} = \xi + U^{-1/2}\mathbf{X},$$

onde $\xi: q \times 1$ é um vetor de constantes, $\mathbf{X} \sim \text{SN}_q(\mathbf{0}, \Omega, \lambda)$ e U é uma variável aleatória com densidade $h(u|v)$ sendo que \mathbf{X} e U são independentes e $P(U > 0) = 1$.

Note que se $U = 1$ então o vetor \mathbf{Y} terá distribuição normal assimétrica com parâmetros ξ, Ω e λ . Na Figura 2.1, temos o gráfico de uma densidade normal assimétrica bivariada.

Pela Definição 4 temos

$$\mathbf{Y}|U = u \sim \text{SN}_q(\xi, u^{-1}\Omega, \lambda), \quad U \sim h(u|v). \quad (2.7)$$

Podemos ver, pela Definição 3, que as distribuições que são fruto da Definição 4 para diferentes $h(u|v)$ são misturas de escala de distribuições normais assimétricas, em que U é o fator de escala e $h(u|v)$ é a densidade da mistura. Assim temos que a densidade marginal de \mathbf{Y} é

$$\text{SNI}(\mathbf{y}|\xi, \Omega, \lambda, v) = 2 \int_0^\infty \text{N}_q(\mathbf{y}|\xi, u^{-1}\Omega) \Phi(u^{1/2} \lambda' \Omega^{-1/2} (\mathbf{y} - \xi)) h(u|v) du. \quad (2.8)$$

Como comentamos, dependendo da distribuição de U , podemos ter diferentes distribuições, como a normal assimétrica quando $U = 1$, ou a normal assimétrica contaminada quando U tem distribuição binária com $P(U = \tau) = \rho = 1 - P(U = 1)$, onde $0 < \tau, \rho < 1$.

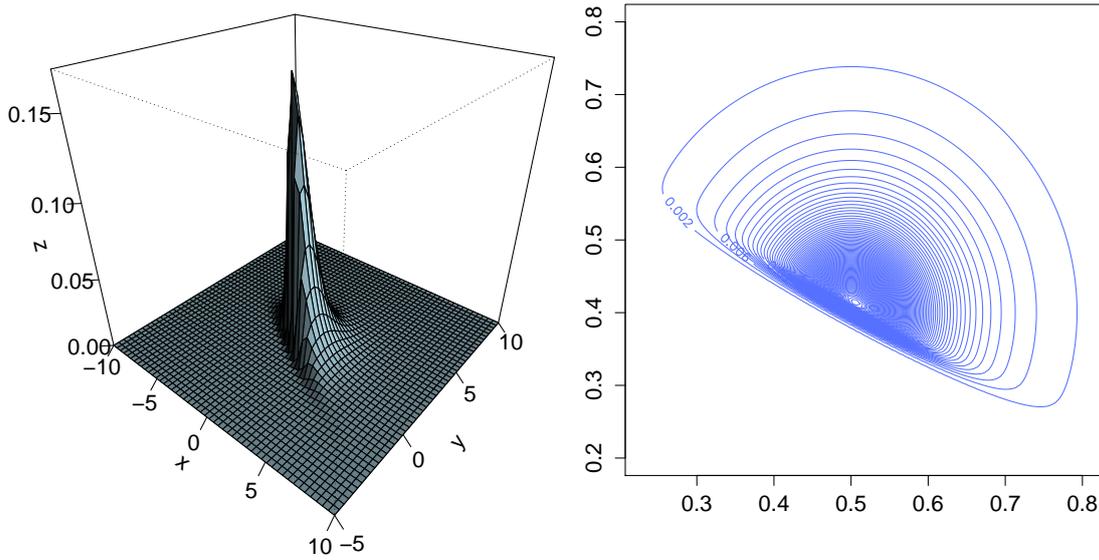


Figura 2.2: Densidade da distribuição ST com $\xi = (0, 1)'$, $\Omega = \text{diag}(2, 1.5)'$, $\lambda = (5, 10)'$ e $\nu = 2$.

Essas distribuições fazem parte da família SNI, bem explorada em Cabral *et al.* [7]. Porém, não são as únicas pertencentes à essa família. Neste trabalho trataremos a modelagem através de duas densidades que veremos a seguir.

2.1 Distribuição t de Student assimétrica multivariada (ST)

Essa distribuição é muito usada em modelos com muitos outliers, pois além de incorporar a assimetria ainda combina uma estrutura de caudas pesadas, o que a torna muito mais flexível que a normal ou a t de Student.

A distribuição t de Student Assimétrica (ST, do inglês skew-t) multivariada com ν graus de liberdade, com notação $ST_q(\xi, \Omega, \lambda, \nu)$, pode ser encontrada a partir da marginal (2.8), tendo U com distribuição $\text{Gamma}(\nu/2, \nu/2)$. A densidade do vetor \mathbf{Y} é dada por

$$f(\mathbf{y}) = 2t_q(\mathbf{y}|\xi, \Omega, \nu) T\left(\left(\frac{q+\nu}{d+\nu}\right)^{1/2} \lambda' \Omega^{-1/2} (\mathbf{y}-\xi) \mid \nu+p\right), \quad \mathbf{y} \in \mathbb{R}^q, \quad (2.9)$$

onde $d = (\mathbf{y} - \xi)' \Omega^{-1} (\mathbf{y} - \xi)$ é a distância de Mahalanobis ao quadrado, $t_q(\cdot|\xi, \Omega, \nu)$ denota a densidade t de Student multivariada e $T(\cdot|\nu)$ a função de distribuição t de Student univariada com parâmetros 0, 1 e ν . Note que quando λ é o vetor nulo então temos a distribuição t usual (simétrica).

Em relação aos momentos de $\mathbf{Y} \sim \text{ST}_q(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu)$ temos

$$\begin{aligned} \mathbb{E}[\mathbf{Y}] &= \boldsymbol{\xi} + (\nu/\pi)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta}, \quad \nu > 1; \\ \text{Var}[\mathbf{Y}] &= \frac{\nu}{\nu-2} \boldsymbol{\Omega} - (\nu/\pi) \left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \right)^2 \boldsymbol{\Omega} \boldsymbol{\delta} \boldsymbol{\delta}', \quad \nu > 2, \end{aligned}$$

Para esclarecermos a parametrização usada, ter $U \sim \text{Gamma}(r, s)$ significa que U tem a seguinte densidade

$$f(u) = \frac{s^r u^{r-1} e^{-su}}{\Gamma(r)}, \quad u > 0. \quad (2.10)$$

Para compreendermos o comportamento da distribuição t de Student assimétrica é interessante observarmos a Figura 2.2 onde temos um gráfico de uma ST bivariada. Nesta figura fica clara as caudas mais pesadas da distribuição t de Student Assimétrica comparada com a distribuição Normal Assimétrica que vimos na Figura 2.1.

2.2 Distribuição Slash assimétrica multivariada (SSL)

Outra distribuição flexível do ponto de vista de caudas pesadas e assimetria muito usada na literatura é a distribuição Slash Assimétrica (SSL, do inglês Skew-Slash) com notação $\text{SSL}_q(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu)$. Ela pode ser encontrada a partir da marginal (2.8) quando U tem distribuição $\text{Beta}(\nu, 1)$, tendo densidade

$$f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1} \mathbf{N}_q(\mathbf{y}|\boldsymbol{\xi}, u^{-1}\boldsymbol{\Omega}) \Phi(u^{1/2} \boldsymbol{\lambda}' \boldsymbol{\Sigma}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})) du, \quad \mathbf{y} \in \mathbb{R}. \quad (2.11)$$

Quanto aos momentos da SSL, temos

$$\begin{aligned} \mathbb{E}[\mathbf{Y}] &= \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}} \frac{2\nu}{2\nu-1} \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta}, \quad \nu > 1/2, \text{ e} \\ \text{Var}[\mathbf{Y}] &= \frac{\nu}{\nu-1} \boldsymbol{\Omega} - \frac{2}{\pi} \left(\frac{2\nu}{2\nu-1} \right)^2 \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta} \boldsymbol{\delta}' \boldsymbol{\Omega}^{1/2}, \quad \nu > 1. \end{aligned}$$

Para esclarecermos a parametrização usada, ter $U \sim \text{Beta}(a, b)$ significa que U tem a seguinte densidade

$$f(u) = u^{a-1} (1-u)^{b-1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \quad 0 < u < 1. \quad (2.12)$$

2.3 Representação estocástica no modelo de mistura

Existe outra forma de definir as distribuições que pertencem a família SNI, ela é dada no próximo teorema. Porém antes de apresentá-lo, vamos definir a distribuição *normal truncada* (NT). Dizemos que T tem uma distribuição NT no intervalo (a, b) com parâmetros ξ e η^2 , denotando por $T \sim \text{NT}(\xi, \eta^2, (a, b))$, quando T tiver a mesma distribuição de $W|(a < W < b)$ para alguma variável aleatória $W \sim \text{N}(\xi, \eta^2)$. Se $a = 0$, $b = +\infty$, $\xi = 0$ e $\eta^2 = 1$, então a chamamos de *half-normal* ($\text{HN}(0,1)$).

Teorema 1. Se $\mathbf{X} \sim \text{SN}_q(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ então existem variáveis T e \mathbf{V} tais que

$$\mathbf{X} = \boldsymbol{\xi} + \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta} T + \boldsymbol{\Omega}^{1/2} (\mathbf{I} - \boldsymbol{\delta} \boldsymbol{\delta}')^{1/2} \mathbf{V},$$

onde $T \sim \text{HN}(0,1)$, $\mathbf{V} \sim \text{N}_q(\mathbf{0}, \mathbf{I})$ e $\boldsymbol{\delta} = \boldsymbol{\lambda} / (1 + \boldsymbol{\lambda}' \boldsymbol{\lambda})^{1/2}$, para T e \mathbf{V} independentes

Este teorema pode ser encontrado em Arellano-Valle & Genton [1] ou sua forma univariada em Henze [19]. Uma forma mais atrativa de se utilizar o Teorema 1 e que ajuda muito em nossos propósitos é usando a seguinte reparametrização

$$\boldsymbol{\Delta} = \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta}, \quad \boldsymbol{\Gamma} = \boldsymbol{\Omega}^{1/2} (\mathbf{I} - \boldsymbol{\delta} \boldsymbol{\delta}') \boldsymbol{\Omega}^{1/2} = \boldsymbol{\Omega} - \boldsymbol{\Delta} \boldsymbol{\Delta}', \quad (2.13)$$

ou seja, podemos reescrever o teorema da seguinte forma

$$\mathbf{X} = \boldsymbol{\xi} + \boldsymbol{\Delta} T + \boldsymbol{\Gamma}^{1/2} \mathbf{V}, \quad \text{onde } T \sim \text{HN}(0,1) \text{ e } \mathbf{V} \sim \text{N}_q(\mathbf{0}, \mathbf{I}). \quad (2.14)$$

Note que $(\boldsymbol{\lambda}, \boldsymbol{\Omega}) \rightarrow (\boldsymbol{\Delta}, \boldsymbol{\Gamma})$ é uma relação um a um. Então podemos recuperar $\boldsymbol{\lambda}$ e $\boldsymbol{\Omega}$ com as expressões

$$\boldsymbol{\lambda} = \frac{(\boldsymbol{\Gamma} + \boldsymbol{\Delta} \boldsymbol{\Delta}')^{-1/2} \boldsymbol{\Delta}}{[1 - \boldsymbol{\Delta}' (\boldsymbol{\Gamma} + \boldsymbol{\Delta} \boldsymbol{\Delta}')^{-1} \boldsymbol{\Delta}]^{1/2}}, \quad \boldsymbol{\Omega} = \boldsymbol{\Gamma} + \boldsymbol{\Delta} \boldsymbol{\Delta}'. \quad (2.15)$$

Entretanto agora os parâmetros a serem estimados são $(\boldsymbol{\xi}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$, o que nos dá a seguinte notação $\mathbf{X} \sim \text{SN}_q(\boldsymbol{\xi}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$. Esta reparametrização tem função crucial no processo inferencial. Através dela podemos enunciar os seguintes teoremas.

Teorema 2. $\mathbf{Y} \sim \text{SN}_q(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ admite a seguinte representação estocástica

$$\mathbf{Y}|T = t \sim \text{N}_q(\boldsymbol{\xi} + \boldsymbol{\Delta} t, \boldsymbol{\Gamma}); \quad T \sim \text{HN}(0,1), \quad (2.16)$$

onde $\boldsymbol{\Delta}$ e $\boldsymbol{\Gamma}$ são dados em (2.13).

Demonstração: Aplicação direta do Teorema 1 reparametrizado dado em (2.14)

Teorema 3. $\mathbf{Y} \sim \text{SNI}_q(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu)$ admite a seguinte representação estocástica

$$\mathbf{Y}|T = t, U = u \sim \text{N}_q(\boldsymbol{\xi} + \boldsymbol{\Delta}t, u^{-1}\boldsymbol{\Gamma}); \quad T \sim \text{HN}(0, u^{-1}); \quad U \sim h(u|\nu), \quad (2.17)$$

onde $\boldsymbol{\Delta}$ e $\boldsymbol{\Gamma}$ são dados em (2.13).

Antes de demonstrarmos o Teorema 3 vamos especificar que a notação $\stackrel{d}{=}$ significa que as variáveis tem a mesma distribuição.

Demonstração. Usando (2.7) e (2.14) temos

$$\mathbf{Y}|U = u \stackrel{d}{=} \boldsymbol{\xi} + \boldsymbol{\Delta}T^* + u^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{V}, \quad \text{onde} \quad T^* = u^{-1/2}T \sim \text{HN}(0, u^{-1}),$$

ou seja,

$$\mathbf{Y}|U = u, T^* = t \stackrel{d}{=} \boldsymbol{\xi} + \boldsymbol{\Delta}t + u^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{V}.$$

Como $\mathbf{V} \sim \text{N}_q(\mathbf{0}, I)$ então

$$\mathbf{Y}|U = u, T^* = t \sim \text{N}_q(\boldsymbol{\xi} + \boldsymbol{\Delta}t, u^{-1}\boldsymbol{\Gamma}); \quad T^*|U = u \sim \text{HN}(0, u^{-1}); \quad U \sim h(u|\nu).$$

□

A importância desses teoremas está totalmente ligada à facilidade que teremos para construir um algoritmo de Gibbs usando as representações estocásticas.

O resultado dos Teoremas 2 e 3 nos permite obter a distribuição de uma transformação linear $\mathbf{A}\mathbf{Y}$, onde \mathbf{A} é uma matriz $p \times q$. Se usarmos a parametrização alternativa (2.13), para $\mathbf{Y} \sim \text{SNI}_q(\boldsymbol{\xi}, \boldsymbol{\Gamma}, \boldsymbol{\Delta}, \nu)$, então $\mathbf{A}\mathbf{Y} \sim \text{SNI}_q(\mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}', \mathbf{A}\boldsymbol{\Delta}, \nu)$. Em seguida, podemos utilizar as relações (2.15) para recuperar a parametrização original.

Definição 5. Uma mistura finita de distribuições cujas componentes são SNI é dada por (2.6) onde $g_i(\mathbf{y}) = \text{SNI}_q(\mathbf{y}|\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i, \boldsymbol{\lambda}_i, \nu_i)$.

Chamamos este modelo de FM-SNI. Para o conceito de misturas vamos definir uma variável classificadora $w \stackrel{d}{=} S$ equivalente a (2.5). A variável w não é observável, mas se fosse funcionaria como uma variável que informa a qual das G populações da mistura o vetor de observações \mathbf{y} pertence. Este fato em conjunto com os Teoremas 2 e 3 nos permite enunciar o seguinte resultado, uma representação em dados aumentados (representação estocástica) para o modelo FM-SNI.

Teorema 4. Suponha que a distribuição do vetor aleatório \mathbf{Y} : $q \times 1$ é uma mistura de densidades onde cada subgrupo i tem distribuição $\text{SNI}_q(\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i, \boldsymbol{\lambda}_i, \nu_i)$, $i = 1, \dots, G$. Então

\mathbf{Y} admite a seguinte representação

$$\begin{aligned}
 \mathbf{Y}|w=i, T=t, U=u &\sim N_q(\boldsymbol{\xi}_i + \boldsymbol{\Delta}_i t, u^{-1}\boldsymbol{\Gamma}_i); \\
 T|U=u &\sim \text{HN}(0, u^{-1}); \\
 U &\sim h(u|v); \\
 P(w=i) &= p_i, \quad i=1, \dots, G,
 \end{aligned} \tag{2.18}$$

onde $\boldsymbol{\Delta}_i = \boldsymbol{\Omega}_i^{1/2} \boldsymbol{\delta}_i$, $\boldsymbol{\delta}_i = \boldsymbol{\lambda}_i / (1 + \boldsymbol{\lambda}_i' \boldsymbol{\lambda}_i)^{1/2}$, $\boldsymbol{\Gamma}_i = \boldsymbol{\Omega}_i - \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i'$, $p_i \geq 0$, $i=1, \dots, G$, com $\sum_{i=1}^G p_i = 1$.

Na Figura 2.3 vemos um gráfico de densidade para um modelo FM-SNI com duas densidade bivariadas ST na mistura.

