Inferência bayesiana em modelos lineares mistos t-assimétricos

Cristian Luis Bayes Rodríguez

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

Programa: Estatística
Orientador: Profa. Dra. Márcia D'Elia Branco

Durante o desenvolvimento deste trabalho o autor recebeu auxílio FAPESP

São Paulo, março de 2009

Inferência bayesiana em modelos lineares mistos t-assimétricos

Este exemplar corresponde à redação final da dissertação/tese devidamente corrigida e defendida por Cristian Luis Bayes Rodríguez e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Profa. Dra. Márcia D'Elia Branco (orientador) - IME-USP.
- Prof. Dr. Heleno Bolfarine - IME-USP.
- Profa. Dra. Rosangela Helena Loschi - UFMG.
- Prof. Dr. Marc Georges Genton - TAU.
- Prof. Dr. Reinaldo Boris Arellano Valle - PUC - Chile.

Agradecimentos

- Aos meus pais, Luis e Bertila, por todo o carinho, motivação e apoio em todos os sentidos que sempre me deram.
- À minha orientadora Prof. Dra. Márcia D'Elia Branco, por sua compreensão, dedicação e orientação, que de uma maneira clara e objetiva sempre me direcionou para um bom desenvolvimento desse trabalho.
- Aos membros da banca examinadora: Prof. Dr. Heleno Bolfarine, Profa. Dra. Rosangela Helena Loschi, Prof. Dr. Marc Georges Genton e Prof. Dr. Reinaldo Boris Arellano Valle pelos seus valiosos aportes e sugestões.
- Aos companheiros da república: Christian Paz, Jesús, Karina, Javier, Tiago, Christian Noriega, Gustavo e Hilda que se tornaram minha família aqui em São Paulo.
- Aos amigos Lourdes, Juan Carlos, Jorge, Jannet, Víctor Hugo, Vladimir, Eduardo, Thiago, Iván pelo prazer de suas amizades, conversas e trocas de conhecimentos.
- A Erika pelo apoio, preocupação e carinho nesta fase da minha vida.
- Ao Prof. Dr. Vladimir Belitsky pela sua ajuda e amizade durante o mestrado.
- Ao IME e à USP pela oportunidade de fazer meu doutorado nesta Instituição, e à FAPESP pelo apoio financeiro.
- Aos amigos que incentivaram minha vinda para o IME-USP, como o Prof. Victor Maehara e o Prof. Jorge Chue do Departamento de Estatística - UNALM, ao Prof. Antonio Bravo e muitos outros.

Finalmente, agradeço a Deus, por sua proteção, e por ter me dado força e perseverança ao longo de todo este período.

Resumo

Esta tese compreende um estudo das propriedades da distribuição t-assimétrica fundamental, uma das vantagens desta distribuição é o fato que permite modelar dados que apresentam assimetria e curtose diferentes da distribuição normal. Modelos lineares mistos são muito utilizados na análise de dados com medidas repetidas porque permitem modelar a correlação entre sujeitos. Uma suposição usual é a normalidades dos efeitos aleatórios e dos erros. Neste trabalho, estendemos este modelo assumindo a distribuição t-assimétrica tanto para os erros como para os efeitos aleatórios, são analisados várias possíveis formas de estender este modelo, por exemplo, (i) considerando que os erros e os efeitos aleatórios seguem conjuntamente uma distribuição t-assimétrica e (ii) assumindo que eles são independentes e seguem a distribuição t-assimétrica. Resultam como casos especiais destas suposições os modelos simétricos t-Student e normal e o modelo normal-assimétrico. Para a obtenção das estimativas desenvolvemos algoritmos do tipo MCMC. Aplicações a diversos conjuntos de dados são apresentadas.

Palavras-chave: distribuição t-assimétrica, modelos lineares mistos, inferência bayesiana.

Abstract

This work approaches a study of the properties of the fundamental skew-t distribution. One advantage of this class of distribution is the exibility to work with asymmetrical and heavy tailed data. Linear mixed models are frequently used in repeated measures data analysis. However, the usual assumption of normality for the error and the random effects makes inference vulnerable to the presence of outliers. Here we propose to consider a skew-t distribution for both errors and random effects. We analize several cases to extend this model, for example, (i) considering errors and random effects with a joint skew-t distribution and (ii) considering that errors and random effects are independent and follow a skew-t distribution. A Bayesian inference approach is adopted using MCMC algorithms. The models developed are illustrated using several real data sets.
Keywords: skew-t distribution, linear mixed models, bayesian inference.

Sumário

Lista de Figuras ix

Lista de Tabelas xi

1 Introdução 1

1.1 Organização 2

2 Distribuição t-assimétrica 5

2.1 Distribuição t-assimétrica fundamental 5

2.2 Distribuição t-assimétrica tipo I 11

2.2.1 Distribuição Normal assimétrica tipo I 12

2.3 Distribuição t-assimétrica tipo II 13

2.3.1 Distribuição Normal assimétrica tipo II 14

2.4 Interpretação dos parâmetros 15

3 Modelo de Regressão Multivariada 19

3.1 Introdução 19

3.2 Modelo de Regressão Multivariada com a distribuição t-assimétrica 19

3.3 Distribuições *a Priori* 20

3.4 O Modelo Hierárquico e o Algoritmo MCMC 21

3.5 Comparação de Modelos 22

3.6 Estudo de Simulação 23

4 Modelo Linear Misto t-assimétrico Dependente 25

4.1 Introdução 25

4.2 Modelo Linear Misto com a distribuição t-assimétrica 26

4.3 Distribuições *a Priori* 29

4.4 O Modelo Hierárquico e o Algoritmo MCMC 29

4.5 Aplicação 31

5 Modelo Linear Misto t-assimétrico Independente 35

5.1 Introdução 35

5.2 Modelo Linear Misto t-assimétrico independente 35

5.3 Distribuições *a Priori* 37
5.4 O Modelo Hierárquico e o Algoritmo MCMC 38
5.5 Aplicação 40
5.5.1 Análise de Sensibilidade 41
5.5.2 Resultados 42

6 Estudo de Robustez **47**
6.1 Introdução 47
6.2 Influência nos efeitos fixos 47
6.3 Influência Global 49

7 Modelo com Erros das Variáveis de Intercepto Nulo **53**
7.1 Introdução 53
7.2 Modelo com Erros nas variáveis com intercepto nulo 53
7.3 Distribuição *a Priori* 55
7.4 O Modelo Hierárquico e o Algoritmo MCMC 55
7.5 Aplicação 57
7.6 Estudo de Robustez 61

8 Problemas de inferência em modelos t-assimétricos multivariados **63**
8.1 Modelo 63
8.2 A Distribuição *à Posteriori* 64

9 Considerações Finais **67**
9.1 Perspectiva Futuras 68

A CPO e Pseudo-Fator de Bayes **71**

B Distribuição condicional dos graus de liberdade **73**

Referências Bibliográficas **75**

Lista de Figuras

2.1 Coeficiente de assimetria γ_1 como uma função de λ para graus de liberdade $\nu =$ 4(linha solida), 5(linha tracejada), 10(linha pontilhada) e 20(linha ponto traço) . . . 16

4.1 Histograma dos níveis de colesterol 31
4.2 Histograma dos interceptos e inclinações dos perfis individuais 32
4.3 CPO para os modelos: (1) St, (2) SN, (3) t e (4) N dos dados de Framningham . . 33

5.1 o peso no nascimento (em gramas) da crias de ratos 41
5.2 Distribuição *a priori* dos parâmetros graus de liberdade e distribuição *a posteriori* dos efeitos fixos para as distribuições *a priori* I (linha solida), II (linha tracejada), III (linha pontilhada), IV (linha ponto traço) 43
5.3 Distribuição *a priori* dos parâmetros de assimetria e distribuição *a posteriori* dos efeitos fixos para as distribuições *a priori* I (linha solida), II (linha tracejada), III ((linha pontilhada) e IV (linha ponto traço) 44
5.4 Distribuição *a priori* e distribuição *a posteriori* dos parâmetros de graus de liberdade para as distribuições *a priori* I (linha solida), II (linha tracejada), III ((linha pontilhada) e IV (linha ponto traço) 45
5.5 Distribuição *a priori* e *a posteriori* dos parâmetros de assimetria para as para as distribuições *a priori* I (linha solida), II (linha tracejada), III ((linha pontilhada) e IV (linha ponto traço) 45
5.6 CPO para os modelos: (1) N−N, (2) N−t, (3) N−SN e (4) N−St 46

6.1 Média *a posteriori* (linha tracejada) e intervalo HPD de 95% (linha solida) para β_0 sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação 48
6.2 Média *a posteriori* (linha tracejada) e intervalo HPD de 95% (linha solida) para β_1 sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação 49
6.3 Amplitude do intervalo HPD de 95% para β_0 para o modelo dependente (linha preta) e para o modelo dependente (linha tracejada) sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação . 50

6.4 Amplitude do intervalo HPD de 95% para β_1 para o modelo dependente (linha preta) e para o modelo dependente (linha tracejada) sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação . 51
6.5 Distancia de Kullback–Leibler entre $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y})$ e $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}(\Delta))$ para os modelos: N (linha solida), SN (linha tracejada), t (linha pontilhada) e St (linha ponto-traço) . 52

7.1 Histograma do índice de placa dental no *baseline* 58
7.2 Box plot de λ_x. 60
7.3 Média *a posteriori* (linha tracejada) e HPD de 95% de probabilidade (linha sólida) para β_{1_1} dos modelos N, SN, St para diferentes contaminações Δ de uma observação 61
7.4 Média *a posteriori* (linha tracejada) e HPD de 95% de probabilidade (linha sólida) para β_{2_1} dos modelos N, SN, St para diferentes contaminações Δ de uma observação 62

Lista de Tabelas

3.1 Estudo de Simulação: Porcentagem de vezes o critérios escolhem os modelos das colunas sob diferentes cenários 24

4.1 Casos Especiais do MLM t-assimétrico dependente 30
4.2 Medidas de comparação de modelos para os dados de Framingham 32
4.3 Média *a posteriori*, desvio padrão (D.P) e intervalo HPD de 95% de probabilidade para o modelo St 33

5.1 Casos Especiais do MLM t-assimétrico independente 37
5.2 Medidas de comparação de modelos para os dados toxicológicos 45
5.3 Média *a posteriori* e intervalo HPD de 95%− para os modelos N $-St$ e $St - St$ . . . 46

7.1 Critérios de comparação de modelos. 59
7.2 Média *a posteriori*, desvio padrão e intervalo HPD de 95% para os modelos SN e St. 60
7.3 Média *a posteriori* e HPD de 95% para o Modelo 2 61

Capítulo 1

Introdução

Em muitas situações práticas a usual suposição de normalidade não é satisfeita devido a falta de simetria dos dados. Azzalini (1985) propõe como alternativa a utilização de uma família mais geral de distribuições, de forma que consiga modelar a assimetria dos dados e além disso, incluir a distribuição normal como um caso particular. Esta família de distribuições foi denominada normal assimétrica.

A distribuição normal assimétrica univariada surgiu independentemente em vários artigos estatísticos, entre os principais trabalhos pode-se destacar Roberts (1966), O'Hagan e Leonard (1976) e Aigner, Lovell e Schmidt (1977). Entretanto, foi Azzalini (1985) que introduziu formalmente esta distribuição, estudou suas propriedades e mostrou que a distribuição tem problemas na estimação do parâmetro que controla a assimetria, pelos métodos de estimação usuais (métodos dos momentos e de máxima verossimilhança). Métodos alternativos tem sido estudados por Sartori (2003) na abordagem clássica, e na abordagem bayesiana, por Liseo e Loperfido (2004). Estes últimos propõem a utilização da priori de Jeffreys, entretanto esta priori tem uma expressão difícil de trabalhar. Bayes e Branco (2007) desenvolvem uma aproximação para a priori de Jeffreys e estudam a estimação dos parâmetros da distribuição normal assimétrica sob o enfoque bayesiano.

Distribuições de probabilidade com caudas mais pesadas que a distribuição normal tem sido propostas na literatura como uma alternativa para modelagem dos erros no modelo de regressão. Tais distribuições tem a vantagem de incorporar observações consideradas discrepantes sob a normalidade. Um trabalho pioneiro nesta área é Zellner (1976), onde a distribuição t-Student multivariada é considerada. Essas idéias foram estendidas por Osiewalski e Steel (1993) considerando a classe das distribuições elípticas (Kelker, 1970). Em Branco, Bolfarine, Iglesias e Arellano-Valle (2000) é apresentado uma proposta do uso das distribuições elípticas para o problema de calibração controlada, além da generalização para a classe das elípticas de alguns resultados em regressão bayesiana sob normalidade. Posteriormente, Branco e Dey (2001), Branco e Dey (2002) e Sahu, Dey e Branco (2003) propuseram novas classes de distribuições multivariadas que são assimétricas e incluam como casos particulares distribuições simétricas multivariadas conhecidas, como a normal, a t-Student, a Pearson II e a classe das distribuições elípticas. Em Sahu et al. (2003) algumas aplicações são consideradas sob modelos de regressão bayesianos. Utilizando-se a distribuição t-assimétrica para os erros, isto nos permite incorporar ao modelo parâmetros que quantificam a assimetria e curtose

dos erros.

Arellano-Valle e Genton (2005) propõem uma classe bastante geral de distribuições assimétricas denominada esférica assimétrica fundamental. Arellano-Valle e Azzalini (2006) e Arellano-Valle, Branco e Genton (2006) construem classes ainda mais gerais de distribuições que tem como caso particular a esférica assimétrica fundamental, estas novas classes estão baseadas em mecanismos de seleção que geram distribuições assimétricas.

A flexibilidade da distribuição t-assimétrica para modelagem de dados, faz com que ela seja uma alternativa interessante em diversas áreas de aplicação. Para alguns exemplos, ver Genton (2004).

Nesta tese de doutorado, estudamos a distribuição t-assimétrica fundamental, proposta como un exemplo das distribuições esféricas fundamentais em Arellano-Valle e Genton (2005), que tem como casos especiais as distribuições t-assimétricas propostas por Branco e Dey (2001) e por Sahu et al. (2003) que denominaremos como tipo I e tipo II respectivamente. Assim os resultados obtidos para a t-assimétrica fundamental serão validos para ambos casos.

Também, propomos trabalhar com modelos lineares mistos com erros t-assimétricos para conseguir modelar a assimetria e curtose muitas vezes presente neste tipo de dados e consideraremos a abordagem bayesiana. Discussão de escolha de distribuições *a priori* e análise de sensibilidade serão alguns dos pontos tratados. Outro aspecto, refere-se a escolha de modelos e a robustez dos modelos a presença de valores *outliers*.

1.1 Organização

O presente trabalho está dividido em nove capítulos. No segundo capítulo, apresentamos definições e propriedades da distribuição t-assimétrica fundamental.

No terceiro capitulo, apresentamos o modelo linear com a distribuição t-assimétrica tipo II, implementamos o algoritmo de Gibbs e obtemos formas conhecidas para as distribuições condicionais completas do modelo, com exceção da distribuição condicional dos graus de liberdade.

No quarto capítulo, apresentamos o modelo linear misto com erros e efeitos aleatórios t-assimétricos dependentes, isto é, o vetor dos erros e os dos efeitos aleatórios seguem uma distribuição conjunta t-assimétrica, estudamos suas propriedades, implementamos o algoritmo de Gibbs e ilustramos o modelo como um conjunto de dados de níveis de colesterol.

No quinto capítulo, apresentamos o modelo linear misto com erros e efeitos aleatórios t-assimétricos independentes, consideramos neste caso a distribuição t-assimétrica tipo I, propomos uma reparametrização do modelo que nos permite obter formas conhecidas para as distribuições condicionais completas do modelo, com exceção da distribuição condicional dos graus de liberdade, e para aplicação consideramos os dados de um estudo toxicológico.

No sexto capítulo, avaliamos a robustez dos modelos apresentado nos capítulos quarto e quinto, consideramos um conjunto de dados dentários já estudados anteriormente na literatura para estudar a influência de um valor *outlier* na distribuição *a posteriori* dos parâmetros.

No sétimo capítulo, apresentamo um modelo com erros nas variáveis de intercepto nulo, considerando a distribuição t-assimétrica tipo I, conseguimos formas conhecidas para as distribuições

condicionais completas e ilustramos a metodologia com um conjunto de dados de um estudo odontológico.

No oitavo capítulo, apresentamos uma discussão sob a dificuldade de estimação do parâmetro graus de liberdade em modelos não-lineares que consideram a distribuição t-assimétrica multivariada para o vetor de erros. Na abordagem bayesiana, iremos mostrar que, que sob certas condições, a distribuição *a posteriori* não é atualizada, isto é, não depende dos dados.

Finalmente no nono capítulo, apresentamos as conclusões dos resultados obtidos neste trabalho e perspectivas de pesquisa futura.

Capítulo 2

Distribuição t-assimétrica

Este capítulo apresenta algumas propriedades da distribuição t-assimétrica fundamental proposta por Arellano-Valle e Genton (2005). Existem múltiplas definições com essa mesma denominação t assimétrica, por exemplo, Azzalini e Capitanio (2003), Sahu et al. (2003), Fernández e Steel (1998) e Jones e Faddy (2003) sendo que as duas ultimas estão definidas somente para o caso univariado. Nesta seção, consideraremos a t-assimétrica fundamental que é um caso especial da classe de distribuições esféricas assimétricas fundamentais, proposta por Arellano-Valle e Genton (2005).

2.1 Distribuição t-assimétrica fundamental

Nesta seção introduzimos a distribuição t-assimétrica fundamental que é um caso especial da classe de distribuições esféricas assimétricas fundamentais, proposta por Arellano-Valle e Genton (2005), apresentada no artigo através de um exemplo. Tal distribuição será explorada com mais detalhes neste trabalho. Uma das vantagens desta definição é o fato de conter como casos especiais a t-assimétrica de Branco e Dey (2001) e a t-assimétrica de Sahu et al. (2003)

Definição 2.1 *Uma vetor aleatório $\boldsymbol{Y}$ de dimensão n tem distribuição t-assimétrica multivariada com parâmetros de posição $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, assimetria $\boldsymbol{D}$ e graus de liberdade ν, se sua função de densidade de probabilidade é dada por*

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& 2^m t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T, \nu) \\ & T_m\left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu}\right)^{1/2} \mid \boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D}, \nu + n\right), \end{aligned} \tag{2.1}$$

sendo que $\mu \in \Re^n$, $\boldsymbol{\Omega}$ matriz $n \times n$ definida positiva, $\boldsymbol{D}$ matriz $n \times m$, $\nu > 0$ e $q(\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{\mu})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$. Denotamos por $t_n(. \mid \boldsymbol{a}, \boldsymbol{B}, c)$ a função de densidade de probabilidade de uma t-Student n-variada com parâmetro de posição $\boldsymbol{a}$, escala $\boldsymbol{B}$ e graus de liberdade c, e por $T_n(. \mid \boldsymbol{B}, c)$ a função de distribuição de uma t-Student n-variada com parâmetro de posição $\boldsymbol{a} = 0$, escala $\boldsymbol{B}$ e graus de liberdade c. Utilizaremos a seguinte notação $\boldsymbol{Y} \sim StF_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$

A seguinte reparametrização nos permite escrever (2.1) como um modelo de locação-escala,

$$\boldsymbol{\Sigma} = \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T \text{ e } \boldsymbol{\Delta} = (\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D}, \tag{2.2}$$

nesse caso $\boldsymbol{\Sigma}$ é um parâmetro de escala, e a densidade é dada por

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2^m t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T_m \left(\boldsymbol{\Delta}^T \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}) \left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu} \right)^{1/2} \mid \boldsymbol{I}_m - \boldsymbol{\Delta}^T \boldsymbol{\Delta}, \nu + n \right), \tag{2.3}$$

Utilizaremos a seguinte notação $\boldsymbol{Y} \sim StF^*(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \nu)$ para a densidade em (2.3).

Alguns casos especiais da família de distribuições definida em (2.1) são

1. Se $m = 1$ e $\boldsymbol{D} = \boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$, $\boldsymbol{\Sigma} = \boldsymbol{\Omega} + \boldsymbol{\eta}\boldsymbol{\eta}^T$ e $\boldsymbol{\lambda} = \frac{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}}{\sqrt{1 - \boldsymbol{\eta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}}}$,então (2.1) se reduz a

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2^m t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T_1 \left(\lambda^T \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}) \left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu} \right)^{1/2} \mid \nu + n \right),$$

   que é a densidade da t-assimétrica obtida em Branco e Dey (2001) e Azzalini e Capitanio (2003).

2. Se $m = n$ e $D = Diag(\eta_1, \ldots, \eta_n)$ temos que

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& 2^n t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}^2, \nu) \\ & \times T_n \left(\boldsymbol{D}^T (\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}) \left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu} \right)^{1/2} \mid \boldsymbol{I}_n - \boldsymbol{D}^T (\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1} \boldsymbol{D}, \nu + n \right), \end{aligned}$$

   assim, obtemos a t-assimétrica definida por Sahu et al. (2003).