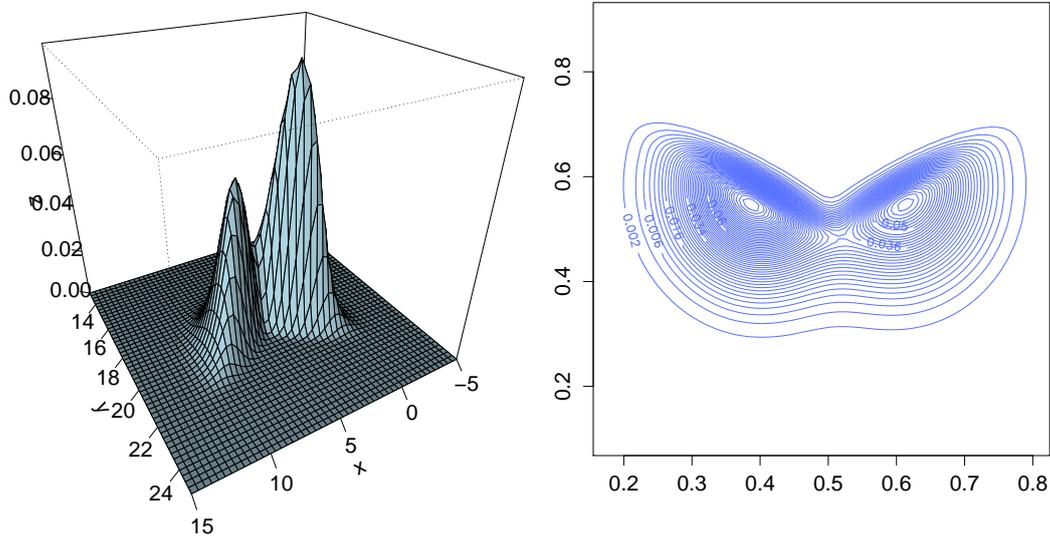


Figura 2.3: Modelo FM-SN com $\boldsymbol{\xi}_1 = (3, 20)'$, $\boldsymbol{\xi}_2 = (7, 20)'$, $\boldsymbol{\Omega}_1 = (\boldsymbol{\Omega}_{11}, \boldsymbol{\Omega}_{12})$, $\boldsymbol{\Omega}_{11} = (20, 6)'$, $\boldsymbol{\Omega}_{12} = (6, 20)$, $\boldsymbol{\Omega}_2 = (\boldsymbol{\Omega}_{21}, \boldsymbol{\Omega}_{22})$, $\boldsymbol{\Omega}_{21} = (20, -6)$, $\boldsymbol{\Omega}_{22} = (-6, 20)$, $\boldsymbol{\lambda}_1 = (-2, 4)$, $\boldsymbol{\lambda}_2 = (2, -4)$ e $p_1 = 0.6$

Capítulo 3

Modelando a covariável por misturas de SNI

Lachos *et al.* [20] estenderam o modelo com erros nas variáveis para uma família de distribuições bem mais ampla que a normal, assumindo que a variável latente x_i pode ser modelada por uma SNI, tendo assim

$$\mathbf{R}_i = \begin{bmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \stackrel{\text{iid}}{\sim} \text{SNI}_{1+p} \left[\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0}_p \end{pmatrix}, \text{block diag}\{\sigma^2, \boldsymbol{\Omega}\}, \begin{pmatrix} \lambda \\ \mathbf{0}_p \end{pmatrix} \right], \quad (3.1)$$

onde temos marginalmente

$$x_i \stackrel{\text{iid}}{\sim} \text{SNI}(\boldsymbol{\mu}, \sigma^2, \lambda, \boldsymbol{\nu}), \quad \boldsymbol{\epsilon}_i \stackrel{\text{ind}}{\sim} \text{NI}_p(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\nu}), \quad i = 1, \dots, n. \quad (3.2)$$

Para provar (3.2) basta verificarmos que

$$x_i = \begin{bmatrix} 1 & \mathbf{0}'_p \end{bmatrix} \mathbf{R}_i, \quad \boldsymbol{\epsilon}_i = \begin{bmatrix} 0 & \mathbf{I}_p \end{bmatrix} \mathbf{R}_i,$$

e usarmos o Teorema 3.

Pela definição 4 e assim como foi feito em (2.7) temos

$$\begin{aligned} \mathbf{R}_i | U_i = u_i &\sim \text{SN}_{1+p} \left[\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0}_p \end{pmatrix}, u_i^{-1} \text{block diag}\{\sigma^2, \boldsymbol{\Omega}\}, \begin{pmatrix} \lambda \\ \mathbf{0}_p \end{pmatrix} \right], \\ U_i &\sim h(u | \boldsymbol{\nu}). \end{aligned} \quad (3.3)$$

Partindo da necessidade de usarmos modelos flexíveis para modelar a covariável latente de um modelo com erros nas variáveis chegamos ao nosso objetivo que é modelar a distribuição da covariável latente x_i por uma distribuição FM-SNI. Com extrema flexibilidade, esse modelo pode tornar possível a acomodação simultânea de assimetria, outliers

e multimodalidade, além de possivelmente termos um modelo com menos componentes comparado com o necessário para obter o mesmo grau de explicação usando a distribuição normal, t de Student ou outra densidade simétrica. Assim, como em Roy & Banerjee [25] a nossa proposta mantém a suposição de que ϵ_i pertence a família NI.

O modelo proposto tem a seguinte representação

$$\mathbf{R}_i | w_i = j \stackrel{\text{iid}}{\sim} \text{SNI}_{1+p} \left[\begin{pmatrix} \mu_j \\ \mathbf{0}_p \end{pmatrix}, \text{block diag}\{\sigma_j^2, \mathbf{\Omega}\}, \begin{pmatrix} \lambda_j \\ \mathbf{0}_p \end{pmatrix}, \mathbf{v} \right], \quad i = 1, \dots, n, \quad (3.4)$$

onde w_i é a variável latente classificadora dada em (2.18), de tal modo que $w_i = j$ tem probabilidade p_j , $j = 1, \dots, G$. Marginalmente temos

$$x_i | w_i = j \sim \text{SNI}(\mu_j, \sigma_j^2, \lambda_j, \mathbf{v}), \quad \epsilon_i \sim \text{NI}_p(\mathbf{0}, \mathbf{\Omega}, \mathbf{v}), \quad i = 1, \dots, n.$$

como a densidade conjunta entre x_i e w_i é dada por

$$f(x_i, w_i = j) = f(x_i | w_i = j)P(w_i = j),$$

temos que

$$f(x_i) = \sum_{j=1}^G p_j f(x_i | w_i = j).$$

Logo temos

$$x_i \sim \sum_{j=1}^G p_j \text{SNI}(\cdot | \mu_j, \sigma_j^2, \lambda_j, \mathbf{v}), \quad \epsilon_i \sim \text{NI}_p(\mathbf{0}, \mathbf{\Omega}, \mathbf{v}), \quad i = 1, \dots, n. \quad (3.5)$$

Chamaremos o modelo de mistura de SNI em modelos com erros nas variáveis de FMSNI-ME. Quando a mistura for de ST chamaremos de FMST-ME e quando for de SSL de FMSSL-ME. Note que os graus de liberdade são os mesmos para cada componente da mistura; essa suposição em um modelo com mistura não é tão restritiva como imaginamos, podemos ver isso por exemplo, em Cabral *et al.* [6], onde o modelo linear misto, que tem uma estrutura semelhante, usa a mesma ideia.

Como x_i é univariada adequamos a reparametrização dada em (2.13) da seguinte forma

$$\delta_j = \lambda_j / (1 + \lambda_j^2)^{1/2}, \quad \Delta_j = \sigma_j \delta_j \text{ e } \gamma_j^2 = \sigma_j^2 (1 - \delta_j^2). \quad (3.6)$$

Como \mathbf{R}_i em (3.4) tem distribuição SNI, então podemos usar o Teorema 4 levando em

consideração a reparametrização dada em (3.6) e escrever

$$\mathbf{R}_i|(w_i = j, U_i = u_i, T_i = t_i) \sim N_{1+p} \left[\begin{pmatrix} \mu_j + \Delta_j t_i \\ \mathbf{0}_p \end{pmatrix}, u_i^{-1} \text{block diag}\{\gamma_j^2, \Omega\} \right], \quad (3.7)$$

onde

$$T_i|U_i = u_i \sim HN(0, u_i^{-1}) \quad \text{e} \quad U_i \sim h(u|\mathbf{v}).$$

Lembrando que até agora construímos apenas as suposições para o modelo. Fica faltando então usarmos as ideias propostas em (3.4) e conseqüentemente em (3.7) e construímos o modelo com erros nas variáveis como vimos em (1.1) e (1.2). Para isso considere o vetor de observações $\mathbf{Z}_i = (X_i, \mathbf{Y}'_i)'$ para cada indivíduo i , com $i = 1, \dots, n$, e defina $\mathbf{a} = (0, \boldsymbol{\alpha}')'$, $\mathbf{b} = (1, \boldsymbol{\beta}')'$ e $\mathbf{B} = [\mathbf{b} \mathbf{I}_p]$. Então, podemos escrever

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i = \mathbf{a} + \mathbf{B}\mathbf{R}_i. \quad (3.8)$$

Note que a relação (3.7) implica em $\boldsymbol{\epsilon}_i|U_i = u_i \sim N_p(\mathbf{0}, u_i^{-1}\Omega)$, com isso facilmente obtemos a seguinte representação estocástica.

Teorema 5. (*Primeira representação estocástica para o modelo FMSNI-ME*)

$$\begin{aligned} \mathbf{Z}_i|(x_i, U_i = u_i) &\sim N_p(\mathbf{a} + \mathbf{b}x_i, u_i^{-1}\Omega); \\ x_i|(w_i = j, U_i = u_i, T_i = t_i) &\sim N(\mu_j + \Delta_j t_i, u_i^{-1}\gamma_j^2); \\ T_i|U_i = u_i &\sim HN(0, u_i^{-1}); \\ U_i &\sim h(u|\mathbf{v}); \\ P(w_i = j) &= p_j, \quad j = 1, \dots, G, \quad i = 1, \dots, n. \end{aligned}$$

Agora, usando a segunda igualdade em (3.8), combinada à equação em (3.7) e com a teoria da distribuição normal usual obtemos a segunda representação estocástica.

Teorema 6. (*Segunda representação estocástica para o modelo FMSNI-ME*)

$$\begin{aligned} \mathbf{Z}_i|(w_i = j, U_i = u_i, T_i = t_i) &\sim N_p[\mathbf{a} + \mu_j \mathbf{b} + \Delta_j \mathbf{b}t_i, u_i^{-1}(\gamma_j^2 \mathbf{b}\mathbf{b}' + \Omega)]; \\ T_i|U_i = u_i &\sim HN(0, u_i^{-1}); \\ U_i &\sim h(u|\mathbf{v}); \\ P(w_i = j) &= p_j, \quad j = 1, \dots, G, \quad i = 1, \dots, n. \end{aligned}$$

Observe que a segunda representação é obtida integrando a variável latente x_i . Uma consequência fundamental é que, combinando este resultado com o Teorema 4, podemos obter uma forma fechada para a distribuição marginal de \mathbf{Z}_i , que apresentamos no próximo teorema.

Teorema 7. *O vetor de observações de \mathbf{Z}_i tem distribuição que é uma mistura de distribuições SNI, ou seja, com densidade*

$$f(\mathbf{z}_i) = \sum_{j=1}^G p_j \text{SNI}_p(\mathbf{z}_i | \boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\Lambda}_j, \nu),$$

onde

$$\begin{aligned} \boldsymbol{\xi}_j &= \mathbf{a} + \mu_j \mathbf{b}, & \boldsymbol{\Lambda}_j &= \Delta_j \mathbf{b} & e \\ \boldsymbol{\Sigma}_j &= \gamma_j^2 \mathbf{b} \mathbf{b}' + \boldsymbol{\Omega} = \begin{pmatrix} \gamma_j^2 + \omega_0^2 & \gamma_j^2 \boldsymbol{\beta}' \\ \gamma_j^2 \boldsymbol{\beta} & \gamma_j^2 \boldsymbol{\beta} \boldsymbol{\beta}' + \boldsymbol{\Omega}_e \end{pmatrix}. \end{aligned} \quad (3.9)$$

Note que a parametrização utilizada em (3.9) é (2.13). A expressão de $f(\mathbf{z}_i)$ é obtida recuperando os parâmetros originais usando as fórmulas (2.15) e em seguida, colocando as suas respectivas expressões em (2.8). Neste trabalho vamos supor que a matriz de covariância $\boldsymbol{\Omega}_e$ é diagonal, $\boldsymbol{\Omega}_e = \text{diag}\{\omega_1^2, \dots, \omega_r^2\}$.

Capítulo 4

Estimação via MCMC

Sob um contexto Bayesiano, além das especificações feitas para as distribuições das variáveis aleatórias $(\mathbf{Z}_i, x_i, T_i, U_i)$ envolvidas no modelo que são dadas pelos Teoremas 5 e 6 condicionadas aos valores dos parâmetros do vetor $\Theta = (\alpha', \beta', \mu', \Delta', \gamma', \omega', \mathbf{p}', \nu)'$, onde $\mu = (\mu_1, \dots, \mu_G)'$, $\Delta = (\Delta_1, \dots, \Delta_G)'$, $\gamma = (\gamma_1^2, \dots, \gamma_G^2)'$, $\omega = (\omega_0^2, \omega_1^2, \dots, \omega_r^2)'$ e $\mathbf{p} = (p_1, \dots, p_G)'$, também temos que especificar distribuições para esses mesmos parâmetros citados. Essas distribuições são chamadas de *distribuições a priori* e as especificaremos através da notação $\pi(\Theta)$. A partir de agora $\pi(\cdot)$ sempre denotará uma densidade; $\pi(\mathbf{Z}|\Theta)$ denotará a distribuição de \mathbf{Z} dado Θ , que é a verossimilhança, enquanto $\pi(\Theta|\mathbf{Z})$ denotará a densidade a posteriori de Θ , por exemplo.

Uma distribuição a priori para os parâmetros de Θ deverá conter e descrever nossa incerteza a respeito desse vetor de parâmetros baseado em um conhecimento prévio a respeito do problema. No nosso caso, usaremos distribuições a priori não informativas.

A escolha de uma distribuição a priori está totalmente ligada à classe que cada parâmetro pertence, sendo elas de locação, escala e assimetria ou forma. Existe um conhecimento que sempre nos ajuda a escolher uma distribuição conveniente para o parâmetro quando, por exemplo, sabemos que o vetor de parâmetros de locação $\mu = (\mu_1, \dots, \mu_n)'$ contém componentes que pertencem à reta, logo podemos definir uma distribuição que expresse isso, como uma normal multivariada. Outro exemplo é para γ_i^2 , $i = 1, \dots, n$, que é um parâmetro de escala positivo; podemos expressar essa positividade atribuindo a ele uma distribuição a priori Gamma Inversa que tem como suporte os reais positivos. Todos os parâmetros das distribuições a priori serão denominados de hiperparâmetros.

As distribuições a priori usadas nesse trabalho seguiram esse pensamento de especificação dependendo da sua classe. Definimos então

$$\alpha \sim N_r(\mathbf{c}_\alpha, \mathbf{D}_\alpha), \quad \beta \sim N_r(\mathbf{c}_\beta, \mathbf{D}_\beta), \quad \mu \sim N(a_\mu, b_\mu), \quad \Delta \sim N(c_\Delta, d_\Delta),$$

onde, em geral, as matrizes de covariância acima são positivas definidas e diagonais.

Para o parâmetro de dispersão γ , adotamos uma distribuição a priori hierárquica defi-

nida por

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

Esse esquema hierárquico é uma adaptação de uma sugestão dada por Richardson & Green [23] para um modelo de misturas de normais univariadas. Além disso, definimos $\omega_i^{-2} \sim \text{Gamma}(l, m)$, $i = 0, 1, \dots, r$, onde e, g, h, l e m são positivos. Para os vetores de pesos nós usamos uma suposição habitual para modelos de mistura,

$$\mathbf{p} \sim \text{Dir}(\kappa_1, \dots, \kappa_G), \quad (4.1)$$

isto é, uma distribuição Dirichlet com hiperparâmetros conhecidos e positivos.

Para estimação dos parâmetros usaremos um algoritmo que está incluído na classe dos algoritmos MCMC (*Markov Chain Monte Carlo*). Para desenvolvermos o algoritmo de uma forma geral, é necessário utilizar o procedimento de “aumentar” o conjunto de dados, que são as chamadas representações estocásticas, onde usamos o condicionamento em variáveis latentes como um auxílio para transformar o problema de gerar amostras de distribuições complicadas em gerações a partir de distribuições “standard”. No nosso caso utilizamos a representação dada no Teorema 2, onde o problema de gerar a partir de uma SN tranforma-se em um de gerar a partir de uma HN e depois de uma Normal, o que o torna de bem mais fácil solução.

Cada modelo pertencente à família SNI tem um parâmetro de escala com interpretação específica e obviamente merece um tratamento diferenciado para a sua distribuição a priori. Por exemplo, existem várias sugestões para estimar os graus de liberdade do modelo t de Student assimétrico (ver uma excelente discussão em [14]). Nós não iremos tratar esta questão com profundidade, mas adotaremos uma escolha de distribuição a priori que será útil para os nossos propósitos. Ela consiste em uma distribuição a priori exponencial com parâmetro $\lambda > 0$ para ν , e inclui um segundo nível de hierarquia determinado por $\lambda \sim U(\lambda_0, \lambda_1)$, em que $0 < \lambda_0 < \lambda_1$. Isto é, para o modelo t de Student assimétrico o parâmetro graus de liberdade ν terá uma distribuição a priori exponencial com média $1/\lambda$, e λ é dado por uma uniforme no intervalo (λ_0, λ_1) . Já para o modelo slash assimétrico teremos, $\nu \sim \text{Gamma}(\varphi_{sl}, \psi_{sl})$, onde φ_{sl} e ψ_{sl} são positivos.

Outro ponto que devemos atentar é o da suposição de que os parâmetros de Θ sejam independentes entre si a priori

$$\pi(\Theta) = \pi(\alpha)\pi(\beta)\pi(\mu)\pi(\Delta)\pi(\mathbf{p})\pi(\nu|\lambda)\pi(\lambda) \left(\prod_{j=0}^r \pi(\omega_j^2) \right) \left(\prod_{j=1}^G \pi(\gamma_j^2|f) \right) \pi(f).$$

4.1 Algoritmo tipo Gibbs

O modelo FMSNI-ME permite um desenvolvimento direto de vários tipos de algoritmos de Gibbs (Gibbs samplers), através de representações estocásticas dadas nos Teoremas 5 e 6.

O algoritmo que usaremos consiste em uma amostragem marginal dos parâmetros a partir das distribuições condicionais completas usando somente a representação estocástica dada no Teorema 5, isto é, os parâmetros são atualizados um de cada vez.

Sobre propriedades do amostrador de Gibbs e referências envolvendo resultados teóricos mais gerais em torno de algoritmos do tipo MCMC veja os artigos de Casella & George [10] e Gelfand [18] e o livro de Gamerman & Lopes [17].

Para obtermos uma amostra da distribuição a posteriori do modelo FMSNI-ME basta seguirmos os passos do seguinte algoritmo.

Passo 1. Pra cada $i = 1, \dots, n$, gere w_i independentemente a partir da seguinte distribuição discreta

$$P(w_i = j | \mathbf{Z}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{p}, \boldsymbol{\nu}) = \frac{p_j \text{SNI}_p(\mathbf{Z}_i | \boldsymbol{\xi}_j, \boldsymbol{\Sigma}_j, \boldsymbol{\Lambda}_j, \boldsymbol{\nu})}{\sum_{k=1}^G p_k \text{SNI}_p(\mathbf{Z}_i | \boldsymbol{\xi}_k, \boldsymbol{\Sigma}_k, \boldsymbol{\Lambda}_k, \boldsymbol{\nu})}, \quad j = 1, \dots, G,$$

onde $\boldsymbol{\xi}_j$, $\boldsymbol{\Sigma}_j$ e $\boldsymbol{\Lambda}_j$ são dados por (3.9).

Passo 2. Obtenha uma amostra de $\mathbf{p} = (p_1, \dots, p_G)'$ a partir da distribuição $\pi(\mathbf{p} | \mathbf{w})$, que é a distribuição Dirichlet da forma

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

onde o termo m_j indica a quantidade de indivíduos alocados na classe j , para $j = 1, \dots, G$.

Passo 3. Para cada $i = 1, \dots, n$, gere t_i independentemente a partir da distribuição $\pi(t_i | \mathbf{Z}_i, u_i, w_i = j, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\gamma}, \boldsymbol{\omega})$, que é uma NT(μ_{T_i} , $\sigma_{T_i}^2$, (0, 1)) com

$$\sigma_{T_i}^2 = \left(\Delta_j^2 u_i \gamma_j^{-2} + u_i \right)^{-1} \quad \text{e} \quad \mu_{T_i} = \sigma_{T_i}^2 \left[(x_i - \mu_j) \Delta_j u_i \gamma_j^{-2} \right].$$

Passo 4. Para cada $i = 1, \dots, n$, gere u_i independentemente, a partir da distribuição $\pi(u_i | \mathbf{Z}_i, t_i, w_i = j, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\nu})$, que é igual a seguinte distribuição

a) $\text{Gamma}(\text{shape}_{u_i}, \text{rate}_{u_i})$ se $U_i \sim \text{Gamma}(\nu/2, \nu/2)$, ou seja, modelo FMST-ME,

onde

$$\begin{aligned} \text{shape}_{u_i} &= \frac{r+2+\mathbf{v}}{2} \quad \text{e} \\ \text{rate}_{u_i} &= \frac{1}{2} \left[(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) + \frac{1}{\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 + t_i^2 + \mathbf{v} \right]. \end{aligned}$$

b) $\text{TG}(\text{shape}_{u_i}, \text{rate}_{u_i}, (0, 1))$ se $U_i \sim \text{Beta}(\mathbf{v}, 1)$, ou seja, modelo FMSSL-ME, onde

$$\begin{aligned} \text{shape}_{u_i} &= r/2 + 1 + \mathbf{v} \quad \text{e} \\ \text{rate}_{u_i} &= \frac{1}{2} \left[(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) + \frac{1}{\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 + t_i^2 \right], \end{aligned}$$

onde $\text{TG}(\cdot, (a, b))$ é a notação usada para a distribuição gamma truncada no intervalo (a, b) .

Passo 5. Para cada $i = 1, \dots, n$, gere x_i independentemente, a partir da distribuição $\pi(x_i | \mathbf{Z}_i, t_i, u_i, w_i = j, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \gamma, \boldsymbol{\omega})$, que é uma $N(\mu_{x_i}, u_i^{-1} \sigma_{x_i}^2)$, onde

$$\begin{aligned} \mu_{x_i} &= \sigma_{x_i}^2 \left[X_i \omega_0^{-2} + (\mathbf{Y}_i - \boldsymbol{\alpha})' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\beta} + \gamma_j^{-2} (\mu_j + \Delta_j t_i) \right] \quad \text{e} \\ \sigma_{x_i}^2 &= \left(\omega_0^{-2} + \boldsymbol{\beta}' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\beta} + \gamma_j^{-2} \right)^{-1}. \end{aligned}$$

Passo 6. Gere μ_j a partir da distribuição $\pi(\mu_j | \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{T}, \mathbf{U}, \mathbf{x}, \Delta_j, \gamma_j^2, \mathbf{w})$, $j = 1, \dots, G$, independentemente, que é uma $N(\mu_{\mu_j}, \sigma_{\mu_j}^2)$ onde

$$\sigma_{\mu_j}^2 = \left(\gamma_j^{-2} \sum_{i=1}^n u_i + b_{\mu}^{-1} \right)^{-1} \quad \text{e} \quad \mu_{\mu_j} = \sigma_{\mu_j}^2 \left(\gamma_j^{-2} \sum_{i=1}^n (x_i - \Delta_j t_i) u_i + a_{\mu} b_{\mu}^{-1} \right).$$

Passo 7. Gere Δ_j a partir da distribuição $\pi(\Delta_j | \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{T}, \mathbf{x}, \gamma_j^2, \mu_j, \mathbf{w})$, $j = 1, \dots, G$, independentemente, que é a seguinte distribuição $N(\mu_{\Delta_j}, \sigma_{\Delta_j}^2)$ onde

$$\sigma_{\Delta_j}^2 = \left(\gamma_j^{-2} \sum_{i=1}^n u_i t_i^2 + d_{\Delta}^{-1} \right)^{-1} \quad \text{e} \quad \mu_{\Delta_j} = \sigma_{\Delta_j}^2 \left(\gamma_j^{-2} \sum_{i=1}^n u_i (x_i - \mu_j) t_i + d_{\Delta}^{-1} c_{\Delta} \right).$$

Passo 8. Obtenha uma amostra de $\boldsymbol{\alpha}$ a partir da distribuição $\pi(\boldsymbol{\alpha} | \mathbf{Z}, \mathbf{t}, \mathbf{u}, \mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \gamma, \boldsymbol{\omega})$, que é uma densidade normal com média

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \left[\sum_{i=1}^n u_i \boldsymbol{\omega}^{-1} (\mathbf{Y}_i - x_i \boldsymbol{\beta}) + \mathbf{D}_{\boldsymbol{\alpha}}^{-1} \mathbf{C}_{\boldsymbol{\alpha}} \right]$$

e matriz de covariâncias

$$\Sigma_{\alpha} = \left(\omega^{-1} \sum_{i=1}^n u_i + \mathbf{D}_{\alpha}^{-1} \right)^{-1}.$$

Passo 9. Obtenha uma amostra de β a partir da distribuição $\pi(\beta|\mathbf{Z}, \mathbf{t}, \mathbf{w}, \mathbf{u}, \mathbf{x}, \alpha, \mu, \Delta, \gamma, \omega)$, que é a densidade de uma normal com média

$$\mu_{\beta} = \Sigma_{\beta} \left[\Omega_e^{-1} \sum_{i=1}^n u_i x_i (\mathbf{Y}_i - \alpha) + \mathbf{D}_{\beta}^{-1} \mathbf{c}_{\beta} \right]$$

e matriz de covariância

$$\Sigma_{\beta} = \left(\Omega_e^{-1} \sum_{i=1}^n u_i x_i^2 + \mathbf{D}_{\beta}^{-1} \right)^{-1}.$$

Passo 10. Gere γ_j^2 a partir da distribuição $\pi(\gamma_j^2|\mathbf{Z}, \mathbf{t}, \mathbf{w}, \alpha, \mu, \Delta, \omega)$, $j = 1, \dots, G$, independentemente, que é a seguinte distribuição

$$\text{GI} \left(e + m_j/2, f + (1/2) \sum_{i \in A_j} u_i (x_i - \mu_j - \Delta_j t_i)^2 \right),$$

onde $A_j = \{i \in \{1, \dots, n\}, w_i = j\}$ e GI denota a distribuição Gamma Inversa. Vamos aproveitar o espaço para defini-la. Se $X \sim \text{GI}(r, \mu)$ então

$$f(x) = \frac{1}{\Gamma(r)\mu^r} x^{-r-1} e^{-1/x\mu}, \quad x > 0, \quad (4.2)$$

onde r regula a forma e μ é o inverso do parâmetro de escala.

Passo 11. Gere f a partir da distribuição $\pi(f|\gamma)$, que é uma $\text{Gamma}(e + g, h + \sum_{i=1}^G \gamma_i^{-2})$.

Passo 12. Gere ω_k^2 a partir da distribuição $\pi(\omega_k^2|\mathbf{Z}, \mathbf{t}, \mathbf{w}, \mathbf{u}, \mathbf{x}, \alpha, \mu, \Delta, \gamma)$, $k = 0, \dots, r$, independentemente, de acordo com as seguintes condições

(a) Gere ω_0^2 a partir da distribuição

$$\text{GI} \left[n/2 + l, m + (1/2) \sum_{i=1}^n u_i (X_i - x_i)^2 \right];$$

(b) Gere ω_k^2 , $k = 1, \dots, r$, a partir da distribuição

$$\text{GI} \left[n/2 + l, m + (1/2) \sum_{i=1}^n u_i (Y_{ik} - \alpha_k - \beta_k x_i)^2 \right].$$

Passo 13. A amostragem de ν deve ser realizada considerando a escolha da distribuição do fator de escala, onde cada caso merece um tratamento específico. Por exemplo,

(a) Caso t de Student assimétrica (ST):

(a.1) Gere λ com densidade $\pi(\lambda|\nu)$, que é a distribuição TG(2, ν , (λ_0, λ_1)).

(a.2) Usando um passo de Metropolis-Hastings, gere ν com a seguinte distribuição marginal condicional

$$\pi(\nu|\alpha, \beta, \mu, \Delta, \gamma, \omega, \mathbf{p}, \lambda, \mathbf{Z}) \propto \exp(-\lambda \nu) \prod_{i=1}^n \sum_{j=1}^G p_j \text{ST}_p(\mathbf{z}_i|\xi_j, \Sigma_j, \Lambda_j, \nu). \quad (4.3)$$

A proposta é a seguinte: dada a observação $\nu^{(j-1)}$ obtida na iteração $j-1$, gere um candidato a nova observação ν^* por uma distribuição lognormal

$$\text{LN}(\log \nu^{(j-1)}, \delta_\nu^2).$$

A nova observação ν^* é aceita com probabilidade

$$\min \left\{ \frac{\pi(\nu^*|\dots)\nu^*}{\pi(\nu^{(j-1)}|\dots)\nu^{(j-1)}}, 1 \right\},$$

onde $\pi(\nu^*|\dots)$ denota (4.3) avaliada usando os valores atualizados de $\alpha, \beta, \mu, \Delta, \gamma, \omega, \mathbf{p}$ e λ .

(c) Caso Slash Assimétrica: Gere ν com densidade $\pi(\nu|\mathbf{u})$, que é a distribuição

$$\text{Gamma}(n + \phi_{sl}, \psi_{sl} - \sum_{i=1}^n \log u_i).$$

Para os passos de Metropolis-Hastings regularemos a quantidade δ_ν^2 de forma a ter uma taxa de aceitação no intervalo (0.15, 0.3). Todos os detalhes da obtenção das condicionais completas podem ser encontrados no apêndice B.

Capítulo 5

Comparação de Modelos

Até agora estudamos vários tipos de modelagens, mas partindo de um problema real como saberemos qual o melhor modelo para se usar? Considere o problema de comparar vários modelos FMSNI-ME, com diferentes fatores de escala, tamanhos amostrais ou número de componentes na mistura de densidades. Fica clara a necessidade de uma regra pra mensurar qual a melhor modelagem. Na literatura existem vários critérios para a comparação de modelos, principalmente para modelos clássicos (estatística frequentista) como AIC, BIC e EDC, porém com nossa abordagem Bayesiana usaremos um critério chamado DIC (*deviance information criterion*).

5.1 DIC

O critério usado para compararmos os modelos será o DIC_{obs} (DIC Observado). Esse critério é baseado nas amostras MCMC da distribuição a posteriori dos parâmetros, sendo uma versão modificada do critério DIC original (Spiegelhalter *et al.* [28]), já que a definição original do DIC não é adequada para modelos de mistura. Isso se dá porque uma condição básica para o uso correto do DIC original é que a média à posteriori seja uma boa estimativa. Isso não é necessariamente verdade no caso de mistura, por conta de um fenômeno que é conhecido como *label switching*, que em português pode ser traduzido como troca de rótulos. O DIC_{obs} foi sugerido por Celeux *et al.* [11], que em sua Seção 3.1 o chamou de DIC_3 .

O fenômeno *label switching* se origina de uma característica básica dos modelos de misturas que é a invariância a permutações dos índices das componentes. Isto implica que o parâmetro θ_j (referente a componente j) não é identificável marginalmente. Formalmente, isto significa que, para todas permutações $\sigma, \sigma' \in \mathcal{P}$ onde \mathcal{P} é um conjunto de índices permutáveis de $j = 1, \dots, G$, a verossimilhança associada ao ponto $(\theta_{\sigma(1)}, \dots, \theta_{\sigma(k)})$ é a mesma a que é associada ao ponto $(\theta_{\sigma'(1)}, \dots, \theta_{\sigma'(k)})$. Se usamos uma distribuição a priori permutável, também teremos uma distribuição a posteriori permutável. Isto im-

plica que as distribuições a posteriori marginais são todas iguais, inviabilizando qualquer procedimento inferencial. Por exemplo, a esperança a posteriori de θ_1 é idêntica a de θ_2 . Assim, devemos utilizar outros estimadores, diferentes da esperança a posteriori. Algumas referências que tratam da solução desse problema são Sperrin *et al.* [27] e Yao [29].

O DIC observado é dado por

$$\text{DIC}_{\text{obs}} = \overline{\text{D}(\Theta)} + \tau_D,$$

onde $\Theta = (\xi, \Lambda, \Sigma, \nu, \mathbf{p})$ são os parâmetros em foco. Estes parâmetros são funções dos parâmetros originais, ver (3.9). Temos que

$$\overline{\text{D}(\Theta)} = -2\text{E}\{\log[L(\mathbf{z}|\Theta)|\mathbf{z}]\}, \quad \text{onde} \quad L(\mathbf{z}|\Theta) = \prod_{i=1}^n L(\mathbf{z}_i|\Theta),$$

$$L(\mathbf{z}_i|\Theta) = \sum_{j=1}^G p_j \text{SNI}_p(\mathbf{z}_i|\xi_j, \Sigma_j, \Lambda_j, \nu)$$

e $L(\mathbf{z}|\Theta)$ é a função de verossimilhança (3.9). Sabemos que

$$\text{E}\{\log[L(\mathbf{z}|\Theta)|\mathbf{z}]\} = \int \log[L(\mathbf{z}|\Theta)]\pi(\Theta|\mathbf{z})d\Theta,$$

onde $\pi(\Theta|\mathbf{z})$ é a densidade a posteriori de Θ , τ_D é uma espécie de mensuração do número efetivo de parâmetros envolvidos no modelo, sendo dado por

$$\tau_D = \overline{\text{D}(\Theta)} + 2\log(\text{E}[L(\mathbf{z}|\Theta)|\mathbf{z}]).$$

Logo, temos como expressão final

$$\text{DIC}_{\text{obs}} = -4\text{E}\{\log[L(\mathbf{z}|\Theta)|\mathbf{z}]\} + 2\log(\text{E}[L(\mathbf{z}|\Theta)|\mathbf{z}]). \quad (5.1)$$

Seja $\Theta^{(l)}$ a l -ésima amostra da distribuição a posteriori de Θ , ou seja, a amostra MCMC na iteração l do algoritmo, $l = 1, \dots, m$. A esperança $\text{E}\{\log[L(\mathbf{z}|\Theta)|\mathbf{z}]\}$ pode ser aproximada por

$$\overline{D} = \frac{1}{m} \sum_{l=1}^m \sum_{i=1}^n \log[L(\mathbf{z}_i|\Theta^{(l)})].$$

Uma recomendação de Celeux *et al.* (2006), é que o segundo termo de (5.1) seja aproximado pela expressão

$$2 \sum_{i=1}^n \log \widehat{\pi}(\mathbf{z}_i),$$

onde

$$\widehat{\pi}(\mathbf{z}_i) = \frac{1}{m} \sum_{l=1}^m L(\mathbf{z}_i | \Theta^{(l)}),$$

que é a estimativa MCMC da densidade a posteriori, chamada de densidade preditiva $E[L(\mathbf{z}_i | \Theta) | \mathbf{z}]$. É importante ressaltarmos que esta estimativa da densidade através das preditivas é robusta contra o fenômeno *Label Switching*, contornando assim a principal desvantagem de usar o DIC original para modelos de mistura.

A fim de comparar os diferentes modelos, escolheremos como melhores ajustes os modelos que tiverem os menores DIC_{obs} .

Capítulo 6

Aplicação com dados simulados e reais

Agora passamos para a parte prática do trabalho onde iremos verificar a adequação dos modelos da família FMSNI-ME em comparação com outros modelos mais parcimoniosos como o Normal e t de Student. Faremos isso de duas formas, simulando dados artificiais e usando dados reais e comparando as modelagens através do DIC_{obs} .

Os valores iniciais e de hiperparâmetros assim como os algoritmos de geração e estimação para todos os modelos podem ser encontrados no apêndice A.

6.1 Dados simulados

6.1.1 Simulação 1

Nessa simulação construiremos uma amostra de tamanho 200 com a seguinte estrutura e composição de parâmetros:

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i, \quad x_i \sim \sum_{j=1}^2 p_j \text{ST}(\mu_j, \gamma_j^2, \Delta_j, \nu), \quad \boldsymbol{\epsilon}_i | U_i = u_i \sim N_p(\mathbf{0}, u^{-1}\boldsymbol{\Omega})$$

e $U_i \sim \text{Gamma}(\nu/2, \nu/2)$

com

$$\mathbf{a} = (0, 0.4, 0.1), \quad \mathbf{b} = (1, 0.8, 0.9), \quad \boldsymbol{\mu} = (2, 8), \quad \boldsymbol{\Delta} = (-2, 2), \quad (6.1)$$
$$\boldsymbol{\gamma}^2 = (0.1, 0.1), \quad \boldsymbol{\Omega} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.4 \end{pmatrix}, \quad p = (0.7, 0.3), \quad \nu = 3.$$

É interessante salientarmos que a disposição dos dados será em uma matriz \mathbf{Z} com três colunas e $n = 200$ linhas, onde na primeira coluna de \mathbf{Z} encontramos as amostras da variável substituta X e nas demais colunas as dos vetores de respostas \mathbf{Y} .

Nessa primeira parte esperamos recuperar através da estimação Bayesiana os verda-

deiros valores para os parâmetros, dado que usaremos o modelo FMST-ME para modelar dados gerados sob a mesma estrutura, evidenciando assim que o procedimento de estimação funciona para esse tipo de modelo.