3. Se $\nu \to \infty$, a densidade em (2.1) coincide com a distribuição normal assimétrica fundamental definida em Arellano-Valle e Genton (2005),

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& 2^m \phi_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T) \\ & \Phi_m \left(\boldsymbol{D}^T (\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \mid \boldsymbol{I}_m - \boldsymbol{D}^T (\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1} \boldsymbol{D} \right), \end{aligned} \tag{2.4}$$

   sendo $\phi_n(. \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ a função de densidade de probabilidade de uma normal n-variada com vetor de médias $\boldsymbol{\mu}$, matriz de covariância $\boldsymbol{\Sigma}$, e $\Phi_n(. \mid \boldsymbol{\Sigma})$ a função de distribuição de uma normal n-variada com vetor de médias $\boldsymbol{\mu} = 0$ e matriz de covariância $\boldsymbol{\Sigma}$, para a qual utilizaremos a seguinte notação $\boldsymbol{Y} \sim SNF_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D})$

A densidade em 2.1 possui algumas propriedades interessantes que serão apresentadas neste trabalho e para cujas provas o seguinte lema dado por exemplo em Arellano-Valle e Genton (2005) será útil.

Lema 2.1 *Se* $V \sim N_k(\boldsymbol{m}, \boldsymbol{S})$, *então*

$$E\left[\Phi_m(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{V}|\boldsymbol{b}, \boldsymbol{B})\right] = \Phi_m(\boldsymbol{a}|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{m}, \boldsymbol{B} + \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^T),$$

sendo $\Phi_n(. \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ *a função de distribuição de uma normal n-variada com vetor de médias* $\boldsymbol{\mu}$ *e matriz de covariância* $\boldsymbol{\Sigma}$.

A distribuição t-assimétrica fundamental pode ser obtida como uma mistura de uma normal assimétrica fundamental, como é mostrado na seguinte proposição,

Proposição 2.1 *Se*

$$\begin{aligned} \boldsymbol{Y} \mid w &\sim SNF_n\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Omega}}{w}, \frac{\boldsymbol{D}}{w^{1/2}}\right) \\ w &\sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \end{aligned}$$

então $\boldsymbol{Y} \sim StF(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$.

Prova : Seja $g(. \mid, a, b)$ a função de densidade de probabilidade de uma distribuição gama com parâmetros a e b. Então

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) = &\int_0^\infty 2^m \phi_n\left(\boldsymbol{y} \mid \boldsymbol{\mu}, \frac{1}{w}(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)\right) \\ &\Phi_m\left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \mid \frac{1}{w}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D})\right) g\left(w \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dw \\ =&t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T, \nu) \\ &\int_0^\infty \Phi_m\left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \mid \frac{1}{w}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D})\right) \\ &g\left(w \mid \frac{\nu + n}{2}, \frac{\nu + q(\boldsymbol{y})}{2}\right) dw \end{aligned}$$

se escrevemos a função de distribuição acumulada da normal como uma integral e consideramos $w^* = \dfrac{\nu + q(\boldsymbol{y})}{\nu + n} w$ temos que

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =&t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T, \nu) \\ &\int_0^\infty \int_{\boldsymbol{t} > \boldsymbol{D}^T(\Omega + \boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y} - \boldsymbol{\mu})} \phi_m\left(\boldsymbol{t} \mid \boldsymbol{0}, \frac{\nu + n}{\nu + q(\boldsymbol{y})} \frac{1}{w^*}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D})\right) \\ &g\left(w^* \mid \frac{\nu + n}{2}, \frac{\nu + n}{2}\right) d\boldsymbol{t} dw \end{aligned}$$

mudando a ordem de integração temos que

$$
\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) =& t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T, \nu) \\
& \int_{\boldsymbol{t} > \boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu})} \int_0^\infty \phi_m \left(\boldsymbol{t} \mid \boldsymbol{0}, \frac{\nu+n}{\nu+q(\boldsymbol{y})} \frac{1}{w^*} (\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D}) \right) \\
& g\left(w^* \mid \frac{\nu+n}{2}, \frac{\nu+n}{2} \right) dw d\boldsymbol{t} \\
=& t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T, \nu) \\
& \int_{\boldsymbol{t} > \boldsymbol{D}^T(\Omega+\boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu})} t_m \left(\boldsymbol{t} \mid \boldsymbol{0}, \frac{\nu+n}{\nu+q(\boldsymbol{y})} (\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D}), \nu+n \right) d\boldsymbol{t} \\
=& 2^m t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T, \nu) \\
& T_m \left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \left(\frac{\nu+n}{q(\boldsymbol{y})+\nu} \right)^{1/2} \mid \boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D}, \nu+n \right).
\end{aligned}
$$

Na reparametrização (2.2) observamos que a distribuição t-assimétrica fundamental é uma mistura na escala de uma normal assimétrica fundamental,

$$
\begin{aligned}
\boldsymbol{Y} \mid w &\sim SNF_n^* \left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{w}, \boldsymbol{\Delta} \right) \\
w &\sim Gama \left(\frac{\nu}{2}, \frac{\nu}{2} \right).
\end{aligned}
$$

A seguir obtemos a função geradora de momentos (f.g.m) da t-assimétrica fundamental, da qual derivaremos medidas importantes para a caracterização desta distribuição, tais como, média e variância.

Proposição 2.2 *A função geradora de momentos da t-assimétrica fundamental é dada por*

$$
M_{\boldsymbol{Y}}(\boldsymbol{t}) = \int_0^\infty 2^m \exp\left\{ \boldsymbol{t}^T \boldsymbol{\mu} + \frac{\boldsymbol{t}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w} \right\} \Phi_m \left(\frac{\boldsymbol{D}^T\boldsymbol{t}}{w^{1/2}} \right) g\left(\frac{\nu}{2}, \frac{\nu}{2} \right) dw,
$$

com $g(. \mid a, b)$ a densidade de uma Gama (a, b).

Prova : Utilizando a definição de f.g.m. e a proposição 2.1, temos que

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= E\left[e^{\boldsymbol{t}^T \boldsymbol{Y}} \right] \\
&= \int_{\boldsymbol{y} \in \Re^n} \int_0^\infty e^{\boldsymbol{t}^T \boldsymbol{y}} 2^m \phi_n \left(\boldsymbol{y} \mid \boldsymbol{\mu}, \frac{1}{w}(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T) \right) \\
& \Phi_m \left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \mid \frac{1}{w}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D}) \right) g\left(w \mid \frac{\nu}{2}, \frac{\nu}{2} \right) dw d\boldsymbol{y}
\end{aligned}
$$

mudando a ordem de integração temos

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= \int_0^\infty \int_{\boldsymbol{y}\in\Re^n} \exp\left(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}\right) 2^m \phi_n\left(\boldsymbol{y} \mid \boldsymbol{\mu} - (\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}, \frac{1}{w}(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\right) \\
&\quad \Phi_m\left(\boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \mid \frac{1}{w}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D})\right) g\left(w \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dydw \\
&= \int_0^\infty \exp\left(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}\right) \int_{\boldsymbol{y}\in\Re^n} 2^m \phi_n\left(\boldsymbol{y} \mid \boldsymbol{\mu} - (\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}, \frac{1}{w}(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\right) \\
&\quad \Phi_m\left(\boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \mid \frac{1}{w}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D})\right) g\left(w \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dydw
\end{aligned}
$$

Logo, podemos escrever a integral em $\boldsymbol{y}$ como uma esperança,

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= \int_0^\infty \exp\left(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}\right) \\
&\quad E_{\boldsymbol{S}}\left[\Phi_m\left(\boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \mid \frac{1}{w}(\boldsymbol{I}_m - \boldsymbol{D}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)^{-1}\boldsymbol{D})\right)\right] g\left(w \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dydw,
\end{aligned}
$$

sendo que $\boldsymbol{S} \sim N\left(\boldsymbol{\mu} - (\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}, \frac{1}{w}(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\right)$.Se utilizamos o lema 2.1 obtemos

$$
M_{\boldsymbol{Y}}(\boldsymbol{t}) = \int_0^\infty 2\exp\left(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}\right) \Phi_m\left(\frac{\boldsymbol{D}^T\boldsymbol{t}}{w^{1/2}}\right) g\left(w \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dw.
$$

Uma representação estocástica para a distribuição t-assimétrica fundamental é dada na seguinte proposição,

Proposição 2.3 *Seja* $\boldsymbol{Y} \sim StF(\boldsymbol{\mu}, \Omega, D, \nu)$, *então*

$$
\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{D}\frac{|U|}{w^{1/2}} + \boldsymbol{\Omega}^{1/2}\frac{\boldsymbol{V}}{w^{1/2}}, \tag{2.5}
$$

sendo que $U \sim N_m(0, \boldsymbol{I}_m)$, $V \sim N_n(0, \boldsymbol{I}_n)$ *e* $w \sim Gama(\nu/2, \nu/2)$ *são objetos aleatóroios independentes.*

Prova :

$$
\begin{aligned}
E\left[e^{\boldsymbol{t}^T\boldsymbol{Y}}\right] &= E\left[e^{\boldsymbol{t}^T\left(\boldsymbol{\mu}+\boldsymbol{D}\frac{|\boldsymbol{U}|}{w^{1/2}}+\boldsymbol{\Omega}^{1/2}\frac{\boldsymbol{V}}{w^{1/2}}\right)}\right] \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}}E\left[E\left[e^{\boldsymbol{t}^T\left(\boldsymbol{D}\frac{|\boldsymbol{U}|}{w^{1/2}}+\boldsymbol{\Omega}^{1/2}\frac{\boldsymbol{V}}{w^{1/2}}\right)}\right] \mid w\right] \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}}E\left[E\left[e^{\boldsymbol{t}^T\boldsymbol{D}\frac{|\boldsymbol{U}|}{w^{1/2}}}\right] E\left[e^{\boldsymbol{\Omega}^{1/2}\frac{\boldsymbol{V}}{w^{1/2}}}\right] \mid w\right] \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}}E\left[e^{\frac{\boldsymbol{t}^T\boldsymbol{\Omega}\boldsymbol{t}}{2w}} E\left[e^{\boldsymbol{t}^T\boldsymbol{D}\frac{|\boldsymbol{U}|}{w^{1/2}}}\right] \mid w\right] \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}}E\left[e^{\frac{\boldsymbol{t}^T\boldsymbol{\Omega}\boldsymbol{t}}{2w}} e^{\frac{\boldsymbol{t}^T\boldsymbol{D}\boldsymbol{D}^T\boldsymbol{t}}{2w}} 2^m \int_{\boldsymbol{U}>\boldsymbol{0}} e^{\boldsymbol{t}^T\boldsymbol{D}\frac{\boldsymbol{U}}{w^{1/2}}}\phi_m\left(\frac{\boldsymbol{D}\boldsymbol{t}}{w^{1/2}} \mid \boldsymbol{0}, \boldsymbol{I}_m\right) d\boldsymbol{U} \mid w\right] \\
&= e^{\boldsymbol{t}^T\boldsymbol{\mu}}E\left[2^m e^{\frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}} \int_{\boldsymbol{U}>\boldsymbol{0}} \phi_m\left(\boldsymbol{U} \mid -\boldsymbol{D}\boldsymbol{t}, \boldsymbol{I}_m\right) d\boldsymbol{U} \mid w\right] \\
&= E\left[2^m e^{\boldsymbol{t}^T\boldsymbol{\mu}+\frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}} \Phi_m\left(\frac{\boldsymbol{D}\boldsymbol{t}}{w^{1/2}}\right) \mid w\right] \\
&= \int_0^\infty 2^m \exp\left(\boldsymbol{t}^T\boldsymbol{\mu}+\frac{\boldsymbol{t}^T(\boldsymbol{\Omega}+\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{t}}{2w}\right)\Phi_m\left(\frac{\boldsymbol{D}^T\boldsymbol{t}}{w^{1/2}}\right) g\left(\frac{\nu}{2},\frac{\nu}{2}\right) dw.
\end{aligned}
$$

A partir da proposição 2.3 obtemos os momentos da t-assimétrica fundamental.

Proposição 2.4 *Seja $\boldsymbol{Y}$ um vetor aleatório com distribuição $StF_n(\mu, \Omega, \boldsymbol{D}, \nu)$, temos que*

$$
E(\boldsymbol{Y}) = \boldsymbol{\mu} + \kappa\boldsymbol{D}1_m \quad e \quad Var(\boldsymbol{Y}) = \frac{\nu}{\nu-2}\left(\boldsymbol{\Omega}+\left(1-\frac{2}{\pi}\right)\boldsymbol{D}\boldsymbol{D}^T\right)+\frac{2}{\pi}\left(\frac{\nu}{\nu-2}-\frac{\pi}{2}\kappa^2\right)\boldsymbol{D}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{D}^T
$$

sendo $\kappa = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$.

Prova : Segue da proposição 2.3. Arellano-Valle e Genton (2005) também obterem este resultado para uma classe mais geral de distribuições.

Corolário 2.1 *Consideremos a seguinte partição de Y*

$$
\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1}(n_1\times 1) \\ \boldsymbol{Y_2}(n_2\times 1)\end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1}(n_1\times 1) \\ \boldsymbol{\mu_2}(n_2\times 1)\end{pmatrix},
$$

$$
\boldsymbol{\Omega}\begin{pmatrix} \boldsymbol{\Omega}_{11}(n_1\times n_1) & \boldsymbol{\Omega}_{12}(n_2\times n_1) \\ \boldsymbol{\Omega}_{12}(n_2\times n_1) & \boldsymbol{\Omega}_{22}(n_2\times n_2)\end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1(n_1\times m) \\ \boldsymbol{D}_2(n_2\times m)\end{pmatrix},
$$

então Y_1 e Y_2 são não correlacionadas se $\boldsymbol{\Omega}_{12} = \mathbf{0}$, e $\boldsymbol{D}_1 = \mathbf{0}$ ou $\boldsymbol{D}_2 = \mathbf{0}$

Este resultado será utilizado na construção do modelo linear misto do capítulo 4, para obter que os erros e os efeitos aleatórios sejam não correlacionados.

A distribuição de uma transformação linear arbitraria $\boldsymbol{AY} + \boldsymbol{b}$ é dada na seguinte proposição

Proposição 2.5 *Seja $\boldsymbol{A}$ uma matriz $k \times n$ de posto completo e $\boldsymbol{b}$ um vetor de dimensão k. Se $\boldsymbol{Y} \sim StF_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$, então*

$$\boldsymbol{AY} + \boldsymbol{b} \sim StF_k(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Omega A}^T, \boldsymbol{AD}, \nu). \tag{2.6}$$

Prova : Seja $\boldsymbol{t} \in \Re^k$

$$\begin{aligned} M_{STF}(\boldsymbol{AY} + \boldsymbol{b}) &= e^{\boldsymbol{b}} M_{STF}(\boldsymbol{AY}) \\ &= e^{\boldsymbol{b}} \int_0^\infty 2\exp\left(\boldsymbol{t}^T\boldsymbol{A\mu} + \frac{\boldsymbol{At}^T(\boldsymbol{\Omega} + \boldsymbol{DD}^T)\boldsymbol{At}}{2w}\right) \Phi_m\left(\frac{\boldsymbol{D}^T\boldsymbol{A}^T\boldsymbol{t}}{w^{1/2}}\right) g(w)dw \\ &= \int_0^\infty 2\exp\left(\boldsymbol{t}^T(\boldsymbol{A\mu} + \boldsymbol{b}) + \frac{\boldsymbol{At}^T(\boldsymbol{\Omega} + \boldsymbol{DD}^T)\boldsymbol{At}}{2w}\right) \Phi_m\left(\frac{(\boldsymbol{DA})^T\boldsymbol{t}}{w^{1/2}}\right) g(w)dw. \end{aligned}$$

O seguinte resultado mostra que a distribuição t-assimétrica fundamental é fechada por marginalização

Proposição 2.6 *Seja $\boldsymbol{Y} \sim StF(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$ e considerando a seguinte partição de $\boldsymbol{Y}$, $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$ e $\boldsymbol{D}$,*

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1}(n_1 \times 1) \\ \boldsymbol{Y_2}(n_2 \times 1) \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1}(n_1 \times 1) \\ \boldsymbol{\mu_2}(n_2 \times 1) \end{pmatrix},$$

$$\boldsymbol{\Omega} \begin{pmatrix} \boldsymbol{\Omega}_{11}(n_1 \times n_1) & \boldsymbol{\Omega}_{12}(n_2 \times n_1) \\ \boldsymbol{\Omega}_{12}(n_2 \times n_1) & \boldsymbol{\Omega}_{22}(n_2 \times n_2) \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1(n_1 \times m) \\ \boldsymbol{D}_2(n_2 \times m) \end{pmatrix},$$

temos que $\boldsymbol{Y_1} \sim StF_{n_1}(\boldsymbol{\mu_1}, \boldsymbol{\Omega}_{11}, \boldsymbol{D}_1, \nu)$

Prova : Seja $0_{a\times b}$ uma matriz de dimensão $a \times b$ de zeros. Considerando $A = (\ I_{n_1} \quad 0_{n_1 \times n_2}\ )$ na proposição 2.5 obtemos o resultado acima.

2.2 Distribuição t-assimétrica tipo I

Esta distribuição foi proposta por Branco e Dey (2001) e Azzalini e Capitanio (2003), é um caso particular da distribuição t assimétrica fundamental definida em (2.1) quando $\boldsymbol{D} = \boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$, $\boldsymbol{\Sigma} = \boldsymbol{\Omega} + \boldsymbol{\eta\eta}^T$ e $\boldsymbol{\lambda} = \frac{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}}{\sqrt{1-\boldsymbol{\eta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}}}$. Um vetor aleatório Y de dimensão n segue uma distribuição t-assimétrica se sua função de densidade de probabilidade é dada por

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2^m t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \Sigma, \nu) T_1\left(\lambda^T\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})\left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu}\right)^{1/2} \mid \nu + n\right), \tag{2.7}$$

sendo que $\boldsymbol{\mu} \in \Re^n$, $\boldsymbol{\Sigma}$ matriz $n \times n$ definida positiva, $\boldsymbol{\lambda} \in \Re^n$, $\nu > 0$ e $q(\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$. Denotamos por $t_n(. \mid \boldsymbol{a}, \boldsymbol{B}, c)$ a função de densidade de probabilidade de uma t-Student n-variada com parâmetro de posição $\boldsymbol{a}$, escala $\boldsymbol{B}$ e graus de liberdade c, e por $T_n(. \mid \boldsymbol{B}, c)$ a função de distribuição de uma t-Student n-variada com parâmetro de posição $\mathbf{0}$, escala $\boldsymbol{B}$ e graus de liberdade ν. Utilizaremos a seguinte notação $\boldsymbol{Y} \sim \text{St-I}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$

A densidade dada em (2.7) é um caso especial da t-assimétrica fundamental (2.1). Assim, as propriedades desta distribuição podem ser obtidas facilmente a partir dos resultados apresentados na seção 1.

O vetor de médias e matriz de covariância de $\boldsymbol{Y}$, são dados por

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} + \kappa \boldsymbol{\eta} \quad \text{e} \quad Var(\boldsymbol{Y}) = \frac{\nu}{\nu - 2} \boldsymbol{\Sigma} - \kappa^2 \boldsymbol{\eta} \boldsymbol{\eta}^T$$

sendo $\kappa = \sqrt{\frac{\nu}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$.

Podemos escrever a t-assimétrica como uma mistura de uma normal assimétrica tipo I. Assim,

$$\begin{aligned} \boldsymbol{Y} \mid w &\sim \text{SN-I}_n \left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{w}, \boldsymbol{\lambda} \right) \\ w &\sim Gama \left(\frac{\nu}{2}, \frac{\nu}{2} \right), \end{aligned}$$

então $\boldsymbol{Y} \sim \text{St-I}_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu)$.

Se consideramos uma partição de $\boldsymbol{Y}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ e $\boldsymbol{\lambda}$ similar a da proposição 2.6 temos que a distribuição marginal de $\boldsymbol{Y}_1$ é dada por

$$\boldsymbol{Y}_1 \sim \text{St-I}_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_1^*, \nu).$$

sendo $\boldsymbol{\lambda}_1 = \frac{\boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\eta}_1}{\sqrt{1 - \boldsymbol{\eta}_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}_1}}$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)^T = \frac{\boldsymbol{\Sigma}^{1/2} \boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda}}}$

Esta distribuição é fechada por transformações lineares, assim a distribuição de $\boldsymbol{AY} + \boldsymbol{b}$ é

$$\boldsymbol{AY} + \boldsymbol{b} \sim \text{St-I}(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A}^T, \boldsymbol{A\lambda}^*, \nu).$$

sendo $\boldsymbol{\lambda} = \frac{\left(\boldsymbol{A\Sigma A}^T\right)^{-1/2} \boldsymbol{A\eta}}{\sqrt{1 - \boldsymbol{\eta}^T \boldsymbol{A} \left(\boldsymbol{A\Sigma A}^T\right)^{-1} \boldsymbol{A\eta}}}$ e $\boldsymbol{\eta} = \frac{\boldsymbol{\Sigma}^{1/2} \boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda}}}$.