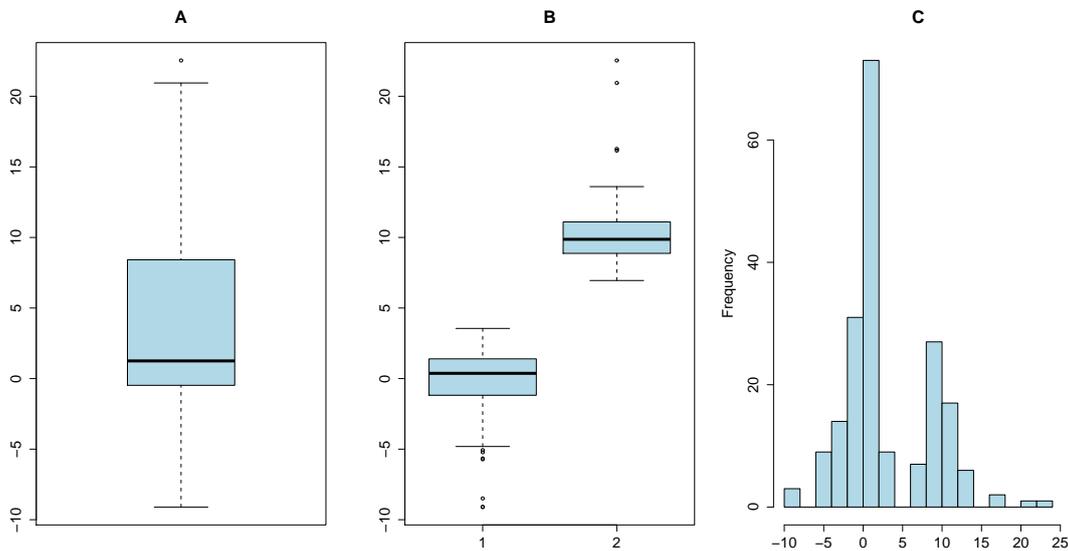


Figura 6.1: Simulação 1 modelo FMST-ME (A) boxplot de X , (B) boxplot da variável substituta X pela classificadora w , (C) histograma de X .

Na Figura 6.1 vemos os boxplots e histograma da variável substituta X_i conhecida também como “instrumento de referência” no âmbito da calibragem de instrumentos, gerada no modelo FMST-ME com duas componentes. No segundo boxplot podemos ver a mesma X_i disposta em classes pela variável w .

Para a estimação usaremos um algoritmo com $m = 25000$ iterações onde em cada iteração o vetor Θ terá seus parâmetros atualizados um de cada vez e ao final usaremos uma técnica chamada *burn-in* onde descartaremos as cinco mil primeiras amostras MCMC. Como a estimação é feita através da média das amostras MCMC, usaremos essa técnica de *burn-in* para que as estimativas não sofram tanto a influência das amostras obtidas no início do processo de geração. A fim de diminuir o efeito da autocorrelação fruto da própria cadeia do algoritmo de Gibbs guardaremos apenas as observações geradas de 5 em 5 interações (*lag*) totalizando ao final uma amostra de tamanho 4 mil.

Na Figura 6.2 vemos gráficos chamados de *traceplots* que nos mostram a trajetória percorrida pela amostra MCMC de um dado parâmetro. Na Figura temos apenas alguns referentes ao modelo FMST-ME, porém nesses já podemos observar o traço homogêneo que percorre todo o suporte do parâmetro em destaque. Os traços indicam que a média amostral é uma boa estimativa, como podemos observar na Figura 6.3.

Além dos modelos FMST-ME e FMSSL-ME serão ajustados também os modelos com erros nas variáveis onde a variável latente é modelada por uma mistura de normais com distribuição normal (FMN-ME), t de Student (FMT-ME) e Normal assimétrica (FMSN-

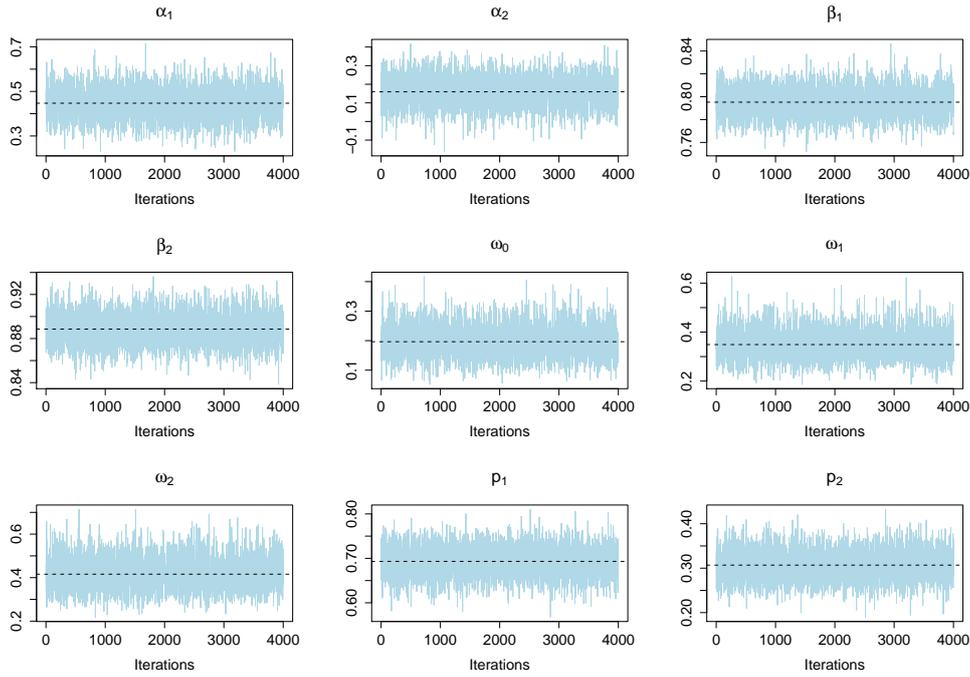


Figura 6.2: Traceplots de alguns parâmetros do modelo FMST-ME. A linha pontilhada corresponde a estimativa do parâmetro.

ME).

Esperamos, é claro, que o ajuste ST seja o melhor, sabendo que os dados gerados de acordo com um modelo FMST-ME. Observamos que isso realmente ocorreu, como podemos ver nas estimativas dispostas na Tabela 6.1 e no gráfico das densidades estimadas para cada modelo (Figura 6.4). Outro modelo que se destacou com um bom ajuste e desvios pequenos foi o FMSSL-ME. Outras medidas interessantes do ponto de vista da estimação são os intervalos de credibilidade, que podemos ver na Tabela 6.2.

Comparando os modelos FMN-ME, FMT-ME, FMSN-ME e FMST-ME, temos dispostos na Tabela 6.3 os valores do critério DIC_{obs} . Vemos então que os melhores ajustes foram para os modelos FMST-ME e FMSS-ME.

6.1.2 Simulação 2

Nesse segundo experimento vamos mostrar a flexibilidade dos modelos FMST-ME e FMSSL-ME, para ajustar um modelo de natureza diferente dos usados até agora. A proposta é criar um modelo de regressão com erros nas variáveis onde a variável latente tenha uma distribuição de misturas de densidades de uma natureza totalmente diferente das SNI. Essa distribuição é a chamada Normal Inversa Gaussiana (NIG), que é uma mistura de escala de uma Normal e uma inversa gaussiana. Então vamos supor que a variável latente tem distribuição mistura de normais inversas gaussianas.

Definição 6. Dizemos que uma variável aleatória U tem distribuição inversa Gaussiana

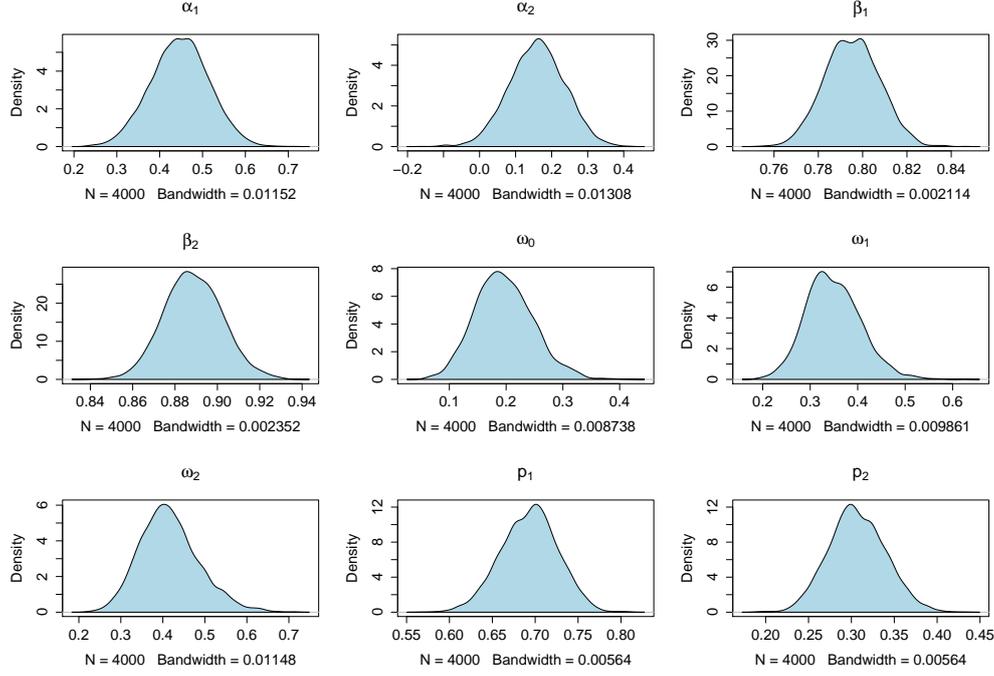


Figura 6.3: Estimador de Kernel de alguns parâmetros do modelo FMST-ME. Para mais detalhes veja a função density no pacote stats do programa R ou a forma teórica em Scott [26].

Tabela 6.1: Estimativas dos parâmetros para vários modelos com erros nas variáveis para os dados simulados FMST-ME. A sigla dp denota o desvio padrão da amostra MCMC e RV o real valor dos parâmetros

Parâmetro	RV	FMN-ME		FMT-ME		FMSN-ME		FMST-ME		FMSSL-ME	
		Estimativa	dp								
α_1	0.4	0.40670	0.0859	0.45385	0.0671	0.40097	0.0848	0.44716	0.0675	0.44648	0.0673
α_2	0.1	0.16581	0.0920	0.15685	0.0728	0.15453	0.0895	0.15949	0.0768	0.17489	0.0776
β_1	0.8	0.79627	0.0140	0.79361	0.0115	0.79952	0.0137	0.79523	0.0123	0.79425	0.0120
β_2	0.9	0.88895	0.0148	0.88715	0.0130	0.89270	0.0143	0.88856	0.0137	0.88623	0.0131
ω_0	0.2	0.16803	0.1136	0.15212	0.0548	0.29943	0.0798	0.19642	0.0510	0.11761	0.0329
ω_1	0.3	1.01743	0.1235	0.33057	0.0566	0.94643	0.1137	0.34900	0.0576	0.20731	0.0370
ω_2	0.4	1.13992	0.1450	0.41010	0.0701	1.05065	0.1276	0.41582	0.0705	0.24997	0.0431
μ_1	2	-0.26749	0.2022	0.26786	0.1624	2.28442	0.1440	1.97809	0.1245	2.01992	0.1409
μ_2	8	10.37460	0.4021	9.70907	0.2161	7.44090	0.2502	7.94272	0.2994	7.89963	0.3447
Δ_1	-2	-	-	-	-	-3.41324	0.2358	-2.2218	0.2302	-1.7937	0.1818
Δ_2	2	-	-	-	-	4.27577	0.4353	2.29373	0.4271	1.75797	0.3417
γ_1^2	0.1	5.35514	0.7135	2.41234	0.4538	0.06024	0.0788	0.10708	0.0948	0.05852	0.0615
γ_2^2	0.1	8.63311	1.8874	2.03686	0.5771	0.06669	0.0995	0.24678	0.2973	0.14936	0.1791
ν	3	-	-	2.98268	0.4615	-	-	3.41924	0.5551	1.17965	0.1487
ρ_1	0.7	0.69112	0.0324	0.69895	0.0331	0.6940	0.0328	0.69317	0.0329	0.69308	0.0323
ρ_2	0.3	0.30887	0.0324	0.30104	0.0331	0.3059	0.0328	0.30682	0.0329	0.30691	0.0323

quando tem densidade dada por

$$g(u) = \frac{\delta}{\sqrt{2\pi}} u^{-3/2} \exp \left\{ -\frac{1}{2} \left(\frac{\delta^2}{u} + \gamma^2 u - 2\delta\gamma \right) \right\}, \quad u > 0,$$

onde $\gamma > 0$ e $\delta > 0$.

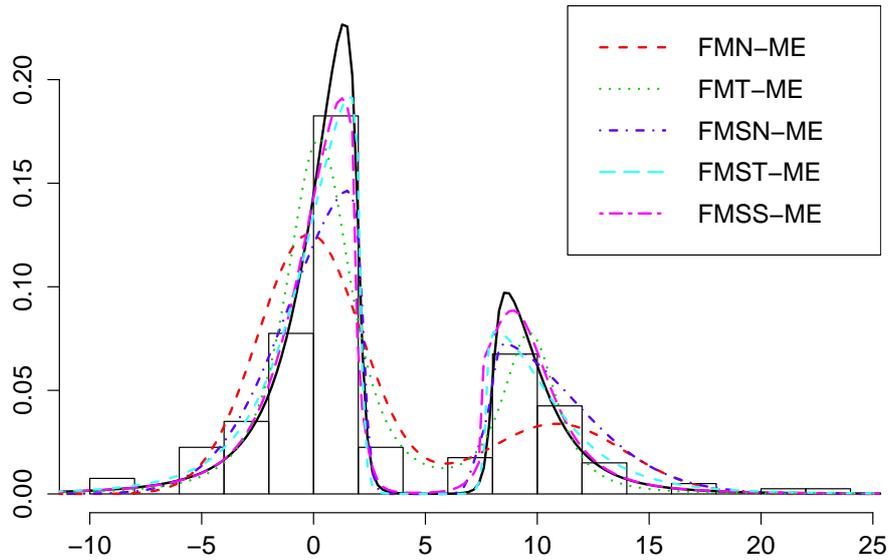


Figura 6.4: Preditiva dos modelos com erros nas variáveis com dados FMST-ME. A linha contínua de cor preta corresponde à densidade real de x .

Tabela 6.2: Intervalos de credibilidade para estimação dos modelos FMST-ME e FMSSL-ME para os dados simulados FMST-ME, 95% de credibilidade.

parâmetro	FMST-ME		FMSSL-ME	
	limite inferior	limite superior	limite inferior	limite superior
α_1	0.3196	0.5817	0.3171	0.5800
α_2	0.0265	0.3266	0.0149	0.3108
β_1	0.7694	0.8169	0.7712	0.8188
β_2	0.8622	0.9138	0.8616	0.9150
ω_0	0.0538	0.1787	0.0991	0.2983
ω_1	0.1426	0.2809	0.2339	0.4591
ω_2	0.1696	0.3363	0.2820	0.5560
μ_1	1.7760	2.3230	1.7440	2.2270
μ_2	7.3040	8.5540	7.4150	8.5830
Δ_1	-2.118	-1.427	-2.685	-1.803
Δ_2	1.0440	2.3320	1.4750	3.2100
γ_1^2	0.0000	0.1808	0.0002	0.2878
γ_2^2	0.0000	0.5168	0.0002	0.8579
ν	0.9176	1.4940	2.4000	4.5050
\mathbf{p}_1	0.6299	0.7557	0.6298	0.7571
\mathbf{p}_2	0.2443	0.3701	0.2429	0.3702

Tabela 6.3: Critério de seleção para vários modelos com erros nas variáveis para os dados simulados FMST-ME.

	FMN-ME	FMT-ME	FMSN-ME	FMST-ME	FMSSL-ME
<i>log-verossimilhança</i>	-1242.812	-1106.296	-1147.723	-1065.219	-1065.235
τ_D	19.91489	21.96576	21.30651	15.66476	16.66865
DIC_{obs}	2505.539	2234.557	2316.752	2146.103	2147.138

Utilizaremos a notação $U \sim IG(\gamma, \delta)$. Uma descrição detalhada pode ser encontrada em Barndorff-Nielsen [4].

Definição 7. Se um vetor aleatório \mathbf{X} p -dimensional tem distribuição NIG então admite a seguinte representação estocástica

$$\mathbf{X}|U = u \sim N_p(\boldsymbol{\mu} + u\boldsymbol{\Delta}\boldsymbol{\lambda}, u\boldsymbol{\Delta}), \quad U \sim \text{IG}(\gamma, \delta),$$

onde $\boldsymbol{\mu}$ e $\boldsymbol{\lambda}$ são vetores de parâmetros p -dimensionais, $\boldsymbol{\Delta}$ é $p \times p$ uma matriz positiva definida, γ e δ são maiores que zero.

Usaremos a notação $\mathbf{X} \sim \text{NIG}(\boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\lambda}, \gamma, \delta)$. Além disso, para assegurar a identificabilidade, usaremos restrição $\det(\boldsymbol{\Delta}) = 1$, isso implica que $\boldsymbol{\Delta} = 1$ no caso univariado. O vetor de médias e a matrix de covariâncias são dados por

$$E(\mathbf{X}) = \boldsymbol{\mu} + (\delta/\gamma)\boldsymbol{\Delta}\boldsymbol{\lambda}, \quad \text{Var}(\mathbf{X}) = (\delta/\gamma^3)(\gamma^2\boldsymbol{\Delta} + \boldsymbol{\Delta}\boldsymbol{\lambda}\boldsymbol{\lambda}'\boldsymbol{\Delta}). \quad (6.2)$$

Observe que

$$\boldsymbol{\lambda} = \mathbf{0}_p \Rightarrow E(\mathbf{X}) = \boldsymbol{\mu} \text{ e } \text{Var}(\mathbf{X}) = (\delta/\gamma)\boldsymbol{\Delta}. \quad (6.3)$$

Nesse caso, se $\boldsymbol{\Delta}$ é diagonal, as coordenadas de \mathbf{X} são não correlacionadas. Os parâmetros da distribuição afetam a sua forma do seguinte modo : o parâmetro $\alpha^2 = \gamma^2 + \boldsymbol{\lambda}'\boldsymbol{\Delta}\boldsymbol{\lambda}$ regula as caudas, onde valores menores implicam em caudas mais pesadas, $\boldsymbol{\Delta}$ está relacionado a escala da distribuição, $\boldsymbol{\mu}$ é o parâmetro de locação e $\boldsymbol{\lambda}$ é o parâmetro que regula a assimetria. Finalmente, quando γ e δ tendem para o infinito, a variável aleatória converge em distribuição para a distribuição normal multivariada.

Agora vamos definir um modelo de regressão com erros nas variáveis que chamaremos de NIG-ME. Assim como antes, vamos considerar a estrutura funcional (1.1) e (1.2). As quantidades $\boldsymbol{\epsilon}_i$ e \mathbf{R}_i também tem a mesma definição. Vamos agora supor que

$$\mathbf{R}_i \stackrel{\text{iid}}{\sim} \text{NIG}_{1+p} [(\boldsymbol{\mu}, \mathbf{0}'_p)', \text{block diag}\{1, \boldsymbol{\Omega}\}, (\boldsymbol{\lambda}, \mathbf{0}'_p)', \gamma, \delta], \quad i = 1, \dots, n,$$

onde $\det(\boldsymbol{\Omega}) = 1$. Pela Definição 7 temos que, para $i = 1, \dots, n$, existe uma variável aleatória U_i tal que

$$\mathbf{R}_i|U_i = u_i \sim N_{1+p} \left[\begin{pmatrix} \boldsymbol{\mu} + u_i\boldsymbol{\lambda} \\ \mathbf{0}_p \end{pmatrix}, \begin{pmatrix} u_i & \mathbf{0}'_p \\ \mathbf{0}_p & u_i\boldsymbol{\Omega} \end{pmatrix} \right], \quad U_i \sim \text{IG}(\gamma, \delta). \quad (6.4)$$

Usando a relação (6.2), podemos provar que

$$\text{Var}(\mathbf{R}_i) = (\delta/\gamma^3) \begin{pmatrix} \gamma^2 + \boldsymbol{\lambda}^2 & \mathbf{0}'_p \\ \mathbf{0}_p & \gamma^2\boldsymbol{\Omega} \end{pmatrix},$$

provando que x_i e $\boldsymbol{\epsilon}_i$ são não correlacionados, que é uma suposição natural e clássica

assumida no modelo de regressão com erros nas variáveis. Considerando a distribuição marginal de x_i e ϵ_i , primeiramente observe que

$$x_i|U_i = u_i \sim N(\mu + u_i\lambda, u_i), \quad \epsilon_i|U_i = u_i \sim N_p(\mathbf{0}_p, u_i\Omega), \quad U_i \sim \text{IG}(\gamma, \delta),$$

que é dada por 6.4, implicando que $x_i \sim \text{NIG}(\mu, 1, \lambda, \gamma, \delta)$ e $\epsilon_i \sim \text{NIG}(\mathbf{0}_p, \Omega, \mathbf{0}_p, \gamma, \delta)$, com o uso da Definição 7. Por (6.3), temos que $E(\epsilon_i) = \mathbf{0}_p$ e $\text{Var}(\epsilon_i) = (\delta/\gamma)\Omega$.

Como anteriormente, para cada indivíduo i , $i = 1, \dots, n$, vamos considerar o vetor de observações $\mathbf{Z}_i = (X_i, \mathbf{Y}'_i)'$ e os vetores $\mathbf{a} = (0, \boldsymbol{\alpha}')'$ e $\mathbf{b} = (1, \boldsymbol{\beta}')'$ e reescrever a forma funcional do modelo como $\mathbf{Z}_i = \mathbf{a} + \mathbf{b}x_i + \epsilon_i$. Assim, podemos escrever a seguinte representação estocástica

$$\mathbf{Z}_i|(x_i, U_i = u_i) \sim N_p(\mathbf{a} + \mathbf{b}x_i, u_i\Omega), \quad x_i|U_i = u_i \sim N(\mu + u_i\lambda, u_i), \quad U_i \sim \text{IG}(\gamma, \delta).$$

Dado que $U_i = u_i$, x_i e ϵ_i são independentes, veja (6.4), temos que $\mathbf{Z}_i|U_i = u_i$ é um vetor com dimensão p tendo distribuição normal,

$$\mathbf{Z}_i|U_i = u_i \sim N_p[\mathbf{a} + \mathbf{b}\mu + u_i\mathbf{b}\lambda, u_i(\mathbf{b}\mathbf{b}' + \Omega)].$$

Observe que $u_i\mathbf{b}\lambda = u_i(\mathbf{b}'\mathbf{b} + \Omega)(\mathbf{b}\mathbf{b}' + \Omega)^{-1}\mathbf{b}\lambda$. Então, temos que a distribuição marginal de \mathbf{Z}_i é

$$\mathbf{Z}_i \sim \text{NIG}(\mathbf{a} + \mathbf{b}\mu, \mathbf{b}\mathbf{b}' + \Omega, (\mathbf{b}\mathbf{b}' + \Omega)^{-1}\mathbf{b}\lambda, \gamma, \delta), \quad i = 1, \dots, n.$$

Podemos obter um modelo mais flexível quanto a multimodalidade introduzindo a cada $i = 1, \dots, n$, a variável classificadora w_i , de tal forma que $P(w_i = j) = p_j$, $\sum_{j=1}^G p_j = 1$ e $p_j \geq 0$, $j = 1, \dots, G$. Assim obtemos a seguinte representação

$$\begin{aligned} \mathbf{Z}_i|(x_i, U_i = u_i) &\sim N_p(\mathbf{a} + \mathbf{b}x_i, u_i\Omega), \\ x_i|(U_i = u_i, w_i = j) &\sim N(\mu_j + u_i\lambda_j, u_i), \\ U_i|w_i = j &\sim \text{IG}(\gamma_j, \delta_j), \\ P(w_i = j) &= p_j, \quad j = 1, \dots, G, \quad i = 1, \dots, n. \end{aligned} \tag{6.5}$$

Nesse caso, a distribuição marginal de x_i é uma mistura finita onde a j -ésima densidade na mistura é uma $\text{NIG}(\mu_j, 1, \lambda_j, \gamma_j, \delta_j)$, $j = 1, \dots, G$. A representação acima é também útil para gerarmos observações deste modelo, que nós chamamos de FMNIG-ME.

Para testar a flexibilidade da nossa metodologia vamos gerar uma amostra de tamanho

$n = 200$ do modelo FMNIG-ME com os seguintes parâmetros

$$\alpha = (0.4, 0.1), \quad \beta = (0.8, 0.9), \quad \mu = (-3, 1), \quad \delta = (0.7, 0.7), \quad (6.6)$$

$$\gamma = (1, 1), \quad \Omega = \mathbf{I}_3, \quad \mathbf{p} = (0.4, 0.6) \quad \text{e} \quad \lambda = (-2, 1).$$

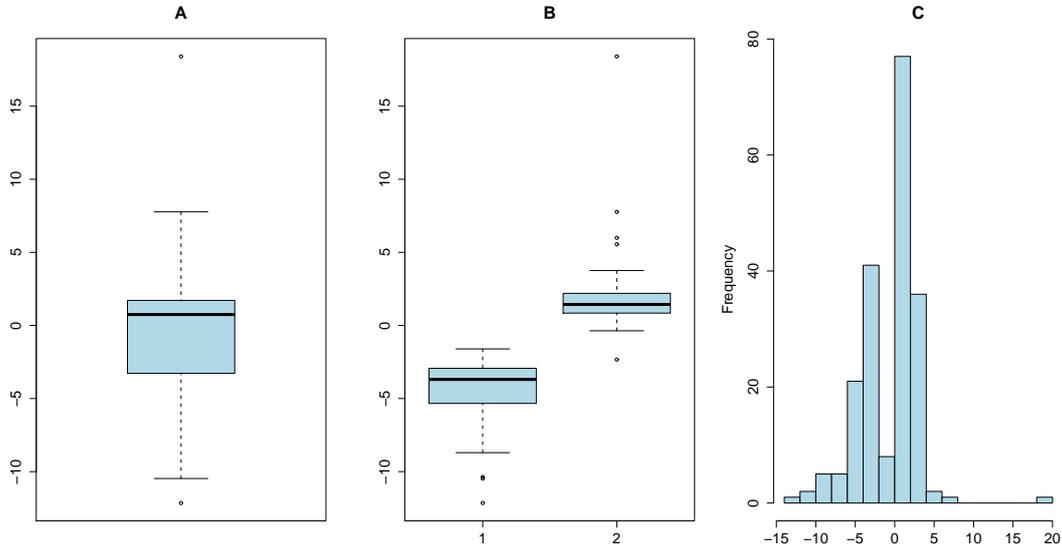


Figura 6.5: Variável X do Modelo FMNIG-ME, (A) boxplot de X , (B) boxplot de X condicionada a variável classificadora w , (C) histograma de X .

Nesse exemplo usaremos os modelos FMST-ME e FMSSL-ME para modelar os dados, onde geraremos uma amostra MCMC de tamanho 25 mil excluindo as primeiras 5 mil e usando $lag=5$, tendo ao final uma amostra de tamanho 4 mil. Usaremos a média amostral para estimar os parâmetros. Nas Figuras 6.6 (modelo FMST-ME) e 6.7 (modelo FMST-ME) vemos que todos os parâmetros apresentados tem densidade simétrica e que percorrem de maneira homogênea os seus suportes, evidenciando assim uma amostra MCMC com pouca autocorrelação, o que é bom no ponto de vista da estimação.

Além dos modelos FMST-ME e FMSSL-ME ajustaremos também os modelos simétricos FMN-ME e FMT-ME e o modelo assimétrico porém sem caudas pesadas FMSN-ME para estudarmos o quão flexíveis são os modelos da família SNI.

Na Tabela 6.4 temos as estimativas dos modelos em estudo. É interessante observarmos como os modelos FMSN-ME e FMN-ME tentam compensar a ausência de peso nas caudas superestimando os parâmetros de escala, mostrando assim a grande necessidade de usar um modelo mais flexível. Além disso podemos ver que os modelos FMST-ME e FMSSL-ME apresentaram pouca variabilidade nas estimativas MCMC com relação aos outros modelos, nos dando assim estimativas mais precisas e intervalos de credibilidade menores como podemos ver na Tabela 6.5.

Na Figura 6.8 podemos ver que os modelos FMST-ME e FMSSL-ME ajustaram-se

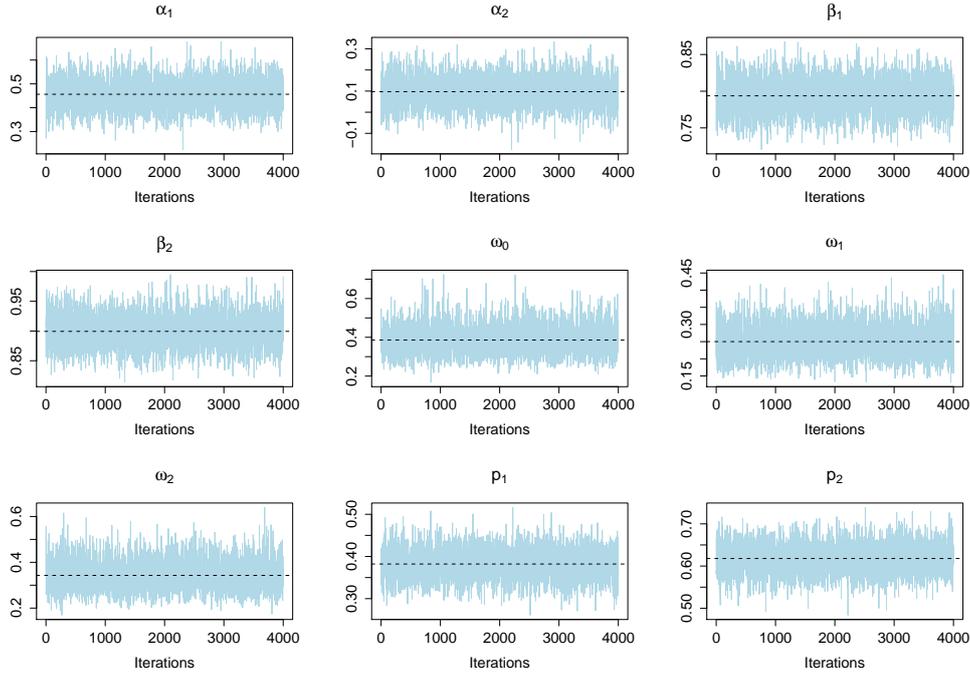


Figura 6.6: Traceplots de alguns parâmetros do modelo FMST-ME, para dados gerados do modelo FMNIG-ME. A linha pontilhada corresponde a estimativa do parâmetro.

Tabela 6.4: Estimativas dos parâmetros para vários modelos com erros nas variáveis para os dados simulados FMNIG-ME. A sigla dp denota o desvio padrão.

Parâmetro	FMN-ME		FMT-ME		FMSN-ME		FMST-ME		FMSSL-ME	
	Estimativa	dp								
α_1	0.43566	0.06999	0.46044	0.05621	0.44083	0.07452	0.45644	0.05957	0.45403	0.05966
α_2	0.15258	0.08764	0.08891	0.06627	0.15806	0.09206	0.09729	0.06740	0.08444	0.06650
β_1	0.73856	0.02111	0.79720	0.02253	0.73938	0.02092	0.79371	0.02260	0.79314	0.02282
β_2	0.89050	0.02578	0.89421	0.02601	0.89173	0.02657	0.89940	0.02674	0.90111	0.02691
ω_0	0.75552	0.13897	0.36135	0.06962	0.76207	0.15486	0.38590	0.07004	0.22204	0.04683
ω_1	0.59784	0.08629	0.21112	0.04500	0.61394	0.09300	0.25004	0.04595	0.14310	0.02886
ω_2	0.93310	0.13018	0.33316	0.06347	0.96950	0.14196	0.34273	0.06337	0.19907	0.04656
μ_1	-2.25382	0.44017	1.39165	0.09590	3.25906	0.22141	-2.63524	0.14426	-2.59830	0.14761
μ_2	1.44849	0.11280	-3.6037	0.16014	-5.85648	6.21951	0.58646	0.12195	0.58595	0.12569
Δ_1	-	-	-	-	-5.03951	0.31863	-1.38087	0.18684	-1.07943	0.17950
Δ_2	-	-	-	-	10.46637	5.71412	1.02692	0.15626	0.76609	0.13027
γ_1^2	15.48714	2.19191	0.46471	0.11073	0.56223	0.40796	0.02899	0.04653	0.01645	0.02615
γ_2^2	0.26252	0.11825	1.05310	0.34693	0.85096	1.70170	0.03971	0.05755	0.02522	0.03524
ν	-	-	2.46853	0.37910	-	-	2.82439	0.41520	1.02979	0.15890
p_1	0.56024	0.04709	0.60981	0.03628	0.93775	0.04378	0.38210	0.03467	0.38244	0.03421
p_2	0.43975	0.04709	0.39018	0.03628	0.06224	0.04378	0.61789	0.03467	0.61755	0.03421

melhor ao comportamento do modelo FMNIG-ME (a densidade de X é estimada pelo histograma) conseguindo resgatar tanto seu comportamento assimétrico e multimodal, como também suas caudas pesadas. Mostrando assim a flexibilidade da nossa metodologia.

Para finalizar, temos a comparação dos modelos, que podemos visualizar na Tabela 6.6. A mesma mostra que a melhor modelagem (menor DIC_{obs}) foi feita através do modelo FMST-ME seguido de muito perto pelo modelo FMSSL-ME, enquanto os outros tiveram um ajuste insatisfatório.

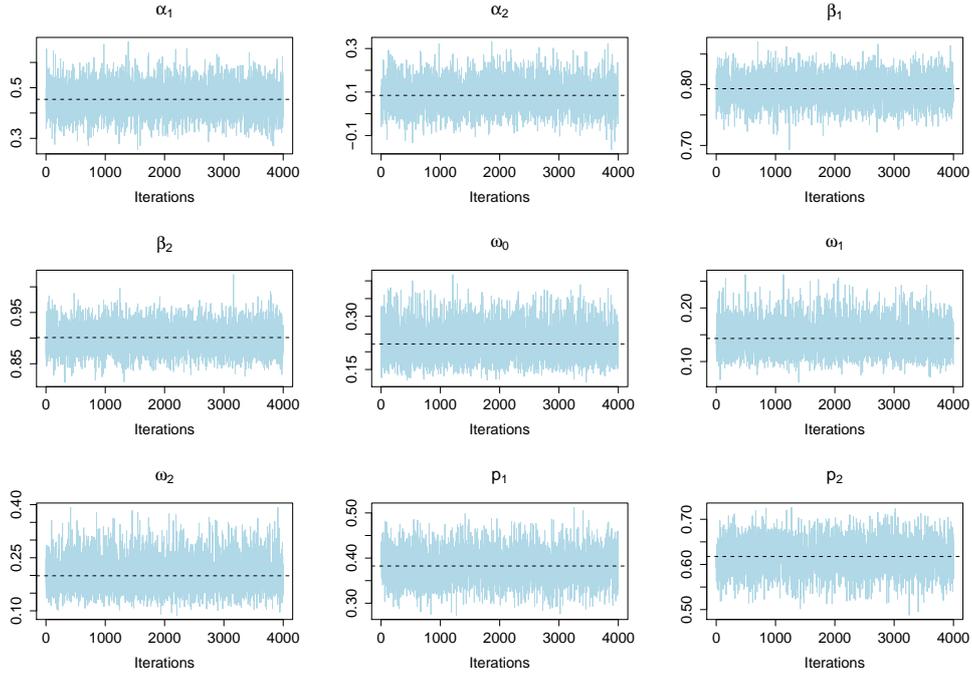


Figura 6.7: Traceplots de alguns parâmetros do modelo FMSSL-ME, para dados gerados do modelo FMNIG-ME. A linha pontilhada corresponde a estimativa do parâmetro.

Tabela 6.5: Intervalos de credibilidade para estimação dos modelos FMST-ME e FMSSL-ME para os dados simulados FMNIG-ME, 95% de credibilidade.

parâmetro	FMST-ME		FMSSL-ME	
	limite inferior	limite superior	limite inferior	limite superior
α_1	0.3446	0.5778	0.3362	0.5690
α_2	-0.0357	0.2282	-0.0475	0.2143
β_1	0.7486	0.8368	0.7496	0.8378
β_2	0.8456	0.9500	0.8501	0.9548
ω_0	0.2503	0.5215	0.1435	0.3222
ω_1	0.1680	0.3433	0.0906	0.1998
ω_2	0.2200	0.4649	0.1208	0.2978
μ_1	-2.9240	-2.382	-2.854	-2.269
μ_2	0.3617	0.8438	0.3385	0.8386
Δ_1	-1.7320	-1.040	-1.440	-0.769
Δ_2	0.7322	1.3330	0.5249	1.0260
γ_1^2	0.00001	0.1218	0.0000	0.0667
γ_2^2	0.00001	0.1649	0.00002	0.0989
ν	2.0600	3.6710	0.7647	1.3770
p_1	0.3143	0.4487	0.3124	0.4480
p_2	0.5513	0.6857	0.5520	0.6876

Tabela 6.6: Critério de seleção para vários modelos com erros nas variáveis para os dados simulados FMNIG-ME.