2.2.1 Distribuição Normal assimétrica tipo I

Se $\nu \to \infty$, a densidade em (2.7) coincide com a distribuição normal assimétrica definida em Azzalini e Dalla Valle (1996) e Branco e Dey (2001),

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2^n \phi_n(\boldsymbol{Y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi_1 \left(\lambda^T \boldsymbol{\Sigma}^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}) \right), \tag{2.8}$$

sendo $\phi_n(. \mid \boldsymbol{a}, \boldsymbol{B})$ a função de densidade de probabilidade de uma normal n-variada com vetor de médias $\boldsymbol{a}$, matriz de covariância $\boldsymbol{B}$, e $\Phi_n(. \mid \boldsymbol{B})$ a função de distribuição acumulada de uma normal n-variada com vetor de médias $\mathbf{0}$ e matriz de covariância $\boldsymbol{B}$. Utilizaremos a seguinte notação $\boldsymbol{Y} \sim \text{SN-I}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{D})$

O vetor de médias e matriz de covariância de $\boldsymbol{Y}$, são dados respectivamente por

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} + \kappa\boldsymbol{\eta} \quad \text{e} \quad Var(\boldsymbol{Y}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\eta}\boldsymbol{\eta}^T.$$

2.3 Distribuição t-assimétrica tipo II

Esta distribuição foi proposta por Sahu et al. (2003) e é um caso particular da distribuição assimétrica fundamental como foi visto na seção 2.1. Um vetor aleatório $\boldsymbol{Y}$ de dimensão n segue uma distribuição t-assimétrica do tipo IIse sua função de densidade de probabilidade é dada por

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& 2^m t_n(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}^2, \nu) \\ & T_n\left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu}\right)^{1/2} \mid I_n - \boldsymbol{D}(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}\boldsymbol{D}, \nu + n\right), \end{aligned} \tag{2.9}$$

sendo que $\boldsymbol{\mu} \in \Re^n$, $\boldsymbol{\Omega}$ é uma matriz $n \times n$ definida positiva, $\boldsymbol{D} = Diag(\boldsymbol{\eta})$, $\boldsymbol{\eta} \in \Re^n$, $\nu > 0$ e $q(\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{\mu})^T(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$. Denotamos por $t_n(. \mid \boldsymbol{a}, \boldsymbol{B}, c)$ a função de densidade de probabilidade de uma t-Student n-variada com parametros de posição $\boldsymbol{a}$, de escala $\boldsymbol{B}$ e graus de liberdade c, e por $T_n(. \mid \boldsymbol{B}, c)$ a função de distribuição acumulada de uma t-Student n-variada com parâmetro de posição $\mathbf{0}$, escala $\boldsymbol{B}$ e graus de liberdade c. Utilizaremos a seguinte notação $\boldsymbol{Y} \sim \text{St-II}_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$

Uma forma de representar esta distribuição é através da seguinte representação estocástica. Seja Y um vetor aleatório tal que

$$\boldsymbol{Y} =^d \boldsymbol{\mu} + \boldsymbol{D}\frac{|\boldsymbol{U}|}{w^{1/2}} + \boldsymbol{\Omega}^{1/2}\frac{\boldsymbol{V}}{w^{1/2}}, \tag{2.10}$$

onde $\boldsymbol{U} \sim N_n(0, I_n)$, $\boldsymbol{V} \sim N_n(0, I_n)$ e $w \sim Gama(\nu/2, \nu/2)$. Então,

$$Y \sim \text{St-II}_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu).$$

A densidade dada em (2.9) é um caso especial da t-assimétrica fundamental (2.1). Assim, as propriedades desta distribuição podem ser obtidas facilmente a partir dos resultados apresentados na seção 1.

A média e matriz de covariância de $\boldsymbol{Y}$, são dados por

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} + \kappa\boldsymbol{\eta} \quad \text{e} \quad Var(\boldsymbol{Y}) = \frac{\nu}{\nu - 2}\left(\boldsymbol{\Omega} + \left(1 - \frac{2}{\pi}\right)\boldsymbol{D}^2\right) + \frac{2}{\pi}\left(\frac{\nu}{\nu - 2} - \frac{\pi}{2}\kappa^2\right)\boldsymbol{\eta}\boldsymbol{\eta}^T \tag{2.11}$$

sendo $\kappa = \sqrt{\frac{\nu}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$. Salientamos que a expressão da covariância é uma correção a apresentada por Sahu et al. (2003), sendo que a expressão dada por esses autores para a covariância é

$$Var(\boldsymbol{Y}) = \frac{\nu}{\nu - 2}\left(\boldsymbol{\Omega} + \left(1 - \frac{2}{\pi}\right)\boldsymbol{D}^2\right). \tag{2.12}$$

Notemos que estaria faltando o ultimo termo de (2.11). Assim, o fato de $\boldsymbol{\Omega}$ ser bloco diagonal não implica mais que os subvetores dessa partição sejam não correlacionado como seria o caso se consideraramos a expressão (2.12).

Podemos escrever a distribuição t-assimétrica tipo II como uma mistura de uma normal assimétrica. Assim,

$$\begin{aligned} \boldsymbol{Y} \mid w &\sim \text{SN-II}_n\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Omega}}{w}, \frac{\boldsymbol{D}}{w^{1/2}}\right) \\ w &\sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \end{aligned}$$

então $\boldsymbol{Y} \sim \text{St-II}_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$.

Se consideramos a partição de $\boldsymbol{Y}$, $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$ e $\boldsymbol{D}$ dada na proposição 2.6 temos que a distribuição marginal de $\boldsymbol{Y}_1$ é dada por

$$\boldsymbol{Y}_1 \sim \text{St-II}_{n1}(\boldsymbol{\mu_1}, \boldsymbol{\Omega}_{11}, \boldsymbol{D}_1, \nu).$$

2.3.1 Distribuição Normal assimétrica tipo II

Se $\nu \to \infty$, a densidade em (2.9) coincide com a distribuição normal assimétrica definida em Sahu et al. (2003),

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) =& 2^n \phi_n(\boldsymbol{Y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} + \boldsymbol{D}^2) \\ & \Phi_n\left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \mid \boldsymbol{I}_n - D^T(\boldsymbol{\Omega} + \boldsymbol{D}\boldsymbol{D}^T)^{-1}D\right), \end{aligned} \tag{2.13}$$

sendo $\phi_n(. \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ a função de densidade de probabilidade de uma normal n-variada com vetor de médias $\boldsymbol{\mu}$, matriz de covariância Σ, e $\Phi_n(. \mid \boldsymbol{\Sigma})$ a função de distribuição de uma normal n-variada com vetor de médias $\boldsymbol{\mu} = 0$ e matriz de covariância $\boldsymbol{\Sigma}$. Utilizaremos a seguinte notação $\boldsymbol{Y} \sim \text{SN-II}_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D})$

Notamos que se $\boldsymbol{\Omega} = Diag(\tau_1^2, ..., \tau_n^2)$ então $y_i \sim SN_1(\mu_i, \tau_i^2, \eta_i)$ independentes. Assim a densidade em (2.8) se reduz

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_{i=1}^{n} 2\phi_1(y_i \mid \mu_i, \tau_i^2 + \eta_i^2)\Phi_1\left(\frac{\eta}{\tau_i}\frac{y_i - \mu_i}{\sqrt{\tau_i^2 + \eta_i^2}}\right). \tag{2.14}$$

Considerando a seguinte partição de $\boldsymbol{Y}$, $\boldsymbol{\mu}$, Σ e D,

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1(n_1 \times 1) \\ \boldsymbol{Y}_2(n_2 \times 1) \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1(n_1 \times 1) \\ \boldsymbol{\mu}_2(n_2 \times 1) \end{pmatrix},$$

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11}(n_1 \times n_1) & \boldsymbol{\Omega}_{12}(n_2 \times n_1) \\ \boldsymbol{\Omega}_{21}(n_2 \times n_1) & \boldsymbol{\Omega}_{22}(n_2 \times n_2) \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1(n_1 \times m) \\ \boldsymbol{D}_2(n_2 \times m) \end{pmatrix} e$$

$\boldsymbol{D}_i = Diag(\boldsymbol{\eta}_i)$, $i = 1, 2$ e $\boldsymbol{\eta}^T = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)$. Se $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21} = 0$ então $\boldsymbol{Y}_1$ e $\boldsymbol{Y}_2$ são independentes e $\boldsymbol{Y}_i \sim \text{SN-II}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Omega}_{ii}, \boldsymbol{D}_i, \nu)$, sendo

2.4 Interpretação dos parâmetros

Analisemos primeiro o caso uniparamétrico, seja $y \sim St_1(\mu, \tau^2, \eta, \nu)$ e se consideremos a parametrização $\lambda = \eta/\tau$ e $\sigma^2 = \eta^2 + \tau^2$, então o coeficiente de assimetria é dado por

$$\begin{aligned} \gamma_1 =& \kappa \frac{\lambda}{\sqrt{1+\lambda^2}} \left(\frac{\nu}{\nu-3} \left(3 - \frac{\lambda^2}{1+\lambda^2} \right) - \frac{3\nu}{\nu-2} + 2\kappa^2 \frac{\lambda^2}{1+\lambda^2} \right) \\ & \times \left(\frac{\nu}{\nu-2} - \kappa^2 \frac{\lambda^2}{1+\lambda^2} \right)^{-3/2} \quad (\text{se } \nu > 3) \end{aligned} \tag{2.15}$$

Observamos que γ_1 é uma função crescente em $|\lambda|$ (ver figura 2.1). Notamos que λ fornece informação sob a direção da assimetria desde que para valores positivos de λ temos assimetria positiva, para valores negativos assimetria negativa e para $\lambda = 0$ obtemos a distribuição t-Student. Também fornece informação sob a magnitude da assimetria por γ_1 ser função crescente de λ. Podemos concluir então que o parâmetro λ caracteriza a assimetria da distribuição.

Se consideramos o caso multivariado da t-assimétrica tipo I, na notação da t-assimétrica fundamental, seja $\boldsymbol{Y} \sim StF_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\eta}, \nu)$, sendo $\boldsymbol{\Omega}$ uma matriz $n \times n$ positiva definida dada por

$$\boldsymbol{\Omega} = \begin{pmatrix} \tau_1^2 & \tau_{12} & \dots & \tau_{1n} \\ \tau_{12} & \tau_2^2 & \dots & \tau_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n1} & \tau_{n2} & \dots & \tau_n^2 \end{pmatrix} \tag{2.16}$$

$\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \Re^n$, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n) \in \Re^n$ e $\nu > 0$. Logo, pela proposição 2.6, temos que cada componente do vetor $\boldsymbol{Y} = (y_1, ...y_n)^T$ tem distribuição t-assimétrica univariada, isto é, $y_i \sim St_1(\mu_i, \tau_i^2, \eta_i, \nu)$.

Logo, podemos definir o vetor $\boldsymbol{\lambda}^{(*)} = (\lambda_1, ..., \lambda_n)^T$, sendo que $\lambda_i = \eta_i/\tau_i$ nos fornece informação sob assimetria da distribuição marginal de y_i, $i = 1, \dots, n$. Notemos que $\boldsymbol{\lambda}^{(*)}$ pode ser escrito na notação matricial como $\boldsymbol{\lambda}^{(*)} = Diag(\boldsymbol{\Omega})^{-1/2}\boldsymbol{\eta}$,

Agora estudaremos as diferentes parametrizações da normal assimétrica tipo I e analisaremos se elas em realidade fornecem informação sob a assimetria das marginais. Na parametrização considerada por Branco e Dey (2001) a quantidade denominada parâmetro de assimetria é dada por

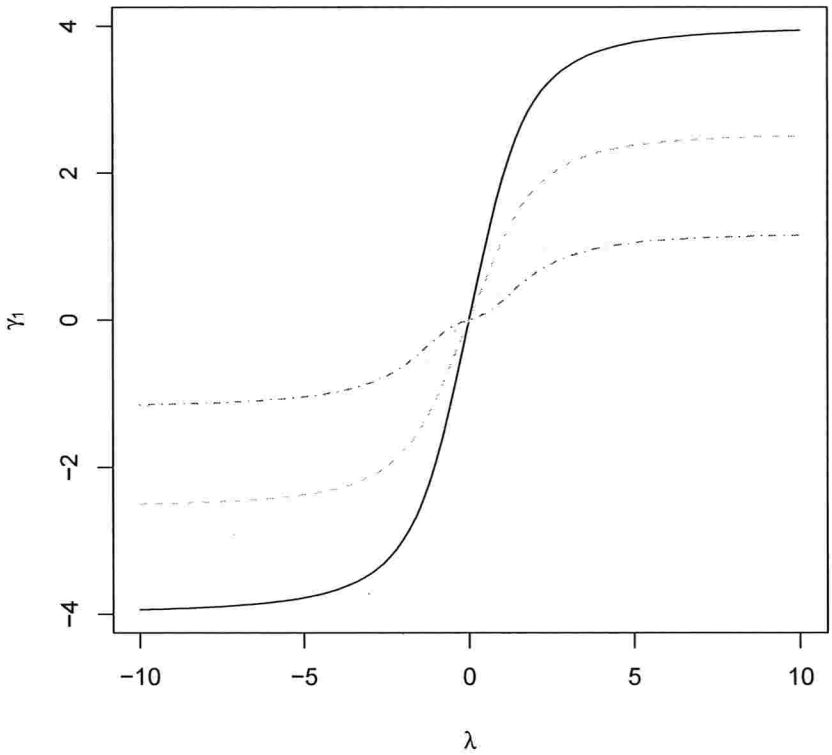

Figura 2.1: Coeficiente de assimetria γ_1 como uma função de λ para graus de liberdade $\nu = 4$(linha solida), 5(linha tracejada), 10(linha pontilhada) e 20(linha ponto traço)

$$\boldsymbol{\lambda}^{(1)} = \frac{\Sigma^{-1}\boldsymbol{\eta}}{\left(1 - \boldsymbol{\eta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\right)^{1/2}}, \tag{2.17}$$

sendo $\boldsymbol{\Sigma} = \boldsymbol{\Omega} + \boldsymbol{\eta}\boldsymbol{\eta}^T$. Por outro lado, na parametrização considerada por Azzalini e Capitanio (1999) essa quantidade é

$$\boldsymbol{\lambda}^{(2)} = \frac{Diag(\boldsymbol{\Sigma})^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}}{\left(1 - \boldsymbol{\eta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\right)^{1/2}}, \tag{2.18}$$

Podemos reescrever $\boldsymbol{\lambda}^{(1)}$ e $\boldsymbol{\lambda}^{(2)}$ em função de $\boldsymbol{\lambda}^{(*)}$. Seja $c = 1/\left(1 - \eta^T\Sigma^{-1}\eta\right)^{1/2}$ então

$$\begin{aligned} \boldsymbol{\lambda}^{(1)} &= c\Sigma^{-1}Diag(\boldsymbol{\Omega})^{1/2}\boldsymbol{\lambda}^{(*)} \text{ e} \\ \boldsymbol{\lambda}^{(2)} &= cDiag(\boldsymbol{\Sigma})^{1/2}\Sigma^{-1}Diag(\boldsymbol{\Omega})^{1/2}\boldsymbol{\lambda}^{(*)}. \end{aligned} \tag{2.19}$$

Notamos que embora $\boldsymbol{\lambda}^{(*)}$ forneça informação sobre a assimetria de cada marginal, isso não ocorre com $\boldsymbol{\lambda}^{(1)}$ e $\boldsymbol{\lambda}^{(2)}$. Vejamos o seguinte exemplo, seja $\nu = 4$,

$$\boldsymbol{\Omega} = \begin{pmatrix} 8 & -1 & 3 \\ -1 & 5 & 4 \\ 3 & 4 & 9 \end{pmatrix} \text{ e } \boldsymbol{\eta} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}. \tag{2.20}$$

Se calculamos os coeficientes de assimetria unidimensionais definidos em (2.1) obtemos

$$\gamma_1 = \begin{pmatrix} 0.74 \\ -1.75 \\ 1.92 \end{pmatrix}, \tag{2.21}$$

isto representa, que y_2 tem assimetria negativa, y_1 e $y_3$tem assimetria positiva e que y_3 apresenta a maior assimetria do vetor.
Então,

$$\boldsymbol{\lambda}^{(*)} = \begin{pmatrix} 0.35 \\ -0.89 \\ 1.00 \end{pmatrix}, \boldsymbol{\lambda}^{(1)} = \begin{pmatrix} -0.18 \\ -0.53 \\ 0.43 \end{pmatrix}, \boldsymbol{\lambda}^{(2)} = \begin{pmatrix} -0.53 \\ -1.60 \\ 1.81 \end{pmatrix} \tag{2.22}$$

Note que que $\boldsymbol{\lambda}^{(1)}$ e $\boldsymbol{\lambda}^{(2)}$ mudam o sinal da assimetria de y_1. Além disso, $\boldsymbol{\lambda}^{(1)}$ indica que a maior assimetria é apresentada por y_2.
Azzalini e Arellano-Valle (2008) apresentaram observação semelhante para o caso da normal assimétrica tipo I.

Consideramos agora a distribuição t-assimétrica multivariada de Sahu et al. (2003). Seja $\boldsymbol{Y} \sim$ St-II$_n(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu)$, com $\boldsymbol{\Omega}$ matriz definida positiva dada como em (2.16), $\boldsymbol{D} = Diag(\boldsymbol{\eta})$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n) \in \Re^n$ e $\nu > 0$.
Então cada componente do vetor $\boldsymbol{Y} = (y_1, ...y_n)^T$ tem distribuição t-assimétrica univariada, isto é, $y_i \sim St_1(\mu_i, \tau_i^2, \eta_i, \nu)$ logo como no caso da t-assimétrica tipo I podemos definir o vetor $\boldsymbol{\lambda}^{(*)} = (\lambda_1, ..., \lambda_n)^T$, sendo que $\lambda_i = \eta_i/\tau_i$ nos fornece informação sob assimetria da distribuição marginal de y_i, $i = 1, \ldots, n$. Assim, consideremos novamente o exemplo anterior, neste caso temos

$$\boldsymbol{\eta} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \boldsymbol{\lambda}^{(*)} = \begin{pmatrix} 0.70 \\ -1.41 \\ 1.00 \end{pmatrix} \text{ e } \gamma_1 = \begin{pmatrix} 0.74 \\ -1.76 \\ 1.92 \end{pmatrix}. \tag{2.23}$$

Como podemos observar a parametrização η preserva o sinal mas não a magnitude da coeficiente de assimetria. Observemos que o coeficiente de assimetria de y_3 é maior em valor absoluto do que as outras variáveis, mas a parametrização η indica que y_2 é mais assimétrico que y_1 e y_2.
Logo, temos que $\boldsymbol{\lambda} = Diag(\boldsymbol{\Omega})^{-1/2}\boldsymbol{\eta}$ preserva o sinal e a magnitude do coeficiente de assimetria de cada marginal tanto para a t-assimétrica tipo I e II. Assim, neste trabalho consideraremos $\boldsymbol{\lambda}^{(*)}$ para a interpretação da assimetria no modelo.

Capítulo 3

Modelo de Regressão Multivariada

3.1 Introdução

Neste capítulo consideraremos o modelo de regressão multivariada sob o modelo t-assimétrico proposto por Sahu et al. (2003). Trataremos o modelo sob o enfoque bayesiano, utilizando as propriedades da distribuição t-assimpetrica (ver capítulo 2), apresentaremos a sua forma hierárquica, a distribuição *a posteriori* aumentada e as distribuições condicionais completas, as quais utilizaremos para implementar o algoritmo de Gibbs e obter inferência *a posteriori*.

Além disso, desenvolveremos um estudo de simulação sob diferentes distribuições *a priori* para os parâmetros de assimetria. Todos os cálculos serão baseados numa amostra de Monte Carlo obtida via amostrador de Gibbs, o qual foi implementado no pacote R e cujo programa encontra-se no apêndice.

3.2 Modelo de Regressão Multivariada com a distribuição t-assimétrica

Suponha que temos m observações independentes $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_m$, sendo que

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{3.1}$$

para $i = 1, ..., m$, sendo que $\boldsymbol{Y}_i$ é um vetor n-dimensional, $\boldsymbol{X}_i$ uma matriz $n \times p$ que contem valores de p variáveis explicativas, $\boldsymbol{\beta} \in \Re^p$ o vetor de parâmetros de regressão e $\boldsymbol{\epsilon}_i$ é um vetor n_i-dimensional de erros. Assumiremos que cada $\boldsymbol{\epsilon}_i$ segue uma distribuição t-assimétrica tipo II

$$\boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} \text{St-II}_n(-\kappa\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu) \tag{3.2}$$

onde $\boldsymbol{\eta} \in \Re^n$, $\boldsymbol{D} = Diag(\boldsymbol{\eta})$, $\nu > 0$ e $\kappa = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$ de modo que $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $i = 1, \ldots, n$.

A função de verossimilhança do modelo definido em (3.1)-(3.2), denotada por $l(\boldsymbol{\theta})$ é dada por

$$l(\boldsymbol{\theta}) = \prod_{i=1}^{m} 2^n t_n(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \kappa\eta \mid 0, \boldsymbol{\Omega} + \boldsymbol{D}^2, \nu)$$
$$T_n\left(\boldsymbol{D}^T(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \kappa\eta)\left(\frac{\nu + n}{q(\boldsymbol{y}) + \nu}\right)^{1/2} \mid I_n - \boldsymbol{D}(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}\boldsymbol{D}, \nu + n\right), \tag{3.3}$$

Casos particulares do modelo linear t-assimétrico definido por (3.1)-(3.2) são

1. Se $\nu \to \infty$ temos o modelo linear normal assimétrico,
2. Se $\boldsymbol{\eta} = \mathbf{0}$ temos o modelo linear t-Student independente,
3. Se $\boldsymbol{\eta} = \mathbf{0}$ e $\nu \to \infty$ temos o modelo linear normal.

Para facilitar a obtenção das distribuições condicionais completas consideramos que o modelo em (3.1) tem intercepto, isto é

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{X}_i^F \end{pmatrix}$$

onde $\boldsymbol{X}_i^F$ é uma matriz de planejamento de dimensão $n \times p - n$ e que

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}$$

sendo que $\boldsymbol{\beta}_0 \in \Re^n$ e $\boldsymbol{\beta}_1 \in \Re^{n-p}$. Logo, podemos reescrever o modelo em (3.1) como

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_i, \tag{3.4}$$

sendo $\boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\beta}_0^{*T} & \boldsymbol{\beta}_1^T \end{pmatrix}^T$ e $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0 - \kappa\boldsymbol{\eta}$ e

$$\boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} \text{St-II}_n(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{D}, \nu). \tag{3.5}$$

onde $\boldsymbol{\eta} \in \Re^n$, $\boldsymbol{D} = Diag(\boldsymbol{\eta})$, $\nu > 0$ e $\kappa = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$ de modo que $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $i = 1, \ldots, n$.

3.3 Distribuições *a Priori*

Para a distribuição normal assimétrica univariada sob a parametrização de Sahu et al. (2003), Bayes e Branco (2007) consideraram $\eta \sim t_1(0, c\tau^2, d)$ e $p(\tau^2) \propto 1/\tau^2$, quando $d = 1/2$ e $c = \pi^2/4$ obtemos uma aproximação da distribuição *a priori* de Jeffreys, e quando $d = 2$ e $c = 1/2$ obtemos *a priori* induzida por assumir uma distribuição uniforme em $\delta = \lambda/\sqrt{1+\lambda^2} \sim U(-1, 1)$ sendo $\lambda = \eta/\tau$.

Para o caso multivariado consideramos uma versão multivariada destas distribuições *priori* considerando a distribuição t-Student multivariada e a distribuição Wishart Inversa. Notemos que

estamos considerando uma distribuição *a priori* própria para $\boldsymbol{\Omega}$,

$$\begin{aligned} \boldsymbol{\eta} &\sim t_n\left(0, c\boldsymbol{\Omega}, d\right) \\ \boldsymbol{\Omega} &\sim Inv - Wishart\left(a, B\right). \end{aligned} \tag{3.6}$$

A distribuição *a priori* para o parâmetro graus de liberdade ν é dada por

$$p(\nu) \propto \nu^{-2} I(\nu > 3), \tag{3.7}$$

como em Liu (1995) e Rosa, Padovani e Gianola (2003). O ponto de truncamento foi escolhido para garantir que a média e a variância da distribuição t-assimétrica sejam finitas.

Para $\boldsymbol{\beta}^*$ assumimos uma distribuição normal multivariada

$$\boldsymbol{\beta}^* \sim N_p\left(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}\right). \tag{3.8}$$

3.4 O Modelo Hierárquico e o Algoritmo MCMC

Uma forma hierárquica de reescrever o modelo dado em (3.4)-(3.5) é utilizando a representação estocástica dada em (2.5), desta forma temos que

$$\begin{aligned} \boldsymbol{Y}_i \mid \boldsymbol{u}_i, w_i &\sim N_n\left(\boldsymbol{X}_i\boldsymbol{\beta}^* + Diag(\boldsymbol{\eta})\boldsymbol{u}_i, \frac{\boldsymbol{\Omega}}{w_i}\right) \\ \boldsymbol{u}_i &\sim N\left(\boldsymbol{0}, \frac{\boldsymbol{I}_n}{w_i}\right) I(\boldsymbol{u}_i > \boldsymbol{0}) \\ w_i &\sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \tag{3.9}$$

A distribuição *a posteriori* aumentada é o produto da função de verossimilhança aumentada dada em (3.9) e a distribuição *a priori* definida em (3.6)-(3.8). Assim, podemos utilizar algoritmos do tipo Markov Chain Monte Carlo (MCMC) para obter amostras da distribuição *a posteriori*. Para implementar o amostrador de Gibbs precisamos das distribuições condicionais completas dadas na proposição 3.1

Proposição 3.1 *Sob o modelo (3.9) e a distribuição a priori dada em(3.6)-(3.8), e reescrevendo a distribuição a priori de $\boldsymbol{\eta}$ dada em (3.6) como uma mistura na escala de uma normal*

$$\begin{aligned} \boldsymbol{\eta} &\sim N_q\left(0, c\frac{\boldsymbol{\Omega}}{v}\right) \\ v &\sim Gamma\left(d/2, d/2\right). \end{aligned} \tag{3.10}$$

As distribuições condicionais completas são dadas por

(i) $\boldsymbol{\beta}^* \mid . \sim N_p(\boldsymbol{m}_{\boldsymbol{\beta}}, \boldsymbol{S}_{\boldsymbol{\beta}})$ *onde* $\boldsymbol{m}_{\boldsymbol{\beta}} = \boldsymbol{S}_{\boldsymbol{\beta}}^{-1}\left(\sum_{i=1}^m w_i \boldsymbol{X}_i^T \boldsymbol{\Omega}_b^{-1}(\boldsymbol{Y}_i - Diag(\boldsymbol{\eta})\boldsymbol{u}_i) + \boldsymbol{\Omega}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)$ *e*
$\boldsymbol{S}_{\boldsymbol{\beta}}^{-1} = w_i \sum_{i=1}^m \boldsymbol{X}_i^T \boldsymbol{\Omega}^{-1}\boldsymbol{X}_i + \boldsymbol{\Omega}_{\boldsymbol{\beta}}^{-1}$;

(ii) $\boldsymbol{\eta}_b \mid . \sim N_q\ (\boldsymbol{m}_{\boldsymbol{\eta}}, \boldsymbol{S}_{\boldsymbol{\eta}})$ *onde* $\boldsymbol{m}_{\boldsymbol{\eta}} = \boldsymbol{S}_{\boldsymbol{\eta}}^{-1}\left(\sum_{i=1}^m w_i Diag(\boldsymbol{u}_{bi})\boldsymbol{\Omega}^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^*)\right)$ *e*
$\boldsymbol{S}_{\boldsymbol{\eta}}^{-1} = \left(\sum_{i=1}^m w_i Diag(\boldsymbol{u}_i)\boldsymbol{\Omega}^{-1}Diag(\boldsymbol{u}_i) + \frac{v}{c}\boldsymbol{\Omega}^{-1}\right)$;

(iii) $\boldsymbol{\Omega} \mid . \sim IW\left(n + a + 1, \sum_{i=1}^m w_i\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^* - Diag(\boldsymbol{\eta})\boldsymbol{u}_i\right)\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^* - Diag(\boldsymbol{\eta})\boldsymbol{u}_i\right)^T + \frac{v_b}{c}\boldsymbol{\eta}\boldsymbol{\eta}^T + B\right)$;

(iv) $v_b \mid . \sim Gamma\left(\frac{d+n}{2}, \frac{1}{2}\left(\frac{\boldsymbol{\eta}\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}}{c} + d\right)\right)$;

(v) $\boldsymbol{u}_i \mid . \sim N_1\left(\boldsymbol{m}_{\boldsymbol{u}_i}, \boldsymbol{S}_{\boldsymbol{u}_i}\right) I(\boldsymbol{u}_i > 0)$ *onde* $\boldsymbol{m}_{\boldsymbol{u}_i} = \boldsymbol{S}_{\boldsymbol{u}_i}^{-1}Diag(\boldsymbol{\eta})\boldsymbol{\Omega}^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^*)$ *e*
$\boldsymbol{S}_{\boldsymbol{u}_i} = \frac{1}{w_i}(\boldsymbol{I}_q + Diag(\boldsymbol{\eta})\boldsymbol{\Omega}^{-1}Diag(\boldsymbol{\eta}))^{-1}$;

(vi) $w_i \mid . \sim Gamma\left(a_{wi}, b_{wi}\right)$ *onde* $a_{wi} = \frac{\nu+n}{2}$ *e*
$b_{wi} = \frac{1}{2}\left(\boldsymbol{Y}_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - Diag(\boldsymbol{\eta})\boldsymbol{u}_i\right)^T \boldsymbol{\Omega}^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - Diag(\boldsymbol{\eta})\boldsymbol{u}_i\right) + \frac{1}{2}\nu$ *e*

(vii) $f(\nu \mid .) \propto \frac{1}{\Gamma(\nu/2)^m}\nu^{-2}\left(\frac{\nu}{2}\right)^{m\frac{\nu}{2}} \exp\left(-\sum_{i=1}^m (w_i - \log w_i)\frac{\nu}{2}\right) I(\nu > 3)$.

Observamos que todas as distribuições condicionais *a posteriori* tem formas conhecidas com exceção dos graus de liberdade. Para gerar amostras da distribuição condicional dos graus de liberdade utilizamos o algoritmo de rejeição adaptativa (Gilks e Wild, 1992), desde que a distribuição é log-concava (ver apêndice B). O amostrador de Gibbs também pode ser facilmente implementado utilizando o pacote WinBUGS (Spiegelhalter, Thomas, Best e Lunn, 2004), basta utilizar a representação hierárquica dada em 3.9.

3.5 Comparação de Modelos

Existem várias metodologias para comparação de modelos e não existe um consenso sob o melhor enfoque para seleção de modelos em inferência bayesiana (Gelman, Carlin e Stern, 2003, capítulo 6). Neste trabalho, consideramos o EAIC (Expected Akaike Information Criterion), EBIC(Expected Bayesian Information Criterion) propostos por Brooks (2002) e o DIC (Deviance Information Criterion) proposto por Spiegelhalter, Best, Carlin e Van der Linde (2002). Essas medidas são fáceis de calcular diretamente das amostras obtidas pelo algoritmo de Gibbs.

Esses critérios são baseados na função desvio $D(\theta)$ sendo

$$D(\theta) = -2\sum_{i=1}^m log f(\boldsymbol{Y}_i \mid \theta).$$

Seja $\theta^1, ..., \theta^B$ uma amostra da distribuição *a posteriori*, p o número de parâmetros,

$$\bar{D} = \frac{1}{B}\sum_{b=1}^B D(\theta^b),\ \ \hat{D} = D(\bar{\theta})\ \text{ e }\ \rho_d = \bar{D} - \hat{D}$$

então

$$\begin{aligned} EAIC &= \bar{D} + 2p, \\ EBIC &= \bar{D} + plog(m), \\ DIC &= \bar{D} + \rho_d. \end{aligned} \tag{3.11}$$

Valores pequenos destes critérios indicam um melhor ajuste.
Outra alternativa de comparação de modelos é utilizar as densidades de validação cruzada (Conditional Predictive Ordinate), isto é, o CPO é a densidade preditiva de uma observação condicionada ao restante dos dados. Para maiores detalhes ver o apêndice A. Uma estimativa de Monte Carlo do CPO da observação $\boldsymbol{Y}_i$ é dada por

$$\widehat{CPO}_i = \left[\frac{1}{B} \sum_{b=1}^{B} \frac{1}{f(\boldsymbol{Y}_i \mid \theta^b)} \right]. \tag{3.12}$$

Como medida de comparação de modelos podemos utilizar o produto dos CPO, isto é, $\prod \widehat{CPO}_i$. Escolhe-se o modelo que corresponde ao maior valor dessa medida. Para comprar dois modelos M1 e M2, podemos calcular o pseudo fator de Bayes (pBF) que é $pBF = \prod \widehat{CPO}_i^{M1} / \prod \widehat{CPO}_i^{M2}$, valores altos do pBF favorecem ao modelo M1.

3.6 Estudo de Simulação

Neste estudo de simulação avaliaremos o desempenho das medias de comparação em detectar a distribuição dos dados, os dados foram gerado do seguinte modelo

$$\boldsymbol{Y} \sim St_2(\boldsymbol{\mu}_0, \boldsymbol{S}_0, \boldsymbol{D}_0, v_0), \tag{3.13}$$

consideramos os casos: Normal, Normal assimétrico, t-Student e t-assimétrico com tamanho de amostra $n = 50, 100$ e 200, geramos 100 replicas e para cada conjunto de dados ajustamos os quatro modelo e calculamos as medidas de comparação. O amostrador de Gibbs foi implementado no pacote R baseados nas distribuições condicionais completas dadas na proposição 3.1, a implementação não foi dificil desde que estas tem formas conhecidas com exceção dos graus de liberdade ν nesse caso consideramos o algoritmo de rejeição adaptativa (ver apêndice B). Para todas as simulações foram geradas 120000 amostras com *burn-in* de 20000. Para reduzir a autocorrelação entre as cadeias, tomamos valores de 10 em 10 para obtermos uma amostra de Monte Carlo aproximadamente independente de tamanho 10000.

Para $\boldsymbol{\mu}_0 = \boldsymbol{0}$, $\boldsymbol{S}_0 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$,$\boldsymbol{\eta}_0 = (2,3)^T$ e $\nu_0 = 4$ os resultados são apresentados na tabela 3.1. Notamos que os o EAIC, EBIC apresentam os melhores resultados, o PBF e o DIC podem escolher modelos mais complexos algumas vezes, mas o DIC no caso normal quase sempre escolhe o modelo normal assimétrico. Podemos concluir que o EAIC e o EBIC conseguem detectar melhor o modelo de onde foram gerados os dados e em menor medida o pBF também, mas o DIC tende

a favorecer modelos complexos e não será considerado neste trabalho. Resultados similares foram encontrados para $n = 100$ e 200.

Tabela 3.1: Estudo de Simulação: Porcentagem de vezes o critérios escolhem os modelos das colunas sob diferentes cenários

Cenário: Normal				
Modelo	N	SN	t	St
EAIC	93	7	0	0
EBIC	99	1	0	0
DIC	2	98	0	0
pBF	76	15	5	4
Cenário: Normal assimétrico				
Modelo	N	SN	t	St
EAIC	1	99	0	0
EBIC	6	94	0	0
DIC	0	74	0	26
pBF	0	85	0	15
Cénario t				
Modelo	N	SN	t	St
EAIC	1	0	99	0
EBIC	2	0	98	0
DIC	0	1	73	27
pBF	0	0	87	13
Cenário: t-assimétrico				
Modelo	N	SN	t	St
EAIC	0	1	0	99
EBIC	0	0	0	100
DIC	0	0	0	100
pBF	0	2	0	98

Capítulo 4

Modelo Linear Misto t-assimétrico Dependente

4.1 Introdução

Modelos lineares mistos são muito utilizados para a análise de medidas repetidas e de dados agrupados, sendo aplicado em áreas tão diversas como, por exemplo, a agricultura, biologia e economia.

O modelo linear misto mais utilizado na literatura foi proposto por Laird e Ware (1982) e é dado por

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{4.1}$$

para $i = 1, ..., m$, em que $\boldsymbol{Y}_i$ é um vetor n_i-dimensional de valores observados para a i-ésima unidade, $\boldsymbol{X}_i$ é uma matriz $n_i \times p$ que contém valores de p variáveis explicativas, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos, $\boldsymbol{Z}_i$ é uma matriz $n_i \times q$ que contém a especificação dos efeitos aleatórios $\boldsymbol{b}_i$ e $\boldsymbol{\epsilon}_i$ é um vetor n_i-dimensional de erros. É comum assumir normalidade para os erros $\boldsymbol{\epsilon}_i$ e para os efeitos aleatórios $\boldsymbol{b}_i$.

A suposição de normalidade pode ser muito restritiva para uma representação adequada da estrutura presente em conjuntos de dados com medidas repetidas ou agrupados. Do ponto de vista prático, o método mais adotado para alcançar a normalidade é a transformação de variáveis. Embora tal metodologia possa dar resultados empíricos razoáveis, este deve ser evitado se um modelo teórico mais conveniente for encontrado (Azzalini e Capitanio, 1999). Assim, interesse considerável é dado na literatura para flexibilizar a suposição de normalidade. Por exemplo, Zhang e Davidian (2001) consideram um modelo semiparamétrico, Pinheiro, Liu e Wu (2001) propõem a distribuição t-multivariada, Ma, Genton e Davidian (2004) consideram a distribuição flexível elíptica assimétrica (Ma e Genton, 2004) para os efeitos aleatórios e Arellano-Valle, Bolfarine e Lachos (2005) consideram a distribuição normal assimétrica multivariada.

Uma distribuição que permite incorporar ao modelo parâmetros que quantificam a assimetria e curtose é a t-assimétrica multivariada que foi proposta por Sahu et al. (2003) que é um caso particular da distribuição assimétrica fundamental apresentada no capítulo 2. Assim, neste trabalho consideramos esta distribuição tanto para os erros, quanto para os efeitos aleatórios.

4.2 Modelo Linear Misto com a distribuição t-assimétrica

Uma suposição usual em modelos lineares mistos é considerar $\boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{R}_i)$ e $\boldsymbol{b}_i \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$, logo podemos escrever

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim N_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{R}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}\right) \tag{4.2}$$

Se consideramos $\boldsymbol{R}_i = \tau_e^2 \boldsymbol{I}_{n_i}$ o modelo em (4.2) se reduz a

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim N_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \tau_e^2 \boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}\right). \tag{4.3}$$

Notemos que os erros $\boldsymbol{\epsilon}_i$ e os efeitos aleatórios $\boldsymbol{b}_i$ são independentes entre si. Além disso, os erros entre sujeitos também são independentes portanto não correlacionados.

Para obter um modelo mais robusto a presença de valores *outlier*, Lange, Little e Taylor (1989) e Pinheiro et al. (2001) consideram que $\boldsymbol{\epsilon}_i$ e $\boldsymbol{b}_i$ tem distribuição t-Student multivariada.

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim t_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \tau_e^2 \boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}, \nu\right). \tag{4.4}$$

Notemos que neste modelo $\boldsymbol{\epsilon}_i$ e $\boldsymbol{b}_i$ não são mais independentes mas eles mantém a propriedade de ser não correlacionados, isto é, $cov[\boldsymbol{\epsilon}_i, \boldsymbol{b}_i] = \mathbf{0}$ e os erros entre sujeitos também são não correlacionados, $cov[\epsilon_{ij}, \epsilon_{ij'}] = 0$, $j \neq j'$. Além disso, temos que a $E[\boldsymbol{\epsilon}_i] = \mathbf{0}$ e a $E[\boldsymbol{b}_i] = \mathbf{0}$.

Uma possibilidade para estender o modelo dado em (4.3) considerando a distribuição t-assimétrica e preservando a suposição que $E[\boldsymbol{\epsilon}_i] = \mathbf{0}$ e a $E[\boldsymbol{b}_i] = \mathbf{0}$ é dada por

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim \text{St-II}_{n_i+q}\left(\begin{pmatrix} -\kappa_e \eta_e \mathbf{1}_{n_i} \\ -\kappa \boldsymbol{\eta_b} \end{pmatrix}, \begin{pmatrix} \tau_e^2 \boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}, \begin{pmatrix} \eta_e \boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_b \end{pmatrix}, \nu\right), \tag{4.5}$$

sendo $\boldsymbol{D}_b = Diag(\boldsymbol{\eta}_b)$, $\boldsymbol{\eta}_b \in \Re^q$. Neste caso, temos que $cov\left[\boldsymbol{\epsilon}_i, \boldsymbol{b}_i\right] = \frac{2}{\pi}\left(\frac{\nu}{\nu-2} - \kappa^2\right)\eta_e \mathbf{1}_{n_i} \boldsymbol{\eta}_b^T$ e $cov[\epsilon_{ij}, \epsilon_{ij'}] = \frac{2}{\pi}\left(\frac{\nu}{\nu-2} - \kappa^2\right)\eta_e^2$, $j \neq j'$. Assim, a covariância entre o erro e o efeito aleatório e a covariância entre sujeitos dependem dos parâmetros de assimetria do modelo.

Logo, para manter as suposições de não correlação entre erro $\boldsymbol{\epsilon}_i$ e o efeito aleatório $\boldsymbol{b}_i$ e de não correlação dos erros entre sujeitos, consideraremos $\eta_e = 0$. Assim, a assimetria é introduzida somente nos efeitos aleatórios e os erros são considerado simétricos. No entanto, como será mostrado a seguir, a distribuição marginal de $\boldsymbol{Y}_i$ será assimétrica.