	FMN-ME	FMT-ME	FMSN-ME	FMST-ME	FMSSL-ME
$\log\text{-verossimilhança}$	-1178.298	-1023.024	-1110.772	-995.3303	-998.3397
τ_D	19.39295	22.34081	17.52553	13.15007	14.60683
DIC_{obs}	2375.989	2068.39	2239.07	2003.811	2011.286

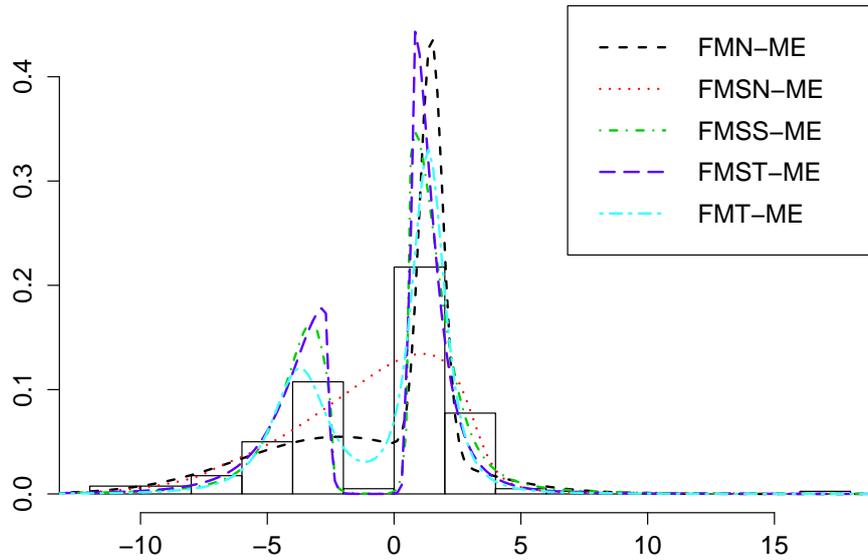


Figura 6.8: Preditiva da variável x nos modelos FMST-ME, FMSS-ME, FMN-ME, FMT-ME e FMSN-ME para dados gerados do modelo FMNIG-ME.

6.2 Aplicação em dados reais

Para ilustrar o método desenvolvido usaremos um conjunto de dados proveniente do trabalho de Chipkevitch *et al.* [13], que pode ser encontrado no apêndice A. O mesmo é constituído por medições do volume testicular de 42 adolescentes, usando cinco diferentes técnicas: ultra-som(US), método gráfico proposto pelo autores (I), a medição dimensional (II), orquidômetro de Prader (III) e orquidômetro anel (IV). Para a nossa abordagem desconsideraremos os instrumentos III e IV e trabalharemos apenas com os demais. A abordagem do ultra-som é assumida como instrumento de medição de referência (X). Galea-Rojas *et al.* [16] analisaram os mesmos dados através de um modelo com erros nas variáveis assumindo normalidade e recomendou que uma transformação de raiz cúbica fosse feita para uma melhor aproximação ao modelo normal.

Em Lachos *et al.* [21] também podemos ver a análise desse conjunto de dados onde são mostradas evidências fortes de que a distribuição da variável latente apresenta um comportamento assimétrico e de cauda pesada. Porém diferente da abordagem de Galea-Rojas *et al.* [16] foram evitadas técnicas de transformação desnecessárias nos dados, usando uma modelagem onde os erros de observação têm distribuições pertencentes à família SNI. Ao final observou-se realmente um melhor ajuste, e não houve mais dúvidas sobre a superioridade dos modelos que levam em consideração as caudas pesadas e a assimetria.

Assim como nos exemplos computacionais usaremos os modelos FMN-ME, FMT-ME, FMSN-ME, FMST-ME e FMSS-ME com uma amostra de 25 mil, com *burning*

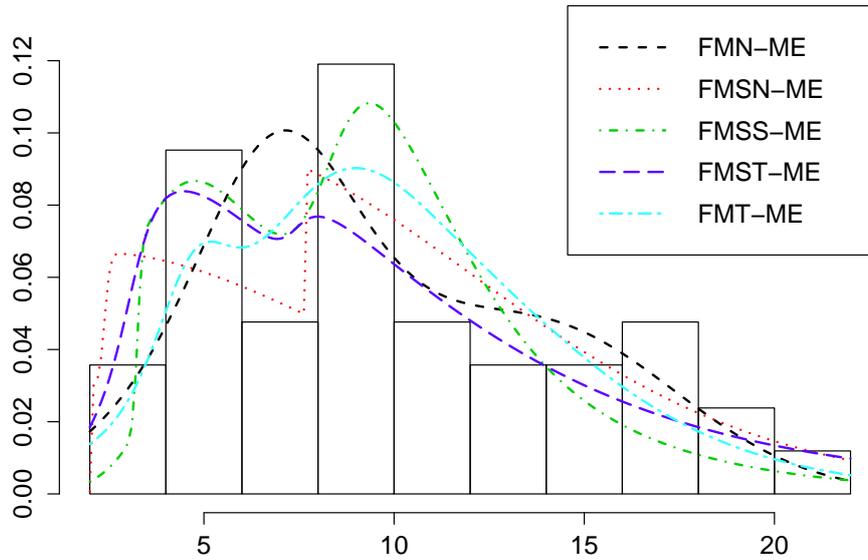


Figura 6.9: Preditiva da variável x nos modelos FMST-ME, FMSS-ME, FMN-ME, FMT-ME e FMSN-ME para o conjunto de dados Chipkevitch (1996).

de 5 mil e *lag* de 5, na Figura 6.9 podemos ver que o histograma nos dá evidências de multimodalidade. Logo parece ser razoável usar um modelo de misturas. Ainda nessa figura podemos ver que as preditivas dos modelos FMST-ME e FMSS-ME evidenciam uma bimodalidade e capturam bem a cauda pesada.

Nessa parte de aplicação em dados reais omitiremos os traceplots. Porém, informamos que sua trajetória foi homogênea e sem *label switching*, assim como nas aplicações com dados artificiais.

Tabela 6.7: Estimativas dos parâmetros para vários modelos com erros nas variáveis para o conjunto de dados Chipkevitch (1996). A sigla dp denota o desvio padrão da amostra MCMC.

Parâmetro	FMN-ME		FMT-ME		FMSN-ME		FMST-ME		FMSS-ME	
	Estimativa	dp	Estimativa	dp	Estimativa	dp	Estimativa	dp	Estimativa	dp
α_1	0.16966	0.64669	0.14493	0.58867	0.25219	0.65137	0.11869	0.51987	0.22550	0.55297
α_2	-0.04083	0.68086	-0.07304	0.59442	0.08638	0.67438	-0.11306	0.56150	-0.07413	0.51870
β_1	0.87743	0.05869	0.87991	0.06084	0.87229	0.05765	0.88557	0.05703	0.87214	0.05609
β_2	0.95303	0.06293	0.95547	0.06315	0.94376	0.05972	0.96478	0.06470	0.95461	0.05419
ω_0	1.13949	0.70256	0.91992	0.49753	1.22246	0.68128	0.88530	0.46403	0.59722	0.32216
ω_1	1.80553	0.63461	1.19360	0.50713	1.73304	0.60673	1.04512	0.42574	0.85895	0.45889
ω_2	1.86214	0.73040	0.87417	0.50897	1.82416	0.66189	0.84806	0.42517	0.57682	0.35621
μ_1	11.12647	4.54372	9.15834	4.52424	5.03266	4.39717	4.64045	3.63728	3.27440	3.96584
μ_2	9.72217	3.97900	8.93623	4.13170	5.17779	5.01398	5.29041	4.45461	5.49006	4.91018
Δ_1	-	-	-	-	6.15640	4.67817	4.65673	4.47016	3.98622	4.28389
Δ_2	-	-	-	-	4.11775	5.98164	4.97795	4.37522	4.24107	4.69861
γ_1^2	12.26406	8.08442	12.88038	23.99096	1.08387	2.37389	2.41739	40.1406	1.24563	36.5881
γ_2^2	11.31506	7.31614	13.31314	38.73909	0.99836	2.16186	1.73710	13.0293	0.78426	3.96993
ν	-	-	6.08686	4.04767	-	-	5.07269	3.44172	2.14340	2.32083
p_1	0.46964	0.24249	0.47045	0.30172	0.54189	0.30845	0.46334	0.32590	0.56846	0.29650
p_2	0.53035	0.24249	0.52954	0.30172	0.45810	0.30845	0.53665	0.32590	0.43153	0.29650

Para o banco de dados de Chipkevitch (1996) ainda podemos observar nas Tabelas 6.9 e 6.8 as estimativas para os parâmetros e os intervalos de credibilidade respectivamente. Usando o critério de classificação de modelos, dispostos na tabela 6.9, concluímos que o melhor ajuste foi o do modelo FMSSL-ME que obteve o menor DIC_{obs} .

Tabela 6.8: Intervalos de credibilidade para estimação dos modelos FMST-ME e FMSSL-ME para os dados de Chipkevitch (1996), 95% de credibilidade.

parâmetro	FMST-ME		FMSSL-ME	
	limite inferior	limite superior	limite inferior	limite superior
α_1	-0.8527	1.1210	-0.9314	1.2460
α_2	-1.153	0.9929	-1.006	0.9995
β_1	0.7830	1.0050	0.7586	0.9823
β_2	0.8440	1.0920	0.8509	1.0630
ω_0	0.0999	1.7460	0.1208	1.2820
ω_1	0.3156	1.8920	0.2212	1.8200
ω_2	0.0114	1.5650	0.0410	1.1840
μ_1	-0.8518	13.000	-5.000	13.000
μ_2	0.6702	19.000	-6.000	16.000
Δ_1	-4.963	11.360	-4.788	13.450
Δ_2	-7.930	11.380	-9.044	11.330
γ_1^2	0.0000	6.9920	0.0000	3.1860
γ_2^2	0.0000	6.9920	0.0000	3.6280
ν	1.6710	9.8620	0.6070	5.9550
p_1	0.0025	0.9689	0.0141	0.9999
p_2	0.0311	0.9975	0.0001	0.9859

Tabela 6.9: Critério de seleção para vários modelos com erros nas variáveis para os dados de Chipkevitch (1996).

	FMN-ME	FMT-ME	FMSN-ME	FMST-ME	FMSSL-ME
<i>log-verossimilhança</i>	-304.4154	-304.3739	-282.3515	-277.65	-277.1748
τ_D	13.17309	16.83206	11.32655	11.30353	10.49962
DIC_{obs}	622.0038	625.5798	576.0296	566.6035	564.8493

Capítulo 7

Conclusão

Levando em consideração os aspectos e a estrutura dos modelos com erros nas variáveis já existentes, fica claro a necessidade de usar uma estrutura que acomode certas características que até então não eram levadas em consideração. Neste trabalho construímos um modelo mais amplo com relação a estrutura dos dados. O principal ganho foi o acréscimo da estrutura de mistura, dando assim a possibilidade de modelagem de dados heterogêneos e com multimodalidade, o que não existia até então. Juntando isso ao uso da classe bem ampla de distribuições da família SNI, conseguimos um modelo altamente flexível que consegue se adequar às mais diferentes estruturas, como vimos na aplicação com dados do modelo FMNIG-ME. Além disso, estimando bem o comportamento da sua própria estrutura, como vimos na aplicação com o modelo FMST-ME.

A fim de mostrar a relevância da nossa metodologia foram comparados vários modelos tradicionais e com formas diferentes, e em todos notamos a superioridade dos modelos com misturas de ST e SSL. Eles se mostraram bem mais flexíveis com bons ajustes e um DIC bem menor, mostrando assim a necessidade de usar um modelo que se incorpore bem a estrutura dos dados e que esteja livre de aproximações ou transformações.

Na modelagem com os dados reais Chipkevitch (1996), analisando os gráficos apresentados, observamos visualmente que a variável substituta tinha uma característica multimodal e de caudas pesadas, o que nos incentivou o uso de um modelo que captasse essa estrutura. Em contrapartida, existia a possibilidade de um ajuste com modelos mais simples, deixando para trás um ajuste com mais parâmetros se o simples fosse satisfatório. Porém, diante do exposto nas aplicações vimos que, apesar de mais simples, os modelos FMN-ME, FMT-ME e FMSN-ME não conseguiram ajustar adequadamente os dados sendo que, em diversas vezes superestimavam parâmetros de escala a fim de resgatar a assimetria e as caudas pesadas e em outras vezes identificavam apenas uma moda na distribuição, entre outras coisas. Quando partimos para os modelos FMST-ME e FMSS-ME ficou claro o seu melhor ajuste, evidenciando um o comportamento bimodal e das caudas pesadas.

Capítulo 8

Apêndice A

8.1 Programas em R para geração da estrutura de dados

8.1.1 Modelo FMST-ME

```
# Função para gerar observações do modelo EIV-FMST
# Neste modelo o efeito aleatório é distribuído como uma mistura de G
# componentes skew-t
#-----
# Início da função
#-----
reiv.STmix <- function(n,alpha,beta,0mega,mu,gam,Delta,nu,peso){
# Função para gerar observações do modelo EIV-FMST (mas com erros de
# observação normais)
# Veja a definição do modelo no artigo "Bayesian Flexible Modelling
# of Random Effects in Linear Mixed Models
# Using Skew-Normal Distribution" de Cabral, Lachos & Madruga
#=====
# Input
# n: tamanho amostral
# X,Z: listas com as matrizes de planejamento. X[[i]] e Z[[i]] são as
# matrizes para os efeitos fixo e aleatório, respectivamente,
# referentes ao indivíduo i
# beta: vetor com os efeitos fixos
# sigma2: variância dos erros de observação
#-----
# Parâmetros para a modelagem dos efeitos aleatórios por misturas
#-----
# mu: vetor G x 1. A coordenada i contém o parâmetro de locação da classe i
```

```

# v: array de 3 dimensões tal que, para i fixo, om[, ,i] é a matriz de escala
# para a classe i (no caso em que m=1 deve ser um vetor)
# Delta: vetor G x 1. A coordenada i contém o parâmetro de assimetria da classe i
# gam: vetor G x 1. A coordenada i contém o parâmetro de dispersão da classe i
# peso: o vetor de pesos na mistura
# beta: vetor com os efeitos fixos
# nu: graus de liberdade
#=====
library(mnormt)
library(msm) # para distribuições truncadas
p <- length(alpha) + 1
G <- length(peso)
#-----
w <- matrix(0,n,1) # 0 elemento i desta lista é a observação
# associada ao indivíduo i
x <- matrix(0,n,1) # variável latente
Z <- matrix(0,n,p) # variável latente
#=====
a <- matrix(c(0,alpha))
b <- matrix(c(1,beta))
w <- matrix(sample(G,size=n,replace=TRUE,prob=peso)) # vetor de alocações
for (i in 1:n){
s <- rgamma(1,shape=nu/2,rate=nu/2)
t <- rtnorm(1,0,1/sqrt(s), lower=0, upper=Inf)
x[i] <- matrix(rnorm(1,mu[w[i]]+Delta[w[i]]*t,sqrt((1/s)*gam[w[i]])))
Z[i,] <- matrix(rmnorm(1,a+b*x[i],(1/s)*Omega))
}
X <- Z[,1]
Y <- Z[,-1]
return(list(X=X,Y=Y,x=x,w=w))
}

```

8.1.2 Modelo NIGM-ME

```

# Função para gerar observações do modelo de regressão linear com
# erros nas variáveis
# onde a variável latente é uma mistura de normais inversas Gaussianas
#=====
# Input

```

```

#=====
# n: tamanho amostral
# alpha, beta: vetores de regressão de dimensão r;
# Omega: matriz p x p associada à escala dos erros de observação, onde p=r+1;
# mu: vetor de parâmetros associados à locação da variável latente
# para cada sub-população. dim(mu)=G, onde
# G é o número de sub-populações e mu[i] é o parâmetro de locação
# associado à sub-população i;
# lambda: analogamente, é um vetor de dimensão G contendo os parâmetros
# reguladores de assimetria para cada sub-população;
# gama, delta: constantes positivas, parâmetros do fator de escala -
# que tem distribuição Gaussiana inversa
# pii: vetor de dimensão G com os pesos na mistura
#-----
# Início da função
#-----
reiv.NIGmix <- function(n,alpha,beta,Omega,mu,lambda,gama,delta,pii){
library(mnormt)
library(SuppDists)
p <- length(alpha) + 1
G <- length(pii)      # número de componentes na mistura
#-----
x <- matrix(0,n,1)      # vetor que conterà os valores gerados da
                        # variável latente
Z <- matrix(0,n,p)      # vetor que conterà as observações geradas
#=====
a <- matrix(c(0,alpha))
B1 <- matrix(c(1,beta))
u <- vector()
w <- matrix(sample(G,size=n,replace=TRUE,prob=pii))
# w é o vetor de alocações, se z[i]=j então o indivíduo i é alocado à classe j
nu <- delta/gama
psi <- delta^2
for (i in 1:n){
u[i] <- rinvGauss(1,nu,psi)
x[i] <- rnorm(1,mu[w[i]]+u[i]*lambda[w[i]],sqrt(u[i]))
Z[i,] <- matrix(rmnorm(1,a+B1*x[i],u[i]*Omega))
}
return(list(Z=Z,w=w,x=x))

```

```
}
```

8.2 Programas em WinBugs para estimação dos modelos

8.2.1 Modelo FMN-ME

```
#####  
### MODELO DE REGRESSÃO COM ERRO PARA MISTURAS DE N ###  
#####  
  
### GERANDO VEROSSIMILHANÇA ###  
  
model  
{  
  
for(i in 1:n){  
  w[i] ~ dcat(peso[] )  
t[i] ~ dnorm(0, 1) I(0,)  
  x[i] ~ dnorm(mean.x[i],prec.x[i])  
  mean.x[i] <- mu[w[i]]  
  prec.x[i] <- igamma2[w[i]]  
  
for(j in 1:p){  
mean.z[i,j] <- a[j]+b[j]*x[i]  
prec.z[i,j] <- iOmega[j]  
z[i,j] ~ dnorm(mean.z[i, j], prec.z[i,j])  
}  
}  
  
### GERANDO DENSIDADES A PRIORI ###  
  
b[1] <- 1  
a[1] <- 0  
peso[1:G] ~ ddirch(theta[])  
  
for(k in 2:p){  
b[k] ~ dnorm(0.0,0.01) #beta  
a[k] ~ dnorm(0.0,0.01) #alpha
```

```

    }

for(l in 1:p){
iOmega[l] ~ dgamma(0.01, 0.01)
Omega[l] <- 1/iOmega[l]
}

for (j in 1:G){
gamma2[j] <- 1/igamma2[j]
            igamma2[j] ~ dgamma(3,f)
mu[j] ~ dnorm(0, 0.01)
}

f ~ dgamma(0.1,0.1)
}

```

8.2.2 Modelo FMSN-ME

```

#####
### MODELO DE REGRESSÃO COM ERRO PARA MISTURAS DE SN ###
#####

### GERANDO VEROSSIMILHANÇA ###

model
{

for(i in 1:n){
    w[i] ~ dcat(peso[] )
t[i] ~ dnorm(0, 1) I(0,)
    x[i] ~ dnorm(mean.x[i],prec.x[i])
    mean.x[i] <- mu[w[i]]+Delta[w[i]]*t[i]
    prec.x[i] <- igamma2[w[i]]

for(j in 1:p){
mean.z[i,j] <- a[j]+b[j]*x[i]
prec.z[i,j] <- iOmega[j]
z[i,j] ~ dnorm(mean.z[i, j], prec.z[i,j])
}
}
}

```

```

}
}

### GERANDO DENSIDADES A PRIORI ###

b[1] <- 1
a[1] <- 0
peso[1:G] ~ ddirch(theta[])

for(k in 2:p){
b[k] ~ dnorm(0.0,0.01) #beta
a[k] ~ dnorm(0.0,0.01) #alpha
}

for(l in 1:p){
iOmega[l] ~ dgamma(0.01, 0.01)
Omega[l] <- 1/iOmega[l]
}

for (j in 1:G){
gamma2[j] <- 1/igamma2[j]
igamma2[j] ~ dgamma(3,f)
mu[j] ~ dnorm(0, 0.01)
Delta[j] ~ dnorm(0,0.01)
}

f ~ dgamma(0.1,0.1)

}

```

8.2.3 Modelo FMSSL-ME

```

#####
### MODELO DE REGRESSÃO COM ERRO PARA MISTURAS DE SS ###
#####

### GERANDO VEROSSIMILHANÇA ###

```

```

model
{

for(i in 1:n){
  w[i] ~ dcat(peso[] )
t[i] ~ dnorm(0, 1) I(0,)
  x[i] ~ dnorm(mean.x[i],prec.x[i])
  mean.x[i] <- mu[w[i]]+Delta[w[i]]*(t[i]/sqrt(u[i]))
  prec.x[i] <- u[i]*igamma2[w[i]]
  u[i] ~ dbeta(nu,1)      # skew-slash

for(j in 1:p){
mean.z[i,j] <- a[j]+b[j]*x[i]
prec.z[i,j] <- iOmega[j]*u[i]
z[i,j] ~ dnorm(mean.z[i, j], prec.z[i,j])
}
}

### GERANDO DENSIDADES A PRIORI ###

b[1] <- 1
a[1] <- 0
peso[1:G] ~ ddirch(theta[])

for(k in 2:p){
b[k] ~ dnorm(0.0,0.01) #beta
a[k] ~ dnorm(0.0,0.01) #alpha
}

for(l in 1:p){
iOmega[l] ~ dgamma(0.01, 0.01)
Omega[l] <- 1/iOmega[l]
}

for (j in 1:G){
gamma2[j] <- 1/igamma2[j]
  igamma2[j] ~ dgamma(1,f)
mu[j] ~ dnorm(0, 0.01)
Delta[j] ~ dnorm(0,0.01)
}

```

```

}

nu ~ dgamma(0.0001,0.0001)
f ~ dgamma(0.1,0.1)
}

```

8.2.4 Modelo FMST-ME

```

#####
### MODELO DE REGRESSÃO COM ERRO PARA MISTURAS DE ST ###
#####

### GERANDO VEROSSIMILHANÇA ###

model
{

for(i in 1:n){
  w[i] ~ dcat(peso[] )
  t[i] ~ dnorm(0, 1) I(0,)
  x[i] ~ dnorm(mean.x[i],prec.x[i])
  mean.x[i] <- mu[w[i]]+Delta[w[i]]*(t[i]/sqrt(u[i]))
  prec.x[i] <- u[i]*igamma2[w[i]]
  u[i] ~ dgamma(nu1, nu1)      # t-Student

for(j in 1:p){
  mean.z[i,j] <- a[j]+b[j]*x[i]
  prec.z[i,j] <- iOmega[j]*u[i]
  z[i,j] ~ dnorm(mean.z[i, j], prec.z[i,j])
}
}

### GERANDO DENSIDADES A PRIORI ###

b[1] <- 1
a[1] <- 0
peso[1:G] ~ ddirch(theta[])

```

```

for(k in 2:p){
b[k] ~ dnorm(0.0,0.01) #beta
a[k] ~ dnorm(0.0,0.01) #alpha
}

for(l in 1:p){
iOmega[l] ~ dgamma(0.01, 0.01)
Omega[l] <- 1/iOmega[l]
}

for (j in 1:G){
gamma2[j] <- 1/igamma2[j]
igamma2[j] ~ dgamma(1,f)
mu[j] ~ dnorm(0, 0.01)
Delta[j] ~ dnorm(0,0.01)
}

nu ~ dexp(lambda) # t-Student
nu1 <- 0.5*nu # t-Student
f ~ dgamma(0.1,0.1)
lambda ~ dunif(11,mm)
}

```

8.2.5 Modelo FMT-ME

```

#####
### MODELO DE REGRESSÃO COM ERRO PARA MISTURAS DE T ###
#####

### GERANDO VEROSSIMILHANÇA ###

model
{

for(i in 1:n){
w[i] ~ dcat(peso[] )
x[i] ~ dnorm(mean.x[i],prec.x[i])
mean.x[i] <- mu[w[i]]
}
}

```

```

        prec.x[i] <- u[i]*igamma2[w[i]]
        u[i] ~ dgamma(nu1, nu1)          # t-Student

for(j in 1:p){
mean.z[i,j] <- a[j]+b[j]*x[i]
prec.z[i,j] <- iOmega[j]*u[i]
z[i,j] ~ dnorm(mean.z[i, j], prec.z[i,j])
}
}

### GERANDO DENSIDADES A PRIORI ###

b[1] <- 1
a[1] <- 0
peso[1:G] ~ ddirch(theta[])

for(k in 2:p){
b[k] ~ dnorm(0.0,0.01) #beta
a[k] ~ dnorm(0.0,0.01) #alpha
}

for(l in 1:p){
iOmega[l] ~ dgamma(0.01, 0.01)
Omega[l] <- 1/iOmega[l]
}

for (j in 1:G){
gamma2[j] <- 1/igamma2[j]
                igamma2[j] ~ dgamma(1,f)
mu[j] ~ dnorm(0, 0.01)
}

nu ~ dexp(lambda)          # t-Student
nu1 <- 0.5*nu             # t-Student
f ~ dgamma(0.1,0.1)
lambda ~ dunif(1l,mm)
}

```

8.3 Valores iniciais e hiperparâmetros

Os valores iniciais e dos hiperparâmetros foram os mesmo para todos os modelos sob todos os ajustes de conjunto de dados.

```
#dados
dados <- list(z=Z,n=200, G=2, ll=0.04, mm=0.5, p=3,theta=c(1,1))

#valores iniciais
ini <- list()

peso=c(0.6,0.4)
a=c(NA,0.6, 0.3)
b=c(NA,0.4,0.7)
iOmega=c(8,3,1)
mu=c(2,8)
Delta=c(0,4)
igamma2=c(8,6)
nu=4
w=sample(2,size=n,replace=TRUE,prob=peso)

#hiperpâmetros

C_b = C_a = rep(0,r)
D_a = D_b = diag(rep(0.01,r))
a_mu = c_delta = 0
b_mu = d_delta = m = l = 0.01
h = g = 0.1
kappa_1 = kappa_2 = e = 1
lambda_0 = 0.04
lambda_1 = 0.5
phi_sl = varphi_sl = 0.001
```

8.4 Dados Chipkevitch (1996)

```
X Y_1 Y_2
5.0000 7.5000 5.9000
5.7000 5.0000 4.8000
7.4000 5.0000 6.8000
```

2.6000 3.5000 3.1000
5.7000 5.0000 5.0000
6.1000 5.0000 4.4000
6.2000 5.0000 6.0000
10.4000 10.0000 8.8000
9.1000 7.5000 7.9000
14.8000 10.0000 13.0000
16.4000 12.5000 10.3000
9.6000 7.5000 8.2000
15.7000 15.0000 19.8000
3.0000 2.0000 2.0000
16.4000 15.0000 17.3000
17.6000 15.0000 17.3000
10.0000 7.5000 7.9000
4.1000 3.5000 4.4000
12.7000 10.0000 11.4000
2.7000 3.5000 4.1000
10.2000 10.0000 11.1000
16.5000 10.0000 15.3000
4.5000 3.5000 3.9000
5.6000 5.0000 4.5000
11.0000 7.5000 9.7000
9.2000 10.0000 11.3000
8.5000 7.5000 8.8000
5.4000 5.0000 6.1000
6.7000 7.5000 7.2000
5.3000 5.0000 5.9000
20.0000 20.0000 16.3000
18.8000 15.0000 16.3000
13.9000 12.5000 12.2000
9.4000 10.0000 10.3000
9.1000 7.5000 10.8000
14.1000 15.0000 13.0000
9.3000 10.0000 8.4000
20.9000 20.0000 22.1000
11.5000 10.0000 10.6000
9.7000 10.0000 9.7000
13.7000 12.5000 11.6000
8.9000 10.0000 8.1000

Capítulo 9

Apêndice B

9.1 Detalhes do algoritmo

No passo 3 temos a condicional completa da variável latente T_i , $i = 1, \dots, n$ vamos agora ver o seu desenvolvimento

$$\begin{aligned} \pi(T_i | \mathbf{Z}_i, \beta, \alpha, \Omega, \mu_j, \Delta_j, \gamma_j^2, x_i, U_i, w_i = j) &\propto \pi(\mathbf{Z}_i | T_i, \beta, \alpha, \Omega, \mu_j, \Delta_j, \gamma_j^2, x_i, U_i, w_i = j) \\ &\times \pi(x_i | \mu_j, \Delta_j, \gamma_j^2, T_i, U_i, w_i = j) \pi(T_i | U_i). \end{aligned}$$

Assim temos

$$\begin{aligned} &\propto \pi(x_i | \mu_j, \Delta_j, \gamma_j^2, T_i, U_i, w_i = j) \pi(T_i | U_i) \\ &\propto \exp \left\{ -\frac{u_i}{2\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 \right\} \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} I_{(0, \infty)}(T_i) \\ &\propto \exp \left\{ -\frac{1}{2} \left[\frac{u_i}{\gamma_j^2} (-2(x_i - \mu_j) \Delta_j t_i + \Delta_j^2 t_i^2) \right] \right\} \exp \left\{ -\frac{1}{2} u_i t_i^2 \right\} I_{(0, \infty)}(T_i) \\ &\propto \exp \left\{ -\frac{1}{2} \left[-2(x_i - \mu_j) \Delta_j u_i \gamma_j^{-2} t_i + t_i^2 \Delta_j^2 u_i \gamma_j^{-2} + u_i t_i^2 \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[-2(x_i - \mu_j) \Delta_j u_i \gamma_j^{-2} t_i + t_i^2 \left(\Delta_j^2 u_i \gamma_j^{-2} + u_i \right) \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2 \left(\Delta_j^2 u_i \gamma_j^{-2} + u_i \right)^{-1}} \left[-2 \left[\frac{(x_i - \mu_j) \Delta_j u_i \gamma_j^{-2}}{\left(\Delta_j^2 u_i \gamma_j^{-2} + u_i \right)} t_i + t_i^2 \right] \right] \right\}. \end{aligned}$$

Então temos que, $T_i | \mathbf{Z}_i, \beta, \alpha, \Omega, \mu_j, \Delta_j, \gamma_j^2, x_i, U_i, w_i = j \sim HN(\mu_{T_i}, \sigma_{T_i}^2)$ onde

$$\sigma_{T_i}^2 = \left(\Delta_j^2 u_i \gamma_j^{-2} + u_i \right)^{-1} \quad e \quad \mu_{T_i} = \sigma_{T_i}^2 \left[(x_i - \mu_j) \Delta_j u_i \gamma_j^{-2} \right].$$

No passo 4 a) temos a condicional completa da variável latente U_i , $i = 1, \dots, n$, vamos agora ver o seu desenvolvimento

$$\begin{aligned} \pi(U_i | \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, T_i, x_i, \mu_j, \Delta_j, \gamma_j^2, w_i = j) &\propto \pi(\mathbf{Z}_i | U_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Omega}, x_i, w_i = j) \\ &\times \pi(\mathbf{x}_i | T_i, \mu_j, \gamma_j^2, \Delta_j, w_i = j) \pi(T_i | U_i) \pi(U_i). \end{aligned}$$

Assim temos

$$\begin{aligned} &\propto \frac{1}{|\boldsymbol{\Omega} u_i^{-1}|^{1/2}} \exp \left\{ -\frac{u_i}{2} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) \right\} \\ &\times \left(\frac{1}{(\gamma_j^2 u_i^{-1})^{1/2}} \right) \exp \left\{ -\frac{u_i}{2\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 \right\} \frac{2}{u_i^{-1/2}} \exp \left\{ -\frac{u_i}{2} t_i^2 \right\} u_i^{(v/2)-1} \exp \left\{ -\frac{v}{2} u_i \right\} \\ &\propto u_i^{r/2} \exp \left\{ -u_i \left[\frac{1}{2} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) \right] \right\} u_i^{1/2} \exp \left\{ -u_i \left[\frac{1}{2\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 \right] \right\} \\ &\times u_i^{1/2} \exp \left\{ -u_i \left[\frac{1}{2} t_i^2 \right] \right\} u_i^{(v/2)-1} \exp \left\{ -u_i \frac{v}{2} \right\} \\ &\propto u_i^{\frac{r}{2} + 1 + \frac{v}{2} - 1} \exp \left\{ -u_i \left[\frac{1}{2} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) + \frac{1}{2\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 + \frac{1}{2} t_i^2 + \frac{v}{2} \right] \right\}. \end{aligned}$$

Então temos que, $U_i | \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, T_i, x_i, \mu_j, \Delta_j, \gamma_j^2, w_i = j \sim \text{Gamma}(\text{shape}_{u_i}, \text{rate}_{u_i})$

onde

$$\text{shape}_{u_i} = \frac{r+2+v}{2} \quad e \quad \text{rate}_{u_i} = \frac{1}{2} \left[(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) + \frac{1}{\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 + t_i^2 + v \right].$$

Para o passo 4 b) a ideia é a mesma mudando apenas a densidade $\pi(U_i)$ para uma $\text{Beta}(v, 1)$. No passo 5 temos a condicional completa da variável latente x_i , $i = 1, \dots, n$, vamos agora ver o seu desenvolvimento usando a representação estocástica dada no Teo-

rema 5.

$$\begin{aligned}
\pi(x_i|\mathbf{Z}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Omega}, U_i, T_i, \mu_j, \Delta_j, \gamma_j^2) &\propto \pi(\mathbf{Z}_i|x_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Omega}, U_i, T_i, \mu_j, \Delta_j, \gamma_j^2)\pi(x_i|T_i, U_i, \mu_j, \Delta_j, \gamma_j^2) \\
&\propto \exp\left[-\frac{1}{2}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)'(u_i^{-1}\boldsymbol{\Omega})^{-1}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)\right] \exp\left\{-\frac{1}{2\gamma_j^2 u_i^{-1}}(x_i - \mu_j - \Delta_j t_i)^2\right\} \\
&\propto \exp\left[-\frac{u_i}{2}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)'\boldsymbol{\Omega}^{-1}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)\right] \exp\left\{-\frac{u_i}{2\gamma_j^2}(x_i^2 - 2x_i(\mu_j + \Delta_j t_i))\right\} \\
&\propto \exp\left\{-\frac{u_i}{2}\left[-2(\mathbf{Z}_i - \mathbf{a})'\boldsymbol{\Omega}^{-1}\mathbf{b}x_i + x_i^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b} + \gamma_j^{-2}(x_i^2 - 2x_i(\mu_j + \Delta_j t_i))\right]\right\} \\
&\propto \exp\left\{-\frac{u_i}{2}\left[-2x_i\left((\mathbf{Z}_i - \mathbf{a})'\boldsymbol{\Omega}^{-1}\mathbf{b} + \gamma_j^{-2}(\mu_j + \Delta_j t_i)\right) + x_i^2\left(\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b} + \gamma_j^{-2}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[-2x_i\left((\mathbf{Z}_i - \mathbf{a})'\boldsymbol{\Omega}^{-1}\mathbf{b}u_i + u_i\gamma_j^{-2}(\mu_j + \Delta_j t_i)\right) + x_i^2\left(\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}u_i + u_i\gamma_j^{-2}\right)\right]\right\}.
\end{aligned}$$

Então x_i é normal com

$$\begin{aligned}
\text{var}(x_i) &= \left(\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}u_i + u_i\gamma_j^{-2}\right)^{-1} \quad e \\
E(x_i) &= \text{var}(x_i) \left[(\mathbf{Z}_i - \mathbf{a})'\boldsymbol{\Omega}^{-1}\mathbf{b}u_i + u_i\gamma_j^{-2}(\mu_j + \Delta_j t_i)\right]
\end{aligned}$$

ou simplesmente

$$x_i|\mathbf{Z}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Omega}, U_i, T_i, \mu_j, \Delta_j, \gamma_j^2, w_i = j \sim N\left(\frac{(\mathbf{Z}_i - \mathbf{a})'\boldsymbol{\Omega}^{-1}\mathbf{b} + \gamma_j^{-2}(\mu_j + \Delta_j t_i)}{\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b} + \gamma_j^{-2}}, \frac{1}{\left(\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}u_i + u_i\gamma_j^{-2}\right)}\right).$$

Agora vamos desenvolver os termos que estão em função de \mathbf{a} e \mathbf{b} , colocando-os em função de $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$.