Assim, o modelo a ser estudado nesta seção, o qual denominado modelo linear misto t-assimétrico dependente será

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim \text{St-II}_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ -\kappa\boldsymbol{\eta}_b \end{pmatrix}, \begin{pmatrix} \tau_e^2\boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}, \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_b \end{pmatrix}, \nu\right), \tag{4.6}$$

sendo $\boldsymbol{D}_b = Diag(\boldsymbol{\eta}_b)$.
Podemos reescrever (4.6) como uma mistura de uma normal assimétrica

$$\begin{aligned} &\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \mid w_i \sim \text{SN-II}_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ -\kappa\boldsymbol{\eta}_b \end{pmatrix}, \frac{1}{w_i}\begin{pmatrix} \tau_e^2\boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}, \frac{1}{w_i^{1/2}}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_b \end{pmatrix}, \nu\right) \\ &w_i \sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \tag{4.7}$$

Como a matriz de escala é bloco diagonal, temos que, condicional a w_i, $\boldsymbol{\epsilon}_i$ e $\boldsymbol{b}_i$ são independentes, e que os ϵ_{ij} são independentes e identicamente distribuídos. Além disso, marginalmente

$$\boldsymbol{\epsilon}_i \sim t_{n_i}\left(\mathbf{0}, \tau_e^2\boldsymbol{I}_{n_i}, \nu\right) \text{ e } \boldsymbol{b}_i \sim \text{St-II}_q\left(-\kappa\boldsymbol{\eta}_b, \boldsymbol{\Psi}, \boldsymbol{D}_b, \nu\right) \tag{4.8}$$

desde que para cada $i = 1, ...m$, $\boldsymbol{\epsilon}_i$ e $\boldsymbol{b}_i$ são afetados por w_i em geral eles não são independentes. A independência corresponde ao caso em que $w_i = 1\ \forall i$, isto é, o modelo t-assimétrico se reduz a modelo normal assimétrico.
Logo, se aplicamos a proposição 2.5 obtemos distribuição conjunta de $\boldsymbol{Y}_i$ e $\boldsymbol{b}_i$ que é dada por

$$\begin{pmatrix} \boldsymbol{Y}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim \text{St-II}_{n_i+q}\left(\begin{pmatrix} \boldsymbol{X}_i\beta - \kappa\boldsymbol{Z}_i\boldsymbol{\eta}_b \\ -\kappa\boldsymbol{\eta}_b \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i + \tau_e^2\boldsymbol{I}_{n_i} & \boldsymbol{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\boldsymbol{Z}_i^T & \boldsymbol{\Psi} \end{pmatrix}, \begin{pmatrix} \mathbf{0} & \boldsymbol{Z}_i\boldsymbol{D}_b \\ \mathbf{0} & \boldsymbol{D}_b \end{pmatrix}, \nu\right). \tag{4.9}$$

Utilizando a proposição 2.6, a distribuição marginal de $\boldsymbol{Y}_i$ é dada por

$$\boldsymbol{Y}_i \sim StF_{n_i}\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{D}_i, \nu\right). \tag{4.10}$$

sendo

$$\begin{aligned} \boldsymbol{\mu}_i &= \boldsymbol{X}_i\beta - \kappa\boldsymbol{Z}_i\boldsymbol{\eta}_b \\ \boldsymbol{\Sigma}_i &= \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T + \tau_e^2\boldsymbol{I}_{n_i} \text{ e} \\ \boldsymbol{D}_i &= \begin{pmatrix} \mathbf{0} & \boldsymbol{Z}_i\boldsymbol{D}_b \end{pmatrix}. \end{aligned} \tag{4.11}$$

A densidade da distribuição marginal de $\boldsymbol{Y}_i$

$$\begin{aligned} &f(\boldsymbol{Y}_i \mid \theta) = 2^{n_i+q}t_n(\boldsymbol{Y}_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i + \boldsymbol{D}_i\boldsymbol{D}_i^T, \nu) \\ &T_{n_i+q}\left(\boldsymbol{D}_i^T(\boldsymbol{\Sigma}_i + \boldsymbol{D}_i\boldsymbol{D}_i^T)^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)\left(\frac{\nu+n}{q(\boldsymbol{Y}_i)+\nu}\right)^{1/2} \mid \boldsymbol{I}_{ni+q} - \boldsymbol{D}_i^T(\boldsymbol{\Sigma}_i + \boldsymbol{D}_i\boldsymbol{D}_i^T)^{-1}\boldsymbol{D}_i, \nu+n\right), \end{aligned} \tag{4.12}$$

Utilizando a proposição 2.4 obtemos a esperança e a variância de $\boldsymbol{Y}_i$ dadas pelas seguintes expressões,

$$E[\boldsymbol{Y}_i] = \boldsymbol{X}_i\boldsymbol{\beta} \tag{4.13}$$

e

$$Var[\boldsymbol{Y}_i] = \boldsymbol{Z}_i Var[\boldsymbol{b}_i]\boldsymbol{Z}_i + Var[\boldsymbol{\epsilon}_i], \tag{4.14}$$

sendo

$$Var[\boldsymbol{b}_i] = \frac{\nu}{\nu-2}\left(\boldsymbol{\Psi} + \left(1-\frac{2}{\pi}\right)\boldsymbol{D}_b^2\right) + \frac{2}{\pi}\left(\frac{\nu}{\nu-2} - \frac{\pi}{2}\kappa^2\right)\boldsymbol{\eta}_b\boldsymbol{\eta}_b^T$$

e

$$Var[\boldsymbol{\epsilon}_i] = \frac{\nu}{\nu-2}\tau_e^2\boldsymbol{I}_{n_i}.$$

Casos particulares do modelo linear misto t-assimétrico definido em (4.6) são

1. Se $\nu \to \infty$ temos o MLM normal assimétrico dado por Arellano-Valle, Bolfarine e Lachos (2007),
2. Se $\boldsymbol{\eta}_b = \boldsymbol{0}$ temos o MLM t-Student estudado em Pinheiro et al. (2001),
3. Se $\boldsymbol{\eta}_b = \boldsymbol{0}$ e $\nu \to \infty$ temos o MLM Normal.

Para facilitar a obtenção das distribuições condicionais completas consideraremos, o caso mais utilizado do modelo linear misto, quando as colunas de $\boldsymbol{Z}_i$ estão contidas em $\boldsymbol{X}_i$. Assim, sejam as partições

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_i^G & \boldsymbol{X}_i^F \end{pmatrix} \text{ e } \boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{X}_i^G \end{pmatrix} \tag{4.15}$$

onde $\boldsymbol{X}_i^G$ é uma matriz de planejamento $n_i \times q$, $\boldsymbol{X}_i^F$ é uma matriz de planejamento $n_i \times (p-q)$ e consideramos a seguinte partição de $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^G \\ \boldsymbol{\beta}^F \end{pmatrix} \tag{4.16}$$

onde $\boldsymbol{\beta}^G \in \Re^q$ e $\boldsymbol{\beta}^F \in \Re^{(p-q)}$. Logo, podemos reescrever o modelo definido por (4.1) e (4.6) por

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}^* + \boldsymbol{Z}_i\boldsymbol{b}_i^* + \boldsymbol{\epsilon}_i \tag{4.17}$$

sendo

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i^* \end{pmatrix} \sim \text{St-II}_{n_i+q}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \tau_e^2\boldsymbol{I}_{n_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} \end{pmatrix}, \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_b \end{pmatrix}, \nu\right), \tag{4.18}$$

sendo $\boldsymbol{D}_b = Diag(\boldsymbol{\eta}_b)$, $\boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\beta}^{G*} \\ \boldsymbol{\beta}^F \end{pmatrix}$ e $\boldsymbol{\beta}^{G*} = \boldsymbol{\beta}^G - \kappa\boldsymbol{\eta}_b$.

Utilizaremos esta reparametrização do modelo para obter as condicionais completas, notemos que depois podemos obter facilmente amostras de $\boldsymbol{\beta}$ a partir $\boldsymbol{\beta}^*$.

4.3 Distribuições *a Priori*

Consideramos as seguinte distribuições *a priori*, similares as discutidas no capítulo 3,

$$\begin{aligned} \frac{1}{\tau_e^2} &\sim Gamma\left(a_0, b_0\right). \\ \boldsymbol{\eta}_b &\sim t_q\left(0, c1\Psi, d1\right) \\ \boldsymbol{\Psi} &\sim Inversa - Wishart\left(a_1, B_1\right). \end{aligned} \tag{4.19}$$

Assume-se como distribuição *a priori* para o parâmetro graus de liberdade ν a seguinte distribuição:

$$p(\nu) \propto \nu^{-2} I(\nu > 3), \tag{4.20}$$

Para os efeitos fixos $\boldsymbol{\beta}$ assumimos uma distribuição normal multivariada

$$\boldsymbol{\beta}^* \sim N_p\left(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}\right). \tag{4.21}$$

4.4 O Modelo Hierárquico e o Algoritmo MCMC

Uma forma hierárquica de reescrevermos o modelo definido em (4.17) e (4.18) é utilizando a representação estocástica dada em (2.5), desta forma temos que

$$\begin{aligned} \boldsymbol{Y}_i \mid \boldsymbol{b}_i^*, w_i &\sim N_{n_i}\left(\boldsymbol{X}_i\boldsymbol{\beta}^* + \boldsymbol{Z}_i\boldsymbol{b}_i^*, \frac{\tau_e^2}{w_i}\boldsymbol{I}_{ni}\right) \\ \boldsymbol{b}_i^* \mid \boldsymbol{u}_{bi}, w_i &\sim N_q\left(\boldsymbol{0}, \frac{\boldsymbol{\Psi}}{w_i}\right) \\ \boldsymbol{u}_{bi} &\sim N_q\left(\boldsymbol{0}, \frac{\boldsymbol{I}_q}{w_i}\right) I(\boldsymbol{u}_{bi} > \boldsymbol{0}) \\ w_i &\sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \tag{4.22}$$

Alguns modelos na literatura podem ser vistos como casos especiais de (4.22), dependendo dos valores dos parâmetros $\boldsymbol{\eta}_b$ e ν, e as variáveis latentes $\boldsymbol{u}_{bi} = (u_{bij})_{j=1,\ldots,n_i}$ e w_i, alguns destes modelos são apresentados na tabela 4.1.

A distribuição *a posteriori* aumentada é o produto da função de verossimilhança aumentada dada em (4.22) e a distribuição *a priori* definida em (4.19)-(4.21). Assim, podemos utilizar o algoritmo de Markov Chain Monte Carlo (MCMC) para obter amostras da distribuição *a posteriori*.

Tabela 4.1: Casos Especiais do MLM t-assimétrico dependente

$\boldsymbol{\eta}_b$	ν	u_{bij}	w_i	Modelo
$\mathbf{0}$	∞	0	1	Normal
$\mathbf{0}$	ν	0	w_i	Multivariada t [a]
$\boldsymbol{\eta}_b$	∞	u_{bi}	1	Normal assimétrica tipo I [b]
$\boldsymbol{\eta}_b$	∞	u_{bij}	1	Normal assimétrica tipo II [c]
$\boldsymbol{\eta}_b$	ν	u_{bi}	w_i	t-assimétrica tipo I

[a] Pinheiro et al. (2001)
[b] Arellano-Valle, Bolfarine e Lachos (2005)
[c] Arellano-Valle et al. (2007)

As distribuições condicionais completas são dadas na seguinte proposição

Proposição 4.1 *Sob o modelo (4.22) e assumindo-se as distribuições a priori dadas em(4.19)-(4.21), e reescrevendo a distribuição a priori de $\boldsymbol{\eta}_b$ dada em (4.19) como a seguinte mistura na escala de uma normal,*

$$\begin{aligned} \boldsymbol{\eta}_b &\sim N_q\left(0, c_1\frac{\boldsymbol{\Psi}}{v_b}\right) \\ v_b &\sim Gamma\left(d_1/2, d_1/2\right), \end{aligned} \tag{4.23}$$

segue que as distribuições condicionais completas são dadas por

(i) $\boldsymbol{\beta} \mid . \sim N_p(\boldsymbol{m}_\beta, \boldsymbol{S}_\beta)$ *onde* $\boldsymbol{m}_\beta = \boldsymbol{S}_\beta^{-1}\left(\sum_{i=1}^m w_i \boldsymbol{X}_i^T\left(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{b}_i\right)/\tau_e^2 + \boldsymbol{\Omega}_\beta^{-1}\boldsymbol{\mu}_\beta\right)$ *e* $\boldsymbol{S}_\beta^{-1} = \frac{w_i}{\tau_e^2}\sum_{i=1}^m \boldsymbol{X}_i^T\boldsymbol{X}_i + \boldsymbol{\Omega}_\beta^{-1}$;

(ii) $\boldsymbol{b}_i \mid . \sim N_q(\boldsymbol{m}_i, \boldsymbol{S}_i)$ *onde* $\boldsymbol{m}_i = w_i\boldsymbol{S}_i^{-1}\left(\sum_{i=1}^m \boldsymbol{Z}_i^T\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)/\tau_e^2 + \boldsymbol{D}_b\boldsymbol{u}_{bi}\boldsymbol{\Psi}^{-1}\right)$ *e* $\boldsymbol{S}_i^{-1} = w_i\left(\sum_{i=1}^m \frac{\boldsymbol{Z}_i^T\boldsymbol{Z}_i}{\tau_e^2} + \boldsymbol{\Psi}^{-1}\right)$;

(iii) $\boldsymbol{\eta}_b \mid . \sim N_q(\boldsymbol{m}_b, \boldsymbol{S}_b)$ *onde* $\boldsymbol{m}_b = \boldsymbol{S}_b^{-1}\left(\sum_{i=1}^m w_i\boldsymbol{b}_i^T Diag(\boldsymbol{u}_{bi})\boldsymbol{\Psi}^{-1}\right)$ *e* $\boldsymbol{S}_b^{-1} = \left(\sum_{i=1}^m w_i Diag(\boldsymbol{u}_{bi})\boldsymbol{\Psi}^{-1}Diag(\boldsymbol{u}_{bi}) + \frac{v_b}{c_0}\boldsymbol{\Psi}^{-1}\right)^{-1}$;

(iv) $\frac{1}{\tau_e^2} \mid . \sim Gamma\left(\frac{n+1}{2} + a_0, \frac{1}{2}\sum_{i=1}^m\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\right)w_i + \frac{1}{2}\frac{\eta_e^2 v_e}{c_0} + b_0\right)$;

(v) $\boldsymbol{\Psi} \mid . \sim IW\left(q + a_1 + 1, \sum_{i=1}^m w_{bi}\left(\boldsymbol{b}_i - \boldsymbol{D}_b\boldsymbol{u}_{bi}\right)\left(\boldsymbol{b}_i - \boldsymbol{D}_b\boldsymbol{u}_{bi}\right)^T + \frac{v_b}{c_0}\boldsymbol{\eta}_b\boldsymbol{\eta}_b^T + B_1\right)$;

(vi) $v_b \mid . \sim Gamma\left(\frac{d_1+q}{2}, \frac{1}{2}\left(\frac{\boldsymbol{\eta}_b\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_b}{c_1} + d_1\right)\right)$;

(vii) $\boldsymbol{u}_{bi} \mid . \sim N_q\left((\boldsymbol{I}_q + \boldsymbol{D}_b\boldsymbol{\Psi}^{-1}\boldsymbol{D}_b)^{-1}\boldsymbol{D}_b\boldsymbol{\Psi}^{-1}\boldsymbol{b}_i, \frac{1}{w_i}(\boldsymbol{I}_q + \boldsymbol{D}_b\boldsymbol{\Psi}^{-1}\boldsymbol{D}_b)^{-1}\right)I(\boldsymbol{u}_{bi} > \mathbf{0})$;

(viii) $w_i \mid . \sim Gamma\left(a_{wi}, b_{wi}\right)$ *onde* $a_{wi} = \frac{\nu + n_i + 2q}{2}$ *e* $b_{wi} = \frac{1}{2}\frac{\left(\boldsymbol{Y}_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\right)^T\left(\boldsymbol{Y}_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\right)}{\tau_e^2} + \frac{1}{2}\left(\boldsymbol{b}_i - \boldsymbol{D}_b\boldsymbol{u}_{bi}\right)^T\boldsymbol{\Psi}^{-1}\left(\boldsymbol{b}_i - \boldsymbol{D}_b\boldsymbol{u}_{bi}\right) + \frac{1}{2}\boldsymbol{u}_{bi}^T\boldsymbol{u}_{bi} + \frac{1}{2}\nu$

(ix) $f(\nu \mid .) \propto \frac{1}{\Gamma(\nu/2)^m}\nu^{-2}\left(\frac{\nu}{2}\right)^{m\frac{\nu}{2}}\exp\left(-\sum_{i=1}^m(w_i - \log w_i)\frac{\nu}{2}\right)I(\nu > 3)$.

As distribuições condicionais completas *a posteriori* são fáceis de derivar e têm formas conhecidas com exceção dos graus de liberdade, assim o amostrador de Gibbs pode ser facilmente implementado em pacotes como o R. Também, pode ser facilmente implementado utilizando o pacote WinBUGS (Spiegelhalter, Thomas, Best e Lunn, 2004) utilizando a representação estocástica dada em (4.22).

4.5 Aplicação

Nesta seção, ilustramos este modelo com um conjunto de dados longitudinais de níveis de colesterol coletados como parte do conhecido estudo do coração de Framingham, estes dados foram analisados por Zhang e Davidian (2001). Consideramos o mesmo modelo linear misto utilizado por estes autores, dado por

$$y_{ij} = \beta_0 + \beta_1 t_i + \beta_2 idade_i + \beta_3 sexo_i + b_{i0} + b_{i1} t_{ij} + e_{ij}, \tag{4.24}$$

sendo y_{ij} o nível de colesterol medido no j-ésimo tempo para o sujeito i dividido por 100; t_i é $(tempo - 5)/10$, sendo que o tempo é mensurado em anos desde o *baseline*; $sexo_i$ é uma variável indicadora para o sexo masculino; $idade_i$ é a idade no *baseline*; β_0, β_1, γ_0 e γ_1 são os efeitos fixos ; $\boldsymbol{b}_i = (b_{i0}, b_{i1})$ é o efeito aleatório para a sujeito i; e e_{ij} é o erro aleatório, $i = 1, ..., 200$, $j = 1, ..., n_i$. A figura 4.1 apresenta o histograma da variável resposta, observamos que os dados apresentam assimetria. Como o modelo estudado nesta seção introduz assimetria somente nos efeito aleatórios, ajustamos perfis individuais para cada sujeito considerando uma regressão simples sobre o tempo. Na figura 4.2 mostra os histogramas dos interceptos e inclinações estimados por mínimos quadrados. Notamos que o histograma mostra a presença de assimetria nos interceptos.

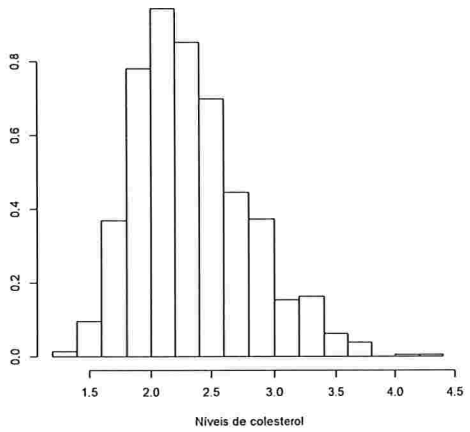

Figura 4.1: Histograma dos níveis de colesterol

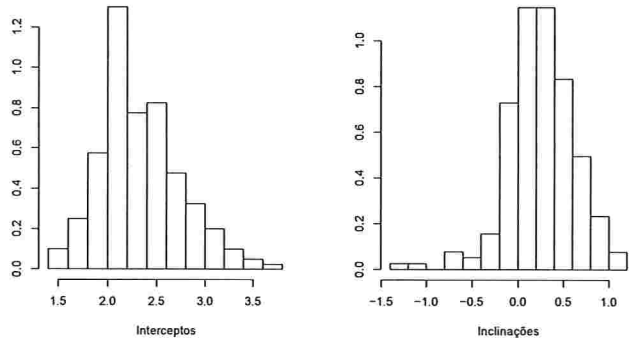

Figura 4.2: Histograma dos interceptos e inclinações dos perfis individuais

Assim, assumimos o seguinte modelo

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim \text{St-II}_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ -\kappa\boldsymbol{\eta}_b \end{pmatrix}, \begin{pmatrix} \tau_e^2 \boldsymbol{I}_{n_i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}, \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_b \end{pmatrix}, \nu\right), \tag{4.25}$$

sendo $\boldsymbol{\epsilon}_i = (e_{ij})_{j=1,\ldots,n_i}$ e $\boldsymbol{b}_i = (b_{i0}, b_{i1})$.

Ajustamos os modelos: t-assimétrico (St), normal assimétrico (SN), t e normal(N). Baseados nas distribuições condicionais completas dadas na proposição 3.1 implementamos o amostrador de Gibbs no pacote R . Para todas as modelos foram geradas 120000 amostras com *burn-in* de 20000. Para reduzir a autocorrelação entre as cadeias, tomamos valores de 10 em 10 para obtermos uma amostra de Monte Carlo de tamanho 10000 aproximadamente independente.

Para a comparação de modelos calculamos as seguintes medidas o pseudo fator de Bayes (Gelfand, 1996) com respeito ao modelo normal, o EAIC (Expected Akaike Information Criterion) e o EBIC (Expected Bayesian Information Criterion) propostos em Brooks (2002), para cada modelo. Estes critérios indicam que o modelo que assume a distribuição t-assimétrica apresenta o melhor ajuste.

Tabela 4.2: Medidas de comparação de modelos para os dados de Framingham

Distribuição	pBF	EAIC	EBIC
St	1.26×10^{12}	301.09	337.38
SN	1.66	352.27	385.26
t	3.53×10^{9}	309.14	338.83
N	1.00	350.02	376.41

Também consideramos os valores do CPO para cada observação (Gelfand, 1996). Na figura 4.3 é apresentado os CPO para os modelos: (1) t-assimétrico; (2) normal assimétrico; (3) t; e (4) normal. Os modelos 1 e 3 são claramente melhores que os outros modelos, o pBF do modelo 4

versus modelo 2 é 120.19 razão pela qual escolhemos a St como o melhor modelo.

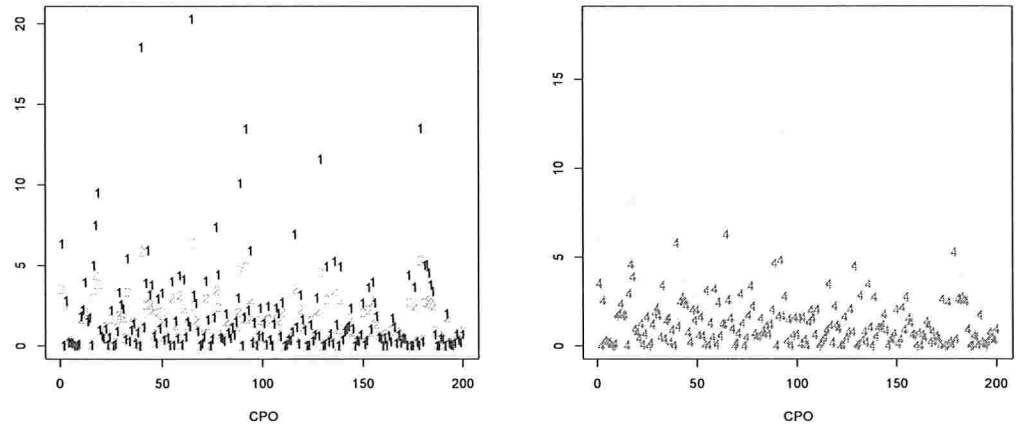

Figura 4.3: CPO para os modelos: (1) St, (2) SN, (3) t e (4) N dos dados de Framningham

Na tabela 4.3 apresentamos a média *a posteriori* e o intervalo HPD de 95%−HPD para o modelo St . Notamos que o HPD para o parâmetro de assimetria do intercepto λ_{b1} não inclue o valor de zero o que confirma a presença de assimetria positiva nos interceptos que foi observado na figura 4.2. A pequena amplitude do intervalo para ao parâmetro graus de liberdade dos erros (ν) reforça a conclusãqo tirada das medidas de de desempenho do modelo, isto é, que os modelos normal ou normal assimétrico não são apropriados para este conjunto de dados.

Tabela 4.3: Média *a posteriori*, desvio padrâo (D.P) e intervalo HPD de 95% de probabilidade para o modelo St

Parâmetro	Média	D.P.	2.5%	97.5%
β_0	1.6646	0.1501	1.3739	1.9610
β_1	0.2798	0.0264	0.2276	0.3311
β_2	0.0160	0.0035	0.0094	0.0229
β_3	-0.0479	0.0506	-0.1453	0.0534
$Var(e_{ij})$	0.0439	0.0037	0.0370	0.0513
$Var(b_{i0})$	0.1576	0.0215	0.1183	0.2001
$Cov(b_{i0}, b_{i1})$	0.0295	0.0121	0.0057	0.0530
$Var(b_{i1})$	0.0786	0.0157	0.0497	0.1099
λ_{b0}	2.9375	1.4838	0.3685	5.9236
λ_{b1}	0.1345	38.3671	-7.7924	10.8332
ν	8.4537	2.0014	5.102	12.5600

Capítulo 5

Modelo Linear Misto t-assimétrico Independente

5.1 Introdução

O modelo linear misto (Laird e Ware,1982) é dado por

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{5.1}$$

para $i = 1, ..., m$, em que $\boldsymbol{Y}_i$ é um vetor n_i-dimensional de valores observados para a i-ésima unidade, $\boldsymbol{X}_i$ é uma matriz $n_i \times p$ que contém valores de p variáveis explicativas, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos, $\boldsymbol{Z}_i$ é uma matriz $n_i \times q$ que contém a especificação dos efeitos aleatórios $\boldsymbol{b}_i$ e $\boldsymbol{\epsilon}_i$ é um vetor n_i-dimensional de erros. É comum assumir normalidade para os erros ϵ_i e para os efeitos aleatórios $\boldsymbol{b}_i$.

Um caso especial e muito utilizado é se consideramos $\boldsymbol{R}_i = \tau_e^2\boldsymbol{I}_{n_i}$

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim N_{n_i+q}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \tau_e^2\boldsymbol{I}_{n_i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} \end{pmatrix}\right). \tag{5.2}$$

Isto é, assumimos que os erros $\boldsymbol{\epsilon}_i$ e os efeitos aleatórios $\boldsymbol{b}_i$ são independentes entre si. Além disso, que os erros entre sujeitos também são independentes portanto não correlacionados.

No capítulo 5, o modelo (5.2) foi estendido para o caso t-assimétrico considerando o caso em que os erros $\boldsymbol{\epsilon}_i$ e os efeitos aleatórios $\boldsymbol{b}_i$ seguem conjuntamente uma distribuição t-assimétrica multivariada. Mas para manter as suposições de que $\boldsymbol{\epsilon}_i$ e $\boldsymbol{b}_i$ sejam não correlacionados e que os erros entre sujeitos também sejam não correlacionados, foi necessário considerar que assimetria estava presente somente nos efeitos aleatórios. Neste capítulo consideraremos o caso em que ϵ_{ij} e $\boldsymbol{b}_i$ são independentes, $\forall i, j$, e que cada um deles segue uma distribuição t-assimétrica proposta por Branco e Dey (2001) que coincide com a apresentada por Azzalini e Capitanio (2003).

5.2 Modelo Linear Misto t-assimétrico independente

Neste trabalho, estendemos o modelo dado em (5.2)

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

considerando,

$$\begin{aligned} \boldsymbol{b}_i &\sim \text{St-I}_q\left(-\kappa_b \boldsymbol{\eta}_b, \boldsymbol{\Psi}, \boldsymbol{\lambda}_b, \nu_b\right), \\ \epsilon_{ij} &\sim \text{St-I}_1\left(-\kappa_e \eta_e, \sigma_e^2, \lambda_e, \nu_e\right), \end{aligned} \tag{5.3}$$

sendo $\kappa_b = (\nu_b/\pi)^{1/2}\Gamma\left(\frac{\nu_b-1}{2}\right)/\Gamma\left(\frac{\nu_b}{2}\right)$ e $\kappa_e = (\nu_e/\pi)^{1/2}\Gamma\left(\frac{\nu_e-1}{2}\right)/\Gamma\left(\frac{\nu_e}{2}\right)$. Assim, temos que

$$E\left(\boldsymbol{b}_i\right) = \boldsymbol{0} \text{ e } E\left(\epsilon_{ij}\right) = 0$$

$i = 1, ..., m$ e $j = 1, ...n_i$. Notemos que $\boldsymbol{b_i}$ é independente de ϵ_{ij} e que os ϵ_{ij} são mutuamente independentes.

Utilizando a representação estocástica dada em (2.5), este modelo pode ser reescrito em forma hierárquica como:

$$\begin{aligned} y_{ij} \mid \boldsymbol{b}_i, u_{eij}, w_{eij} &\sim N_1\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \eta_e(u_{eij} - \kappa_e), \frac{\tau_e^2}{w_{eij}}\right) \\ \boldsymbol{b}_i \mid u_{bi}, w_{bi} &\sim N_q\left(\boldsymbol{\eta_b}(u_{bi} - \kappa_b), \frac{\boldsymbol{\Lambda}}{w_{bi}}\right) \\ u_{eij} &\sim HN\left(0, \frac{1}{w_{eij}}\right) \\ u_{bi} &\sim HN\left(0, \frac{1}{w_{bi}}\right) \\ w_{eij} &\sim Gama\left(\frac{\nu_e}{2}, \frac{\nu_e}{2}\right) \\ w_{bi} &\sim Gama\left(\frac{\nu_b}{2}, \frac{\nu_b}{2}\right) \end{aligned} \tag{5.4}$$

sendo $\boldsymbol{x}_{ij}$ e $\boldsymbol{z}_{ij}$ a j-ésima fila de $\boldsymbol{X}_i$ e $\boldsymbol{Z}_i$ respectivamente; $\sigma_e^2 = \tau_e^2 + \eta_e^2$; $\lambda_e = \frac{\eta_e}{\tau_e}$; $\boldsymbol{\Psi} = \boldsymbol{\Lambda} + \boldsymbol{\eta}_b\boldsymbol{\eta}_b^T$ e $\boldsymbol{\lambda}_b = \dfrac{\boldsymbol{\Psi}^{1/2}\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_b}{\sqrt{1 + \boldsymbol{\eta}_b^T\boldsymbol{\Lambda}_b^{-1}\boldsymbol{\eta}_b}}$.

Alguns modelos na literatura podem ser vistos como casos especiais deste modelo, dependendo dos valores dos parâmetros λ_e, $\boldsymbol{\lambda}_b$, ν_e e ν_b, e das variáveis latentes u_{eij}, u_{bi}, w_{eij} e w_{bi}. Por exemplo, quando $\lambda_e = 0$, $\boldsymbol{\lambda}_b = \boldsymbol{0}$, $u_{eij} = u_{bi} = 0$ e $w_{eij} = w_{bi} = w_i$, sendo que os w_i são mutuamente independentes e seguem uma distribuição gama com parâmetros $\nu/2$ e $\nu/2$, obtemos uma distribuição t-Student conjunta para os efeitos aleatórios e os erros,

$$\begin{pmatrix} \epsilon_i \\ \boldsymbol{b}_i \end{pmatrix} \sim t_{n_i+q}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \sigma_e^2\boldsymbol{I}_{ni} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} \end{pmatrix}, \nu\right)$$

que é o modelo proposto por Pinheiro et al. (2001). Outros exemplo são apresentado na tabela 5.1.

Tabela 5.1: Casos Especiais do MLM t-assimétrico independente

λ_e	$\boldsymbol{\lambda}_b$	ν_e	ν_b	u_{eij}	u_{bj}	w_{eij}	w_{bj}	Modelo
0	$\mathbf{0}$	∞	∞	0	0	1	1	Normal
0	$\mathbf{0}$	ν	ν	0	0	w_i	w_i	Multivariada t [a]
0	$\mathbf{0}$	ν_e	ν_b	0	0	w_{ei}	w_{bj}	Erros t não correlacionados[b]
0	$\mathbf{0}$	ν_e	ν_b	0	0	w_{eij}	w_{bj}	Erros t independentes [b]
λ_e	$\boldsymbol{\lambda}_b$	∞	∞	u_{ei}	u_{bj}	1	1	Normal assimétrica [c]

[a] Pinheiro et al. (2001)
[b] Rosa, Gianola e Padovani (2004)
[c] Arellano-Valle, Bolfarine e Lachos (2005)

Utilizando propriedas da distribuição t-assimétrica obtemos que

$$E[\boldsymbol{Y}_i] = \boldsymbol{X}_i\boldsymbol{\beta} \tag{5.5}$$

e

$$Var[\boldsymbol{Y}_i] = \boldsymbol{Z}_i Var[\boldsymbol{b}_i]\boldsymbol{Z}_i + Var[\boldsymbol{\epsilon}_i], \tag{5.6}$$

sendo

$$Var[\boldsymbol{b}_i] = \frac{\nu_b}{\nu_b - 2}\boldsymbol{\Psi} - \kappa_b^2 \boldsymbol{\eta}_b \boldsymbol{\eta}_b^T$$

e

$$Var[\boldsymbol{\epsilon}_i] = \left[\frac{\nu_e}{\nu_e - 2}\sigma_e^2 - \kappa_e^2 \eta_e^2\right] \boldsymbol{I}_{n_i}.$$

5.3 Distribuições *a Priori*

Para completar a especificação do modelos, devemos considerar distribuições *a priori* para todos os parâmetros. Assim, considerando a revisão feita na seção 4.3, consideramos para os parâmetros dos erros

$$\begin{aligned} \eta_e &\sim t_1(0, c_0\tau_e^2, d_0) \\ \frac{1}{\tau_e^2} &\sim Gamma\,(a_0, b_0)\,. \end{aligned} \tag{5.7}$$

Para os parâmetros dos efeitos aleatórios,

$$\begin{aligned} \boldsymbol{\eta}_b &\sim t_q\,(0, c_1\Psi, d_1) \\ \boldsymbol{\Psi} &\sim IW\,(a_1, B_1)\,. \end{aligned} \tag{5.8}$$

Para os parâmetros graus de liberdade ν_e e ν_b assumimos

$$\begin{aligned} p(\nu_e) &\propto \nu_e^{-2} I(\nu_e > 3), \\ p(\nu_b) &\propto \nu_b^{-2} I(\nu_b > 3), \end{aligned} \tag{5.9}$$

como em Liu (1995) e Rosa et al. (2003).

Finalmente, assumimos uma distribuição normal multivariada para os efeitos fixos $\boldsymbol{\beta}^*$,

$$\boldsymbol{\beta}^* \sim N_p\left(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}\right). \tag{5.10}$$

5.4 O Modelo Hierárquico e o Algoritmo MCMC

Em forma similar ao capítulo 4, para obter formas mais simples para as distribuições condicionais completas, consideraremos o caso usual em modelos lineares mistos em que as colunas de $\boldsymbol{Z}_i$ estão contidas em $\boldsymbol{X}_i$, assim

$$\boldsymbol{X}_i = \begin{pmatrix} \mathbf{1}_{n_i} & \boldsymbol{X}_i^G & \boldsymbol{X}_i^F \end{pmatrix} \text{ e } \boldsymbol{Z}_i = \begin{pmatrix} \mathbf{1}_{n_i} & \boldsymbol{X}_i^G \end{pmatrix} \tag{5.11}$$

onde $\mathbf{1}_{n_i}$ é um vetor de uns de tamanho n_i, $\boldsymbol{X}_i^G$ é uma matriz de planejamento $n_i \times (q-1)$, $\boldsymbol{X}_i^F$ é uma matriz de planejamento $n_i \times (p-q)$ e consideramos a seguinte partição de $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}^G \\ \boldsymbol{\beta}^F \end{pmatrix} \tag{5.12}$$

onde $\beta_0 \in \Re$, $\boldsymbol{\beta}^G \in \Re^{(q-1)}$ e $\boldsymbol{\beta}^F \in \Re^{(p-q)}$. Logo, podemos reescrever o modelo definido por (5.2) e (5.3) por

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}^* + \boldsymbol{Z}_i\boldsymbol{b}_i^* + \boldsymbol{\epsilon}_i^*$$

considerando,

$$\begin{aligned} \boldsymbol{b}_i^* &\overset{\text{iid}}{\sim} \text{St-I}_q\left(\mathbf{0}, \boldsymbol{\Psi}, \boldsymbol{\lambda}_{\boldsymbol{b}}, \nu_b\right), \\ \epsilon_{ij}^* &\overset{\text{iid}}{\sim} \text{St-I}_1\left(0, \sigma_e^2, \lambda_e, \nu_e\right), \end{aligned} \tag{5.13}$$

sendo $\boldsymbol{\beta}^* = \begin{pmatrix} \beta_0^* \\ \boldsymbol{\beta}^{G*} \\ \boldsymbol{\beta}^F \end{pmatrix}$, $\boldsymbol{\eta} = \begin{pmatrix} \eta_{b0} \\ \boldsymbol{\eta}^G \end{pmatrix}$, $\beta_0^* = \beta_0 - \kappa_e\eta_e - \kappa_b\eta_{b0}$ e $\boldsymbol{\beta}^{G*} = \boldsymbol{\beta}^G - \kappa_b\boldsymbol{\eta}_b^G$.

Utilizaremos esta reparametrização do modelo para obter as condicionais completas, notemos que depois podemos obter facilmente amostras de $\boldsymbol{\beta}$ a partir $\boldsymbol{\beta}^*$. Assim a forma hierárquica dada em (5.4) pode ser escrita como:

$$
\begin{aligned}
y_{ij} \mid \boldsymbol{b}_i, u_{eij}, w_{eij} &\sim N_1\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \eta_e u_{eij}, \frac{\tau_e^2}{w_{eij}}\right) \\
\boldsymbol{b}_i \mid u_{bi}, w_{bi} &\sim N_q\left(\boldsymbol{\eta}_b u_{bi}, \frac{\boldsymbol{\Lambda}}{w_{bi}}\right) \\
u_{eij} &\sim HN\left(0, \frac{1}{w_{eij}}\right) \\
u_{bi} &\sim HN\left(0, \frac{1}{w_{bi}}\right) \\
w_{eij} &\sim Gama\left(\frac{\nu_e}{2}, \frac{\nu_e}{2}\right) \\
w_{bi} &\sim Gama\left(\frac{\nu_b}{2}, \frac{\nu_b}{2}\right)
\end{aligned}
\tag{5.14}
$$

Logo, obtemos a densidade *a posteriori* aumentada como o produto da função de verossimilhança aumentada dada em (5.14) e a distribuição *a priori* definida por (5.7)-(5.10).
Consideraremos a seguinte notação $n = \sum_{i=1}^m ni$, $\boldsymbol{u}_{ei} = (u_{ei1}, ..., u_{eini})^T$, $\boldsymbol{w}_{ei} = (w_{ei1}, ..., w_{eini})^T$ e $\boldsymbol{W}_{ei} = Diag(\boldsymbol{w}_{ei})$.