O primeiro termo é,

$$\begin{aligned}
\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b} &= \begin{bmatrix} 1 & \boldsymbol{\beta}' \end{bmatrix} \begin{bmatrix} \frac{1}{\omega_0^2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_e^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\beta} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\omega_0^2} & \boldsymbol{\beta}'\boldsymbol{\Omega}_e^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\beta} \end{bmatrix} = \frac{1}{\omega_0^2} + \boldsymbol{\beta}'\boldsymbol{\Omega}_e^{-1}\boldsymbol{\beta}.
\end{aligned}$$

O segundo termo é,

$$\begin{aligned}
(\mathbf{Z}_i - \mathbf{a})'\boldsymbol{\Omega}^{-1}\mathbf{b} &= \begin{bmatrix} X_i & (\mathbf{Y}_i - \boldsymbol{\alpha})' \end{bmatrix} \begin{bmatrix} \frac{1}{\omega_0^2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_e^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\beta} \end{bmatrix} \\
&= \begin{bmatrix} \frac{X_i}{\omega_0^2} & (\mathbf{Y}_i - \boldsymbol{\alpha})'\boldsymbol{\Omega}_e^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\beta} \end{bmatrix} = \frac{X_i}{\omega_0^2} + (\mathbf{Y}_i - \boldsymbol{\alpha})'\boldsymbol{\Omega}_e^{-1}\boldsymbol{\beta}.
\end{aligned}$$

Assim temos que x_i é normal com

$$E(x_i) = \text{var}(x_i) \left[\frac{X_i}{\omega_0^2} u_i + (\mathbf{Y}_i - \boldsymbol{\alpha})' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\beta} u_i + u_i \gamma_j^{-2} (\mu_j + \Delta_j t_i) \right]$$

e

$$\text{var}(x_i) = \left(\frac{1}{\omega_0^2} u_i + \boldsymbol{\beta}' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\beta} u_i + u_i \gamma_j^{-2} \right)^{-1}.$$

No passo 6 temos a condicional completa de μ_j , $j = 1, \dots, G$, vamos agora ver o seu desenvolvimento usando a representação estocástica dada no Teorema 5

$$\begin{aligned} \pi(\mu_j | \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{T}, \mathbf{U}, \mathbf{x}, \Delta_j, \gamma_j^2, \mathbf{w}) &\propto \pi(\mathbf{Z} | \mu_j, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{T}, \mathbf{x}, \Delta_j, \gamma_j^2, \mathbf{w}) \pi(\mathbf{x} | \mu_j, \mathbf{T}, \mathbf{S}, \gamma_j^2, \Delta_j, \mathbf{w}) \pi(\mu_j) \\ &\propto \pi(\mathbf{x} | \mu_j, \mathbf{T}, \mathbf{S}, \gamma_j^2, \Delta_j) \pi(\mu_j). \end{aligned}$$

Assim temos

$$\begin{aligned} &\propto \prod_{i=1}^n \exp \left\{ -\frac{u_i}{2\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 \right\} \exp \left\{ -\frac{1}{2b_\mu} (\mu_j - a_\mu)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2\gamma_j^2} \sum_{i=1}^n u_i (\mu_j^2 - 2(x_i - \Delta_j t_i) \mu_j) \right\} \exp \left\{ -\frac{1}{2b_\mu} (\mu_j^2 - 2\mu_j a_\mu) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\mu_j^2 \gamma_j^{-2} \sum_{i=1}^n u_i - 2\mu_j \gamma_j^{-2} \sum_{i=1}^n (x_i - \Delta_j t_i) u_i \right) \right\} \exp \left\{ -\frac{1}{2} (b_\mu^{-1} \mu_j^2 - 2\mu_j a_\mu b_\mu^{-1}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\mu_j^2 \left(\gamma_j^{-2} \sum_{i=1}^n u_i + b_\mu^{-1} \right) - 2\mu_j \left(\gamma_j^{-2} \sum_{i=1}^n (x_i - \Delta_j t_i) u_i + a_\mu b_\mu^{-1} \right) \right] \right\}. \end{aligned}$$

Então temos que, $\mu_j | \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{T}, \mathbf{U}, \mathbf{x}, \Delta_j, \gamma_j^2, \mathbf{w} \sim N(\mu_{\mu_j}, \sigma_{\mu_j}^2)$ onde

$$\sigma_{\mu_j}^2 = \left(\gamma_j^{-2} \sum_{i=1}^n u_i + b_\mu^{-1} \right)^{-1} \quad e \quad \mu_{\mu_j} = \sigma_{\mu_j}^2 \left(\gamma_j^{-2} \sum_{i=1}^n (x_i - \Delta_j t_i) u_i + a_\mu b_\mu^{-1} \right).$$

No passo 7 temos a condicional completa de Δ_j , $j = 1, \dots, G$, vamos agora ver o seu desenvolvimento usando a representação estocástica dada no Teorema 5

$$\begin{aligned} \pi(\Delta_j | \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{T}, \mathbf{x}, \gamma_j^2, \mu_j, \mathbf{w}) &\propto \pi(\mathbf{Z}_i | \Delta_j, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{T}, \mathbf{x}, \gamma_j^2, \mu_j, \mathbf{w}) \pi(\mathbf{x} | \mathbf{S}, \mu_j, \gamma_j^2, \Delta_j) \pi(\Delta_j) \\ &\propto \pi(\mathbf{x} | \mathbf{U}, \mu_j, \gamma_j^2, \Delta_j, \mathbf{w}) \pi(\Delta_j). \end{aligned}$$

Assim temos

$$\begin{aligned}
&\propto \prod_{i=1}^n \exp \left\{ -\frac{u_i}{2\gamma_j^2} (x_i - \mu_j - \Delta_j t_i)^2 \right\} \exp \left\{ -\frac{1}{2d_\Delta} (\Delta_j^2 - c_\Delta)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2\gamma_j^2} \sum_{i=1}^n u_i (\Delta_j^2 t_i^2 - 2(x_i - \mu_j) \Delta_j t_i) \right\} \exp \left\{ -\frac{1}{2d_\Delta} (\Delta_j^2 - 2\Delta_j c_\Delta) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\Delta_j^2 \gamma_j^{-2} \sum_{i=1}^n u_i t_i^2 - 2\Delta_j \gamma_j^{-2} \sum_{i=1}^n u_i (x_i - \mu_j) t_i + \Delta_j^2 d_\Delta^{-1} - 2d_\Delta^{-1} \Delta_j c_\Delta \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\Delta_j^2 \left(\gamma_j^{-2} \sum_{i=1}^n u_i t_i^2 + d_\Delta^{-1} \right) - 2\Delta_j \left(\gamma_j^{-2} \sum_{i=1}^n u_i (x_i - \mu_j) t_i + d_\Delta^{-1} c_\Delta \right) \right] \right\}.
\end{aligned}$$

Então temos que, $\Delta_j | \mathbf{Z}, \beta, \alpha, \Omega, \mathbf{U}, \mathbf{T}, \mathbf{x}, \gamma_j^2, \mu_j, \mathbf{w} \sim N(\mu_{\Delta_j}, \sigma_{\Delta_j}^2)$ onde

$$\sigma_{\Delta_j}^2 = \left(\gamma_j^{-2} \sum_{i=1}^n u_i t_i^2 + d_\Delta^{-1} \right)^{-1} \quad e \quad \mu_{\Delta_j} = \sigma_{\Delta_j}^2 \left(\gamma_j^{-2} \sum_{i=1}^n u_i (x_i - \mu_j) t_i + d_\Delta^{-1} c_\Delta \right)$$

No passo 8 temos a condicional completa do vetor α , vamos agora ver o seu desenvolvimento usando a representação estocástica dada no Teorema 5

$$\begin{aligned}
\pi(\alpha | \mathbf{Z}, \beta, \Omega, \mathbf{x}, \mathbf{U}, \mathbf{T}, \mu_j, \Delta_j, \gamma_j^2, \mathbf{w}) &\propto \pi(\mathbf{Z} | \alpha, \beta, \Omega, \mathbf{x}, \mathbf{U}, \mathbf{w}) \pi(\alpha) \\
&\propto \prod_{i=1}^n \exp \left\{ -\frac{u_i}{2} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \Omega^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) \right\} \\
&\quad \times \exp \left\{ -\frac{1}{2} (\alpha - \mathbf{C}_\alpha)' \mathbf{D}_\alpha^{-1} (\alpha - \mathbf{C}_\alpha) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n [-2u_i (\mathbf{Z}_i - \mathbf{b}x_i)' \Omega^{-1} \mathbf{a} + u_i \mathbf{a}' \Omega^{-1} \mathbf{a}] \right\} \\
&\quad \times \exp \left\{ -\frac{1}{2} [\alpha' \mathbf{D}_\alpha^{-1} \alpha - 2\alpha' \mathbf{D}_\alpha^{-1} \mathbf{C}_\alpha] \right\}.
\end{aligned}$$

Desenvolvendo os termos que estão em função de \mathbf{a} e \mathbf{b} .

O primeiro termo é,

$$\begin{aligned}
(\mathbf{Z}_i - \mathbf{b}x_i)' \Omega^{-1} \mathbf{a} &= \begin{bmatrix} X_i - x_i & (\mathbf{Y}_i - x_i \beta)' \end{bmatrix} \begin{bmatrix} \omega_0^{-2} & \mathbf{0} \\ \mathbf{0} & \omega^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \\
&= \begin{bmatrix} (X_i - x_i) \omega_0^{-2} & (\mathbf{Y}_i - x_i \beta)' \omega^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \\
&= (\mathbf{Y}_i - x_i \beta)' \omega^{-1} \alpha = \alpha' \omega^{-1} (\mathbf{Y}_i - x_i \beta),
\end{aligned}$$

o segundo termo é,

$$\begin{aligned}\mathbf{a}'\Omega^{-1}\mathbf{a} &= \begin{bmatrix} 0 & \boldsymbol{\alpha}' \end{bmatrix} \begin{bmatrix} \omega_0^{-2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\omega}^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \boldsymbol{\alpha} \end{bmatrix} \\ &= \begin{bmatrix} 0 & \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \boldsymbol{\alpha} \end{bmatrix} = \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}\boldsymbol{\alpha}.\end{aligned}$$

Voltando a condicional e substituindo os termos temos,

$$\begin{aligned}&\propto \exp \left\{ -\frac{1}{2} \left[-2 \sum_{i=1}^n u_i \boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} (\mathbf{Y}_i - x_i \boldsymbol{\beta}) + \boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} \boldsymbol{\alpha} \sum_{i=1}^n u_i + \boldsymbol{\alpha}' \mathbf{D}_\alpha^{-1} \boldsymbol{\alpha} - 2 \boldsymbol{\alpha}' \mathbf{D}_\alpha^{-1} \mathbf{C}_\alpha \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[-2 \boldsymbol{\alpha}' \left(\sum_{i=1}^n u_i \boldsymbol{\omega}^{-1} (\mathbf{Y}_i - x_i \boldsymbol{\beta}) + \mathbf{D}_\alpha^{-1} \mathbf{C}_\alpha \right) + \boldsymbol{\alpha}' \left(\boldsymbol{\omega}^{-1} \sum_{i=1}^n u_i + \mathbf{D}_\alpha^{-1} \right) \boldsymbol{\alpha} \right] \right\}.\end{aligned}$$

Então temos que, $\boldsymbol{\alpha} | \mathbf{Z}, \boldsymbol{\beta}, \Omega, \mathbf{x}, \mathbf{U}, \mathbf{T}, \mu_j, \Delta_j, \gamma_j^2, \mathbf{w} \sim N_r(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha)$ onde

$$\boldsymbol{\Sigma}_\alpha = \left(\boldsymbol{\omega}^{-1} \sum_{i=1}^n u_i + \mathbf{D}_\alpha^{-1} \right)^{-1} \quad e \quad \boldsymbol{\mu}_\alpha = \boldsymbol{\Sigma}_\alpha \left[\sum_{i=1}^n u_i \boldsymbol{\omega}^{-1} (\mathbf{Y}_i - x_i \boldsymbol{\beta}) + \mathbf{D}_\alpha^{-1} \mathbf{C}_\alpha \right].$$

A condicional completa para passo 9 é obtida a partir do Teorema 5. É possível obter amostras usando o Teorema 6, porém a condicional completa resultante não tem uma forma padrão e, sempre que possível, estamos evitando a utilização de passos de Metropolis-Hastings, assim

$$\begin{aligned}\pi(\boldsymbol{\beta} | \mathbf{Z}, \mathbf{t}, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\gamma}, \boldsymbol{\omega}) &\propto \left[\prod_{i=1}^n \pi(\mathbf{Z}_i | t_i, w_i, u_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\gamma}, \boldsymbol{\omega}) \right] \pi(\boldsymbol{\beta}) \\ &\propto \exp \left\{ -\frac{1}{2} \left[-2 \sum_{i=1}^n u_i x_i \mathbf{b}' \Omega^{-1} (\mathbf{z}_i - \mathbf{a}) + \mathbf{b}' \Omega^{-1} \mathbf{b} \sum_{i=1}^n u_i x_i^2 \right] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left[-2 \boldsymbol{\beta}' \mathbf{D}_\beta^{-1} \mathbf{c}_\beta + \boldsymbol{\beta}' \mathbf{D}_\beta^{-1} \boldsymbol{\beta} \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[-2 \left(\omega_0^{-2} \sum_{i=1}^n u_i x_i X_i + \boldsymbol{\beta}' \sum_{i=1}^n u_i x_i \Omega_e^{-1} (\mathbf{y}_i - \boldsymbol{\alpha}) \right) \right. \right. \\ &\quad \left. \left. + \omega_0^{-2} \sum_{i=1}^n u_i x_i^2 + \boldsymbol{\beta}' \left(\Omega_e^{-1} \sum_{i=1}^n u_i x_i^2 \right) \boldsymbol{\beta} \right] \right\} \exp \left\{ -\frac{1}{2} \left[-2 \boldsymbol{\beta}' \mathbf{D}_\beta^{-1} \mathbf{c}_\beta + \boldsymbol{\beta}' \mathbf{D}_\beta^{-1} \boldsymbol{\beta} \right] \right\} \\ &\quad \exp \left\{ -\frac{1}{2} \left[-2 \boldsymbol{\beta}' \left(\sum_{i=1}^n u_i x_i \Omega_e^{-1} (\mathbf{y}_i - \boldsymbol{\alpha}) + \mathbf{D}_\beta^{-1} \mathbf{c}_\beta \right) + \boldsymbol{\beta}' \left(\Omega_e^{-1} \sum_{i=1}^n u_i x_i^2 + \mathbf{D}_\beta^{-1} \right) \boldsymbol{\beta} \right] \right\}.\end{aligned}$$

Os passos 10 e 11 são simples. Para o passo 12, vamos definir $k \in \{0, 1, \dots, r\}$, onde

$$\pi(\omega_k^2 | \mathbf{Z}, \mathbf{t}, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \gamma) \propto (\omega_k^{-2})^{n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n u_i (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) \right\} \\ \times (\omega_k^{-2})^{l+1} \exp\{-m\omega_k^{-2}\}.$$

Para $k \in \{1, \dots, r\}$, vamos denotar a k -ésima coordenada dos vetores \mathbf{Y}_i , $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ por Y_{ik} , α_k e β_k , respectivamente. Logo podemos reescrever a seguinte expressão,

$$(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) = (X_i - x_i, Y_{i1} - \alpha_1 - \beta_1 x_i, \dots, Y_{ir} - \alpha_r - \beta_r x_i)' \text{diag}\{\omega_0^{-2}, \dots, \omega_r^{-2}\} \\ \times (X_i - x_i, Y_{i1} - \alpha_1 - \beta_1 x_i, \dots, Y_{ir} - \alpha_r - \beta_r x_i) \\ = \omega_0^{-2} (X_i - x_i)^2 + \sum_{k=1}^r \omega_k^{-2} (Y_{ij} - \alpha_k - \beta_k x_i)^2.$$

Assim temos que $\pi(\omega_k^2 | \mathbf{Z}, \mathbf{t}, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta}, \gamma)$ é

$$\propto (\omega_k^{-2})^{n/2} \exp \left\{ -\frac{1}{2} \left[\omega_0^{-2} \sum_{i=1}^n u_i (X_i - x_i)^2 + \sum_{k=1}^r \omega_k^{-2} \sum_{i=1}^n u_i (Y_{ij} - \alpha_k - \beta_k x_i)^2 \right] \right\} \\ \times (\omega_k^{-2})^{l+1} \exp\{-m\omega_k^{-2}\}.$$

Referências Bibliográficas

- [1] Arellano-Valle, R. B. & Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96**, 93–116.
- [2] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- [3] Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**(4), 715–726.
- [4] Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, **24**, 1–13.
- [5] Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, And Applications*. Chapman & Hall/CRC.
- [6] 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.
- [7] 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.
- [8] Carroll, R. J., Maca, J. D. & Ruppert, D. (1999a). Non parametric regression in the presence of measurement error. *Biometrika*, **86**, 541–554.
- [9] Carroll, R. J., Roeder, K. & Wassermann, L. (1999b). Flexible parametric measurement error models. *Biometrics*, **55**, 44–54.
- [10] Casella, G. & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167–174.
- [11] Celeux, G., Forbes, F., Robert, C. P. & Titterton, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, **1**, 651–674.

- [12] Cheng, C. L. & Van Ness, J. W. (1999). *Statistical regression with measurement error*. Arnold, London.
- [13] Chipkevitch, E., Nishimura, R. T., Tu, D. G. S. & Galea-Rojas, M. (1996). Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods. *The Journal of Urology*, **156**, 2050–2053.
- [14] 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.
- [15] Fuller, W. A. (1987). *Measurement Error Models*. John Wiley and Sons, New York.
- [16] Galea-Rojas, M., Bolfarine, H. & de Castro, M. (2002). Local influence in comparative calibration models. *Biometrical Journal*, **44**, 59–81.
- [17] Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo*. Chapman & Hall, second edition.
- [18] Gelfand, A. E. (2000). Gibbs sampling. *Journal of The American Statistical Association*, **95**, 1300–1304.
- [19] Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271–275.
- [20] Lachos, V. H., Garibay, V., Labra, F. V. & Aoki, R. (2009). Robust multivariate measurement error model with skew-normal/independent distributions and Bayesian MCMC implementation. *Statistical Methodology*, **6**, 527–541.
- [21] Lachos, V. H., Labra, F. V., Bolfarine, H. & Ghosh, P. (2010). Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics*, **44**, 541–556.
- [22] Muller, P. & Roeder, K. (1997). A Bayesian semi parametric model for case-control studies with errors in variables. *Biometrika*, **84**, 523–537.
- [23] 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.
- [24] Richardson, S., Leblond, L., Jaussent, I. & Green, P. J. (2002). Mixture models in measurement error problems, with reference to epidemiological studies. *Journal of the Royal Statistical Society, Series A*, **165**, 549–566.
- [25] Roy, S. & Banerjee, T. (2006). A flexible model for generalized linear regression with measurement error. *Annals of the Institute of Statistical Mathematics*, **58**, 153–169.

- [26] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. John Wiley and Sons.
- [27] Sperrin, M., Jaki, T. & Wit, E. (2010). Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics and Computing*, **20**, 357–366.
- [28] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.
- [29] Yao, W. (2012). Model based labeling for mixture models. *Statistics and Computing*, **22**, 337–347.
- [30] Zhang, F. (2011). *Matrix Theory: Basic Results and Techniques*. Springer, second edition.