Proposição 5.1 *Sob a representação hierárquica (5.14) e a distribuição a priori (5.7)-(5.10), e reescrevendo as distribuições a prior de η_e e $\boldsymbol{\eta}_b$ dadas em (5.7) e (5.8) respectivamente, como uma mistura na escala de uma normal*

$$
\begin{aligned}
\eta_e &\sim N_1\left(0, c_0\frac{\tau_e^2}{v_e}\right) \\
v_e &\sim Gama\left(d_0/2, d_0/2\right)
\end{aligned}
\tag{5.15}
$$

e

$$
\begin{aligned}
\boldsymbol{\eta}_b &\sim N_q\left(0, c_1\frac{\boldsymbol{\Lambda}}{v_b}\right) \\
v_b &\sim Gama\left(d_1/2, d_1/2\right).
\end{aligned}
\tag{5.16}
$$

As distribuições condicionais completas são dadas por

(i) $\boldsymbol{\beta}^* \mid . \sim N_p(\boldsymbol{m}_\beta, \boldsymbol{S}_\beta)$ *sendo* $\boldsymbol{m}_\beta = \boldsymbol{S}_\beta^{-1}\left(\sum_{i=1}^m \boldsymbol{X}_i^T\boldsymbol{W}_{ei}\left(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{b}_i^* - \eta_e\boldsymbol{u}_{ei}\right)/\tau_e^2 + \boldsymbol{\Omega}_\beta^{-1}\boldsymbol{\mu}_\beta\right)$ *e* $\boldsymbol{S}_\beta^{-1} = \sum_{i=1}^m \frac{\boldsymbol{X}_i^T\boldsymbol{W}_{ei}\boldsymbol{X}_i}{\tau_e^2} + \boldsymbol{\Omega}_\beta^{-1}$;

(ii) $\boldsymbol{b}_i^* \mid . \sim N_q(\boldsymbol{m}_i, \boldsymbol{S}_i)$ *sendo* $\boldsymbol{m}_i = \boldsymbol{S}_i^{-1}\left(\sum_{i=1}^m \boldsymbol{Z}_i^T\boldsymbol{W}_{ei}\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}^* - \eta_e\boldsymbol{u}_{ei}\right)/\tau_e^2 + w_{bi}u_{bi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_b\right)$ *e* $\boldsymbol{S}_i^{-1} = \sum_{i=1}^m \frac{\boldsymbol{Z}_i^T\boldsymbol{W}_{ei}\boldsymbol{Z}_i}{\tau_e^2} + w_{bi}\boldsymbol{\Lambda}^{-1}$;

(iii) $\eta_e \mid . \sim N_1(m_e, s_e)$ *sendo* $m_e = s_e^{-1}\left(\sum_{i=1}^m\sum_{j=1}^{n_i}\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}^* - \boldsymbol{z}_{ij}\boldsymbol{b}_i^*\right)w_{eij}u_{eij}\right)$ *e* $s_e^{-1} = \left(\sum_{i=1}^m\sum_{j=1}^{n_i}u_{eij}^2w_{eij} + v_e/c_0\right)/\tau_e^2$;

(iv) $\boldsymbol{\eta}_b \mid . \sim N_q(\boldsymbol{m}_b, \boldsymbol{S}_b)$ *sendo* $\boldsymbol{m}_b = \boldsymbol{S}_b^{-1}\left(\sum_{i=1}^m w_{bi}u_{bi}\boldsymbol{b}_i^*\right)$ *e* $\boldsymbol{S}_b^{-1} = \left(\sum_{i=1}^m u_{bi}^2w_{bi} + \frac{v_b}{c_0}\right)^{-1}\boldsymbol{\Lambda}$;

(v) $\frac{1}{\tau_e^2} \mid . \sim Gamma\left(\frac{n+1}{2} + a_0, \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^{n_i}\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}^* - \boldsymbol{z}_{ij}\boldsymbol{b}_i^* - \eta_e u_{eij}\right)w_{eij} + \frac{1}{2}\frac{\eta_e^2 v_e}{c_0} + b_0\right)$;

(vi) $\boldsymbol{\Lambda} \mid . \sim IW\left(q + a_1 + 1, \sum_{i=1}^{m} w_{bi}\left(\boldsymbol{b}_i^* - \boldsymbol{\eta}_b u_{bi}\right)\left(\boldsymbol{b}_i^* - \boldsymbol{\eta}_b u_{bi}\right)^T + \frac{v_b}{c_0}\boldsymbol{\eta}_b\boldsymbol{\eta}_b^T + B_1\right);$

(vii) $v_e \mid . \sim Gamma\left(\frac{d_0}{2}, \frac{1}{2}\left(\frac{\eta_e^2}{\tau_e^2 c_0} + d_0\right)\right);$

(viii) $v_b \mid . \sim Gamma\left(\frac{d_1+q}{2}, \frac{1}{2}\left(\frac{\boldsymbol{\eta}_b\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_b}{c_1} + d_1\right)\right);$

(ix) $u_{eij} \mid . \sim N_1\left(\frac{\eta_e}{\eta_e^2+\tau_e^2}\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}^* - \boldsymbol{z}_{ij}^T\boldsymbol{b}_i^*\right), \frac{1}{w_{eij}}\frac{\tau_e^2}{\eta_e^2+\tau_e^2}\right) I(0,\infty);$

(x) $u_{bi} \mid . \sim N_1\left(\frac{\boldsymbol{\eta}_b\Lambda^{-1}\boldsymbol{b}_i^*}{1+\boldsymbol{\eta}_b^t\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_b}, \frac{1}{w_{bi}}\frac{1}{1+\boldsymbol{\eta}_b^t\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_b}\right) I(0,\infty);$

(xi) $w_{eij} \mid . \sim Gamma\left(\frac{\nu_e}{2} + 1, \frac{1}{2}\left(\frac{\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}^* - \boldsymbol{z}_{ij}\boldsymbol{b}_i^* - \eta_e u_{eij}\right)}{\tau_e^2} + u_{eij}^2 + \nu_e\right)\right);$

(xii) $w_{bi} \mid . \sim Gamma\left(\frac{\nu_b+q+1}{2}, \frac{1}{2}\left(\left(\boldsymbol{b}_i - \boldsymbol{\eta}_b(u_{bi})\right)^T\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{b}_i - \boldsymbol{\eta}_b(u_{bi})\right) + u_{bi}^2 + \nu_b\right)\right)$

(xiii) $f(\nu \mid .) \propto \frac{1}{\Gamma(\nu_e/2)^m}\nu_e^{-2}\left(\frac{\nu_e}{2}\right)^{m\frac{\nu_e}{2}} \exp\left(-\sum_{i=1}^{m}(w_i - \log w_i)\frac{\nu_e}{2}\right).$

(xiv) $f(\nu_b \mid .) \propto \frac{1}{\Gamma(\nu_b/2)^m}\nu_b^{-2}\left(\frac{\nu_b}{2}\right)^{m\frac{\nu_b}{2}} \exp\left(-\sum_{i=1}^{m}(w_i - \log w_i)\frac{\nu_b}{2}\right).$

Agora podemos utilizar algoritmos de Markov Chain Monte Carlo (MCMC) para obter amostra da distribuição *a posteriori*. Para implementar o amostrador de Gibbs precisamos das distribuições condicionais completas dadas na proposição 5.1. Notamos que todas as distribuições condicionais completas tem formas conhecidas com exceção dos graus de liberdade. No entanto, podemos utilizar o algoritmo de rejeição adaptativa (Gilks e Wild, 1992) para obter amostras de ν_e e ν_b (ver apêndice B). O algoritmo de Gibbs também pode ser facilmente implementado utilizando o pacote WinBUGS (Spiegelhalter, Thomas, Best e Lunn, 2004).

5.5 Aplicação

Os dados considerados nesta seção foram previamente analisados por Dempster, Patel, Selwyn e Roth (1984) e Rosa et al. (2003). Os dados são de um estudo toxicológico de reprodução em ratos, a variável resposta é o peso (em gramas) dos filhotes no momento do nascimento. Neste experimento, 30 ninhadas foram aleatoriamente alocadas em três grupos de tratamento do mesmo tamanho: controle, dose baixa e dose alta da substância em teste. Dados de somente 7 ninhadas foram disponíveis para o tratamento de dose alta, e o número de crias por ninhada varia de 2 a 18. Estes dados foram analisados utilizando um modelo misto com duas fontes de variação entre ninhadas e dentro de ninhadas. O modelo considera os seguintes efeitos fixos: tamanho da ninhada, sexo e tratamento. Nós adotamos o mesmo modelo linear misto considerado por Dempster et al. (1984) e Rosa et al. (2003), dado por

$$y_{ij} = \beta_0 + \beta_1 LD_i + \beta_2 HD_i + \gamma L_i + \delta S_{ij} + b_i + e_{ij}, \tag{5.17}$$

sendo y_{ij} o peso no nascimento (em gramas) da cria j da ninhada i; LD_i é uma variável indicadora para o tratamento com dose baixa para a ninhada i; HD_i é uma variável indicadora para o tratamento com dose alta para a ninhada i; L_i é o tamanho da ninhada i; S_{ij} é uma variável indicadora

para o sexo masculino das crias; β_0, β_1 e β_2 são os efeitos fixos do tratamentos; γ é a inclinação para o efeito do tamanho da ninhada; δ é o efeito fixo para o sexo da cria; b_i é o efeito aleatório para a ninhada i; e e_{ij} é o erro dentro da ninhada, $i = 1, ..., 47$, $j = 1, ..., L_i$.
A figura 5.1 apresenta o histograma da variável resposta, que indica que os dados apresentam assimetria e que seria razoável considerar um modelo que considere esta assimetria.

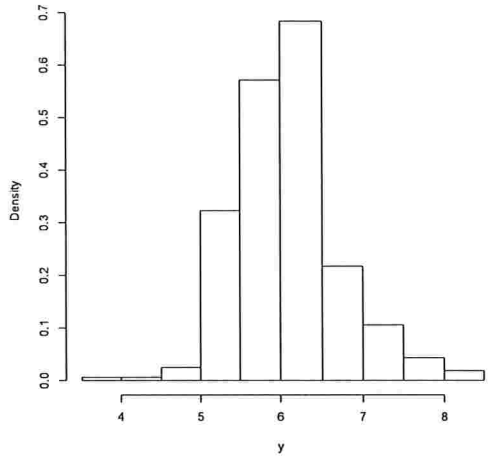

Figura 5.1: o peso no nascimento (em gramas) da crias de ratos

Assumimos que $e_{ij} \sim \text{St-I}_1(-\kappa_e\eta_e, \sigma_e^2, \lambda_e, \nu_e)$ e $b_i \sim \text{St-I}_1(-\kappa_b\eta_e, \sigma_b^2, \lambda_b, \nu_b)$.

5.5.1 Análise de Sensibilidade

Escolhemos as seguintes distribuições *a priori*,

$$\begin{aligned}
&\boldsymbol{\beta}^* \sim N_p\left(\mathbf{0}, 10^{-6}\boldsymbol{I}_n\right), \\
&\tfrac{1}{\tau_e^2} \sim Gamma\,(0.01, 0.01), \\
&\boldsymbol{\Psi}^{-1} \sim Gamma\,(0.01, 0.01), \\
&\eta_e \sim t_1\left(0, c_0\tau_e^2, d_0\right), \\
&\boldsymbol{\eta}_b \sim t_1\left(0, c_1\boldsymbol{\Psi}, d_1\right), \\
&p(\nu_e) \propto \nu_e^{-2} I(\nu_e > 3), \\
&p(\nu_b) \propto \nu_b^{-2} I(\nu_b > 3).
\end{aligned} \tag{5.18}$$

Implementamos um análise de sensibilidade da influência da escolha da distribuição *a priori* na distribuição *a posteriori*. Mas especificamente consideramos diferentes distribuições *a priori* para os parâmetro dos graus de liberdade ν_e and ν_b e para os parâmetros de assimetria η_e e $\boldsymbol{\eta}_b$. Em todos os casos concentramos nossa atenção na estimação dos efeitos fixos.

Para os parâmetros dos graus de liberdade ν_e and ν_b consideramos as seguintes distribuições *a priori*

(I) a distribuição a priori definida em (5.18),

(II) $gamma(1,1)I(3,\infty)$,

(III) $gamma(5,1)I(3,\infty)$ e

(IV) $gamma(10,1)I(3,\infty)$;

e para os parâmetros de assimetria η_e e η_b as distribuições *a priori* definidas em (5.18) com os hiperparâmetros

(I) $c_0 = c_1 = 1/2$ e $d_0 = d_1 = 2$,

(II) $c_0 = c_1 = \pi^2/4$ e $d_0 = d_1 = 1/2$,

(III) $c_0 = c_1 = 100$ e $d_0 = d_1 = 4$ e

(IV) $c_0 = c_1 = 1000$ e $d_0 = d_1 = 20$.

Para todas as distribuições *a priori* consideradas, ajustamos o modelo utilizando o amostrador de Gibbs que foi implementado no pacote R, foram geradas 120000 amostras com *burn-in* de 20000. Para reduzir a autocorrelação entre as cadeias, tomamos valores de 10 em 10 para obtermos uma amostra de Monte Carlo aproximadamente independente de tamanho 10000.

Observamos que a distribuição *a posteriori* dos efeitos fixos aparenta ser em geral robusta a escolha da distribuição *a priori* de ν_e, ν_b, η_e e η_b, ver figuras 5.2 e 5.3. No entanto a distribuição *a posteriori* de ν_e, ν_b, η_e e η_b pode ser influenciada pela escolha da distribuição *a priori*, ver figuras 5.4 e 5.5.

Notamos que a especificação *a priori* para ν_e e ν_b estabelecida em (5.18) é bastante flexível de modo a permitir que distribuições *a posteriori* distintas para ν_e e ν_b. Além disso, para os parâmetros de assimetria η_e e η_b observamos que embora a distribuição *a posteriori* desses parâmetros seja influenciada pela escolha da distribuição *priori* isto não muda a inferência neste caso que os erros apresentam assimetria e os efeitos aleatórios não.

5.5.2 Resultados

Ajustamos vários modelos que consideram diferentes distribuições para os erros e os efeitos aleatórios. Quatro distribuições foram consideradas: t-assimétrica (St), normal assimétrica (SN), t e normal(N). Para a comparação de modelos calculamos as seguintes medidas o pseudo fator de Bayes (Gelfand, 1996), o EAIC (*Expected Akaike Information Criterion*) e o EBIC (*Expected Bayesian Information Criterion*) propostos em Brooks (2002), para cada modelo. Notamos que os efeitos aleatórios aparentam ser normais, ver tabela 5.3.

O pseudo fator de Bayes (pBF), o EAIC e o EBIC para os modelos com efeitos aleatórios normais são apresentados na tabela 4.2. Estes critérios indicam que o modelo que assume erros assimétricos apresenta o melhor ajuste.

Também consideramos os valores do CPO para cada observação (Gelfand, 1996). Na figura 5.6 á apresentado o CPO para os modelos: (1) normal para os erros e os efeitos aleatórios; (2) normal

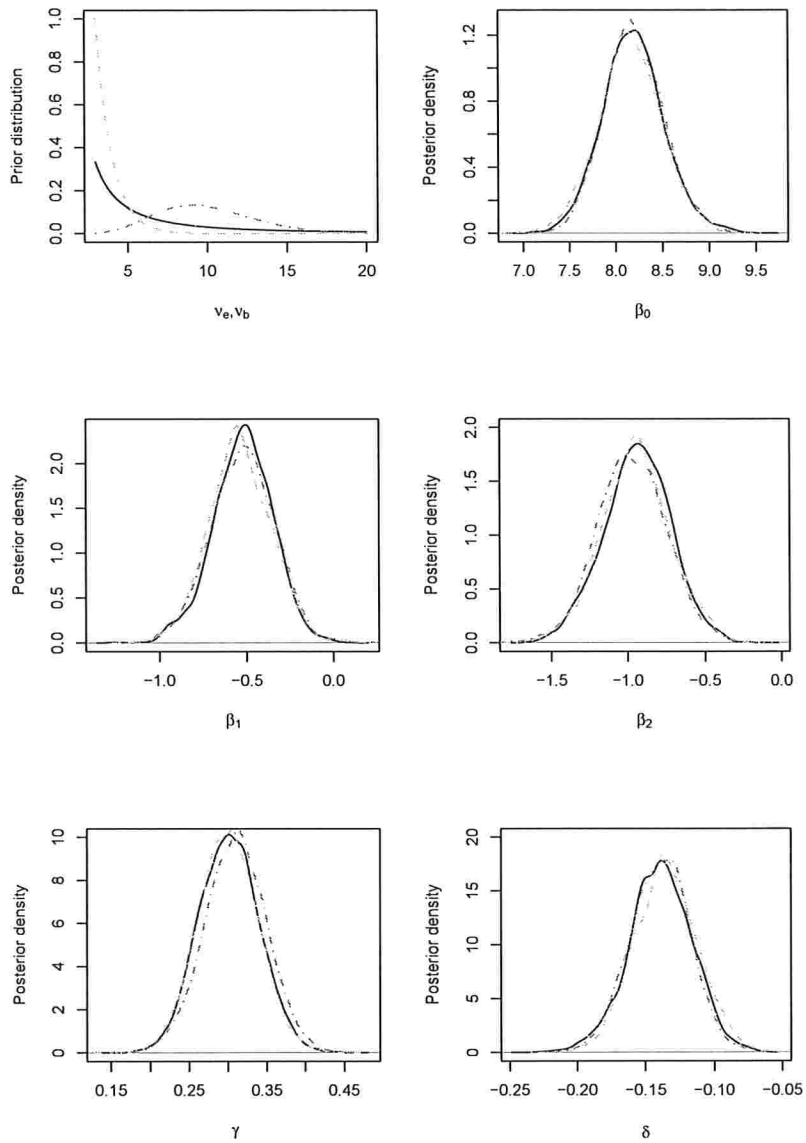

Figura 5.2: Distribuição *a priori* dos parâmetros graus de liberdade e distribuição *a posteriori* dos efeitos fixos para as distribuições *a priori* I (linha solida), II (linha tracejada), III (linha pontilhada), IV (linha ponto traço)

para os efeitos aleatórios e t para os erros; (3) normal para os efeitos aleatórios e SN para os erros; e (4) normal para os efeitos aleatórios e St para os erros. Os modelos 2 e 4 são claramente melhores que os outros modelos, o pBF do modelo 4 versus modelo 2 é 57.14 razão pela qual escolhemos a St como o melhor modelo para os erros.

Na tabela 5.3 apresentamos a média *a posteriori* e o intervalo HPD de 95%−HPD para os

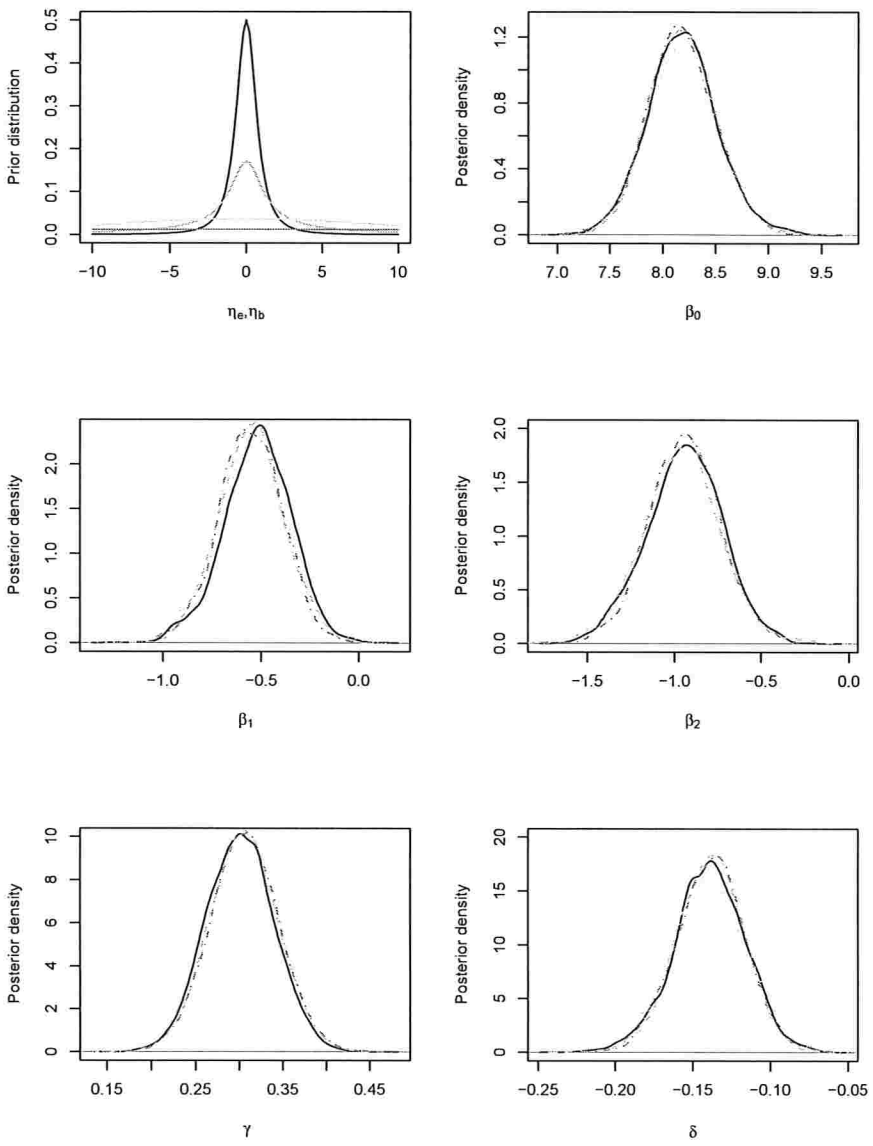

Figura 5.3: Distribuição *a priori* dos parâmetros de assimetria e distribuição *a posteriori* dos efeitos fixos para as distribuições *a priori* I (linha solida), II (linha tracejada), III ((linha pontilhada) e IV (linha ponto traço)

modelos (4) e para o modelo completo $St - St$. No modelo (4) notamos que o HPD para o parâmetro de assimetria λ_e não inclue o valor de zero o que confirma a presença de assimetria negativa nos dados. A pequena amplitude do intervalo HPD de 95% para o parametro graus de liberdade dos erros (ν_e) mostra que os modelos normal ou normal assimétrico não são apropriados neste caso.

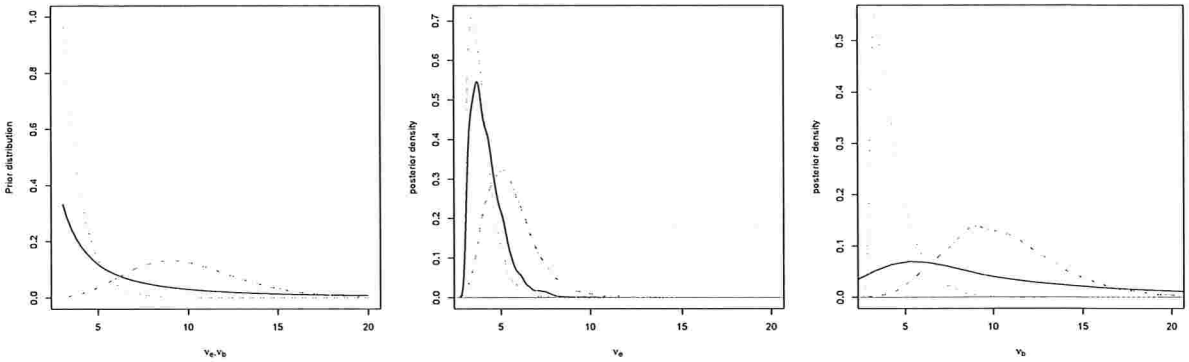

Figura 5.4: Distribuição *a priori* e distribuição *a posteriori* dos parâmetros de graus de liberdade para as distribuições *a priori* I (linha solida), II (linha tracejada), III ((linha pontilhada) e IV (linha ponto traço)

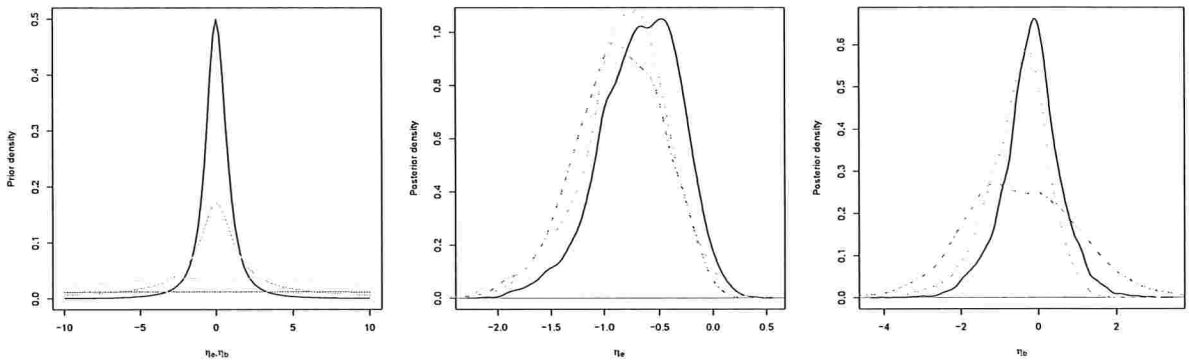

Figura 5.5: Distribuição *a priori* e *a posteriori* dos parâmetros de assimetria para as para as distribuições *a priori* I (linha solida), II (linha tracejada), III ((linha pontilhada) e IV (linha ponto traço)

Tabela 5.2: Medidas de comparação de modelos para os dados toxicológicos

Random Effects	Residuals	pBF	EAIC	EAIC
N	St	2.32×10^{21}	311.32	458.53
N	SN	8.57×10^{9}	361.80	505.23
N	t	4.07×10^{19}	316.61	460.05
N	N	1.00	404.70	544.36

No modelo $St - St$, o HPD do parâmetro de assimetria contem o valor de zero e média *a posteriori* dos graus de liberdade é bastante alta o que nos indica que o modelo normal é adequado para modelar os efeitos aleatórios.

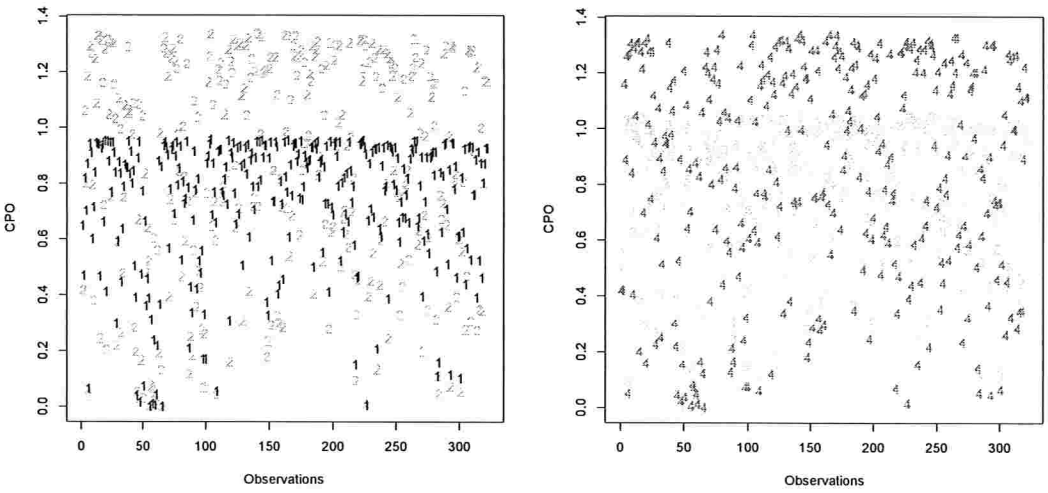

Figura 5.6: CPO para os modelos: (1) N−N, (2) N−t, (3) N−SN e (4) N−St

Tabela 5.3: Média *a posteriori* e intervalo HPD de 95%− para os modelos N $-St$ e $St-St$

	N $-St$	$St-St$
β_0^*	8.19	8.17
	(7.56 , 8.79)	(7.53 , 8.80)
β_1	−0.52	−0.52
	(−0.85 , −0.17)	(−0.88 , −0.18)
β_2	−0.94	−0.94
	(−1.34 , −0.53)	(−1.35 , −0.52)
δ	0.30	0.30
	(0.23 , 0.38)	(0.23 , 0.38)
γ	−0.14	−0.14
	(−0.18 , −0.10)	(−0.18 , −0.09)
$Var[b_i]$	0.14	0.17
	(0.06 , 0.23)	(0.06 , 0.32)
$Var[e_{ij}]$	0.16	0.16
	(0.11 , 0.21)	(0.11 , 0.21)
λ_b		0.04
		(−1.93 , 1.85)
λ_e	−0.77	−0.73
	(−1.58 , 0.00)	(−1.57 , 0.06)
ν_b		86.45
		(3.00 , 105.10)
ν_e	4.13	4.14
	(3.00 , 5.81)	(3.00 , 5.85)

Capítulo 6

Estudo de Robustez

6.1 Introdução

Nesta seção estudamos a influência de um valor *outlier* na distribuição *a posteriori*. Consideramos dois casos: o primeiro investigar qual é a influência na distribuição *a posteriori* dos efeitos fixos e segundo a influência de um ponto *outlier* na distribuição *a posteriori* de todos os parâmetros.
Consideramos o conjunto de dados de um estudo ortodôntico de 16 meninos e 11 meninas entre as idades de 8 aos 14 anos que foram originalmente apresentados por Potthoff e Roy (1964). A variável resposta é a distância (em mm) entre o centro da pituitária e a fissura ptério-maxilar, que foi medida nas idades de 8, 10, 12 e 14 anos. Estes dados também foram analisados por Pinheiro et al. (2001) e Rosa et al. (2004).Como Pinheiro et al. (2001), para o nosso estudo da influencia de um valor *outlier* somente consideramos as observações das meninas do conjunto de dados de Potthoff e Roy (1964). O modelo é dado por

$$y_{ij} = \beta_0 + \beta_1 t_j + \delta_0 s_i + \delta_1 s_i t_j + b_{0i} + b_{1i} t_j + e_{ij}, \tag{6.1}$$

sendo y_{ij} a distância (em mm) entre o centro da pituitária e a fissura ptériomaxilar no tempo j do sujeito i; t_j é a idade mensurada em anos; β_0 e β_1 são o intercepto e a inclinação dos efeitos fixos; $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ é o efeito aleatório para o i-ésimo sujeito; e e_{ij} é o erro entre sujeitos, $i = 1, ..., 11$ e $j = 1, ..., 4$.

6.2 Influência nos efeitos fixos

Ajustamos os modelos dependente e independente considerando as distribuições t-assimétrica, normal assimétrica, t-Student e normal. Consideramos a influência de uma mudança em Δ unidades de uma observação, na distribuição *a posteriori* dos efeitos fixos. Isto é, substituímos a observação y_{ij} pelo valor contaminado $y_{ij}(\Delta) = y_{ij} + \Delta$, re-estimamos o modelo e salvamos a média *a posteriori* e o intervalo HPD de probabilidade de 95%. Neste exemplo contaminamos a quarta observação (idade de 14 anos) da primeira menina, e variamos Δ entre -20mm e 20mm com incrementos de 2mm.

As figuras 6.1 e 6.2 apresentam a média *a posteriori* e o intervalo HPD de probabilidade de 95% para as quatro distribuições para o modelo independente. Como esperado, os modelos t e t-assimétrico são menos sensíveis a mudanças de Δ. Mudanças em Δ tiveram considerável impacto,

para os modelos normal e normal assimétrico, na média *a posteriori* e no HPD. Notamos que a amplitude do intervalo aumenta quando Δ aumenta. Como resultados similares foram encontrados para o caso do modelo dependente, não apresentamos os resultados deste modelo.

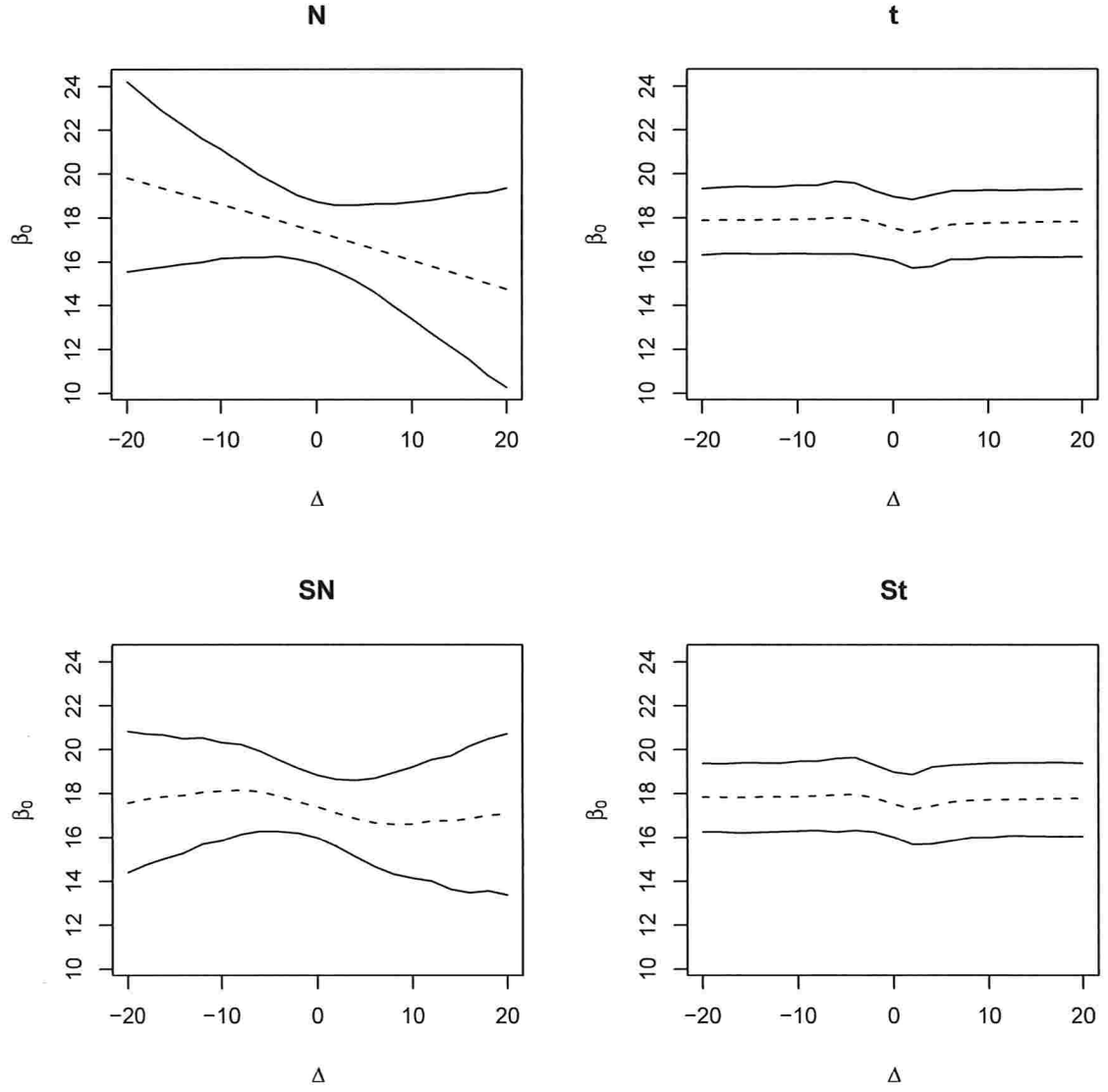

Figura 6.1: Média *a posteriori* (linha tracejada) e intervalo HPD de 95% (linha solida) para β_0 sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação

Para comparação entre os modelos dependente e independente consideramos a amplitude do intervalo HPD de probabilidade 95%. Nas figuras 6.3 e 6.4 apresentamos estes resultados, observamos que no caso do β_1 os dois modelos não apresentam grandes diferenças, mas no caso do intercepto β_0 observamos que para os distribuições t e t-assimétrica a amplitude do intervalo HPD do modelo

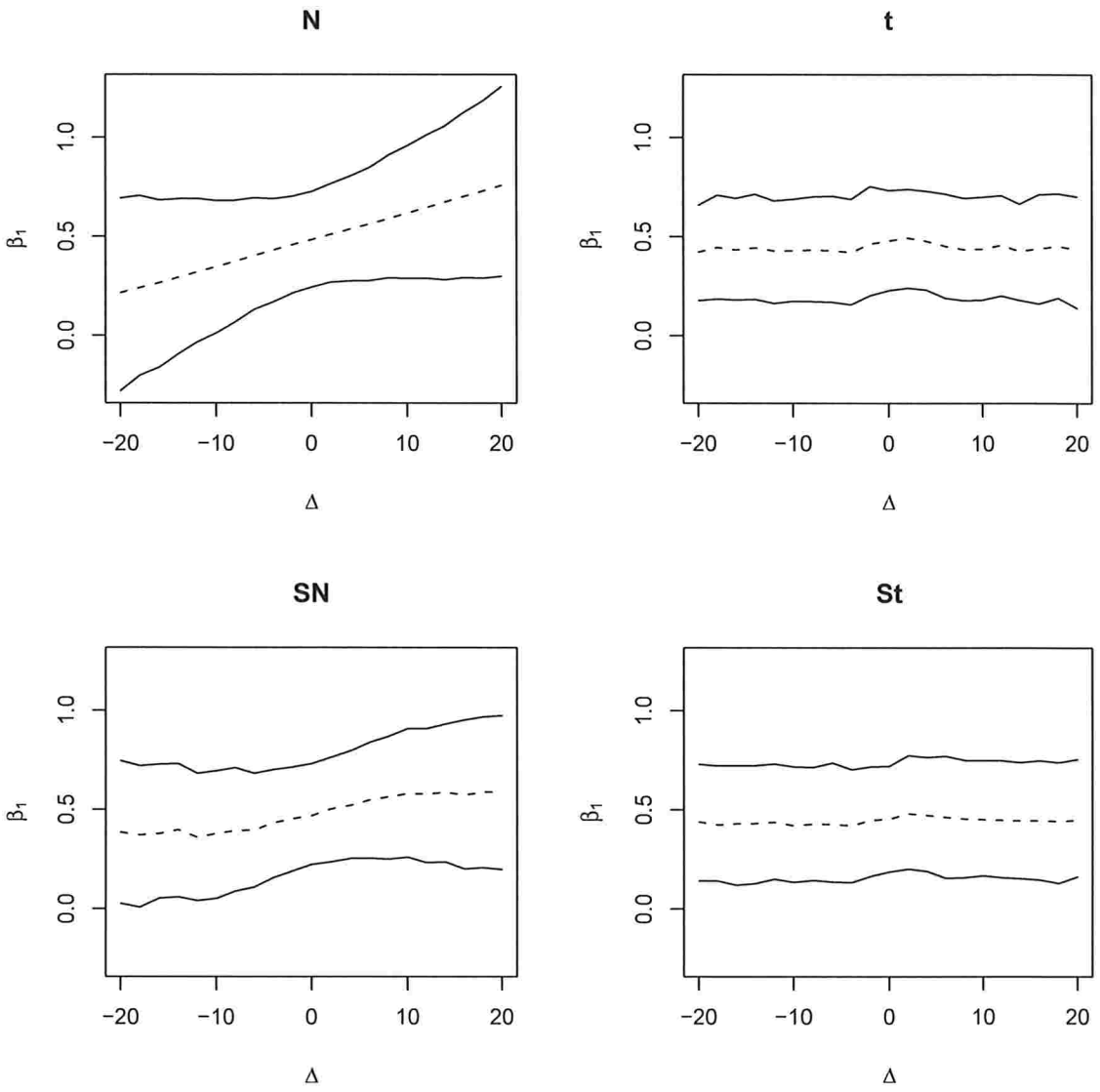

Figura 6.2: Média *a posteriori* (linha tracejada) e intervalo HPD de 95% (linha solida) para β_1 sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação

independente é menos sensível a presença de valores *outlier*.

6.3 Influência Global

Também estudamos a influência de um valor *outlier* na distribuição *a posteriori* de todos os parâmetros. Como medida de influência de uma observação aberrante consideramos a medida de divergência de Kullback–Leibler $D(\Delta)$ entre as distribuições *a posteriori* de $\boldsymbol{\theta}$ dado $\boldsymbol{Y}$ e $\boldsymbol{Y}(\Delta)$ respetivamente, sendo $\boldsymbol{Y}(\Delta)$ os dados quando substituímos a observação y_{ij} pelo valor contaminado

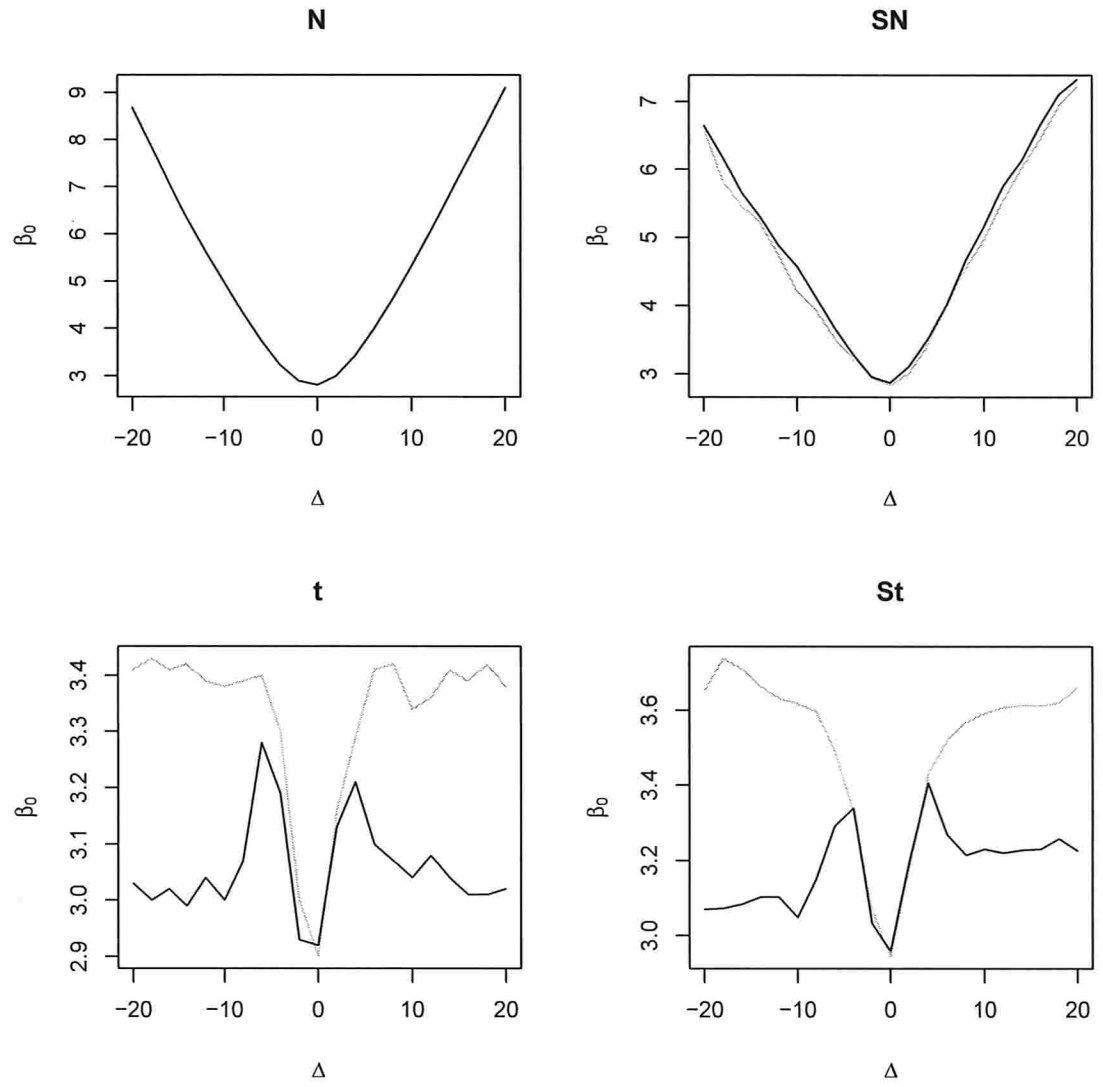

Figura 6.3: Amplitude do intervalo HPD de 95% para β_0 para o modelo dependente (linha preta) e para o modelo dependente (linha tracejada) sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação

$y_{ij}(\Delta) = y_{ij} + \Delta$.

$$D(\Delta) = \int \log\left(\frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{Y})}{\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}(\Delta))}\right) \pi(\boldsymbol{\theta} \mid \boldsymbol{Y}) d\boldsymbol{\theta}.$$

Considerando a ideias de Peng e Dey (1995) pode-se mostrar que uma estimativa de Monte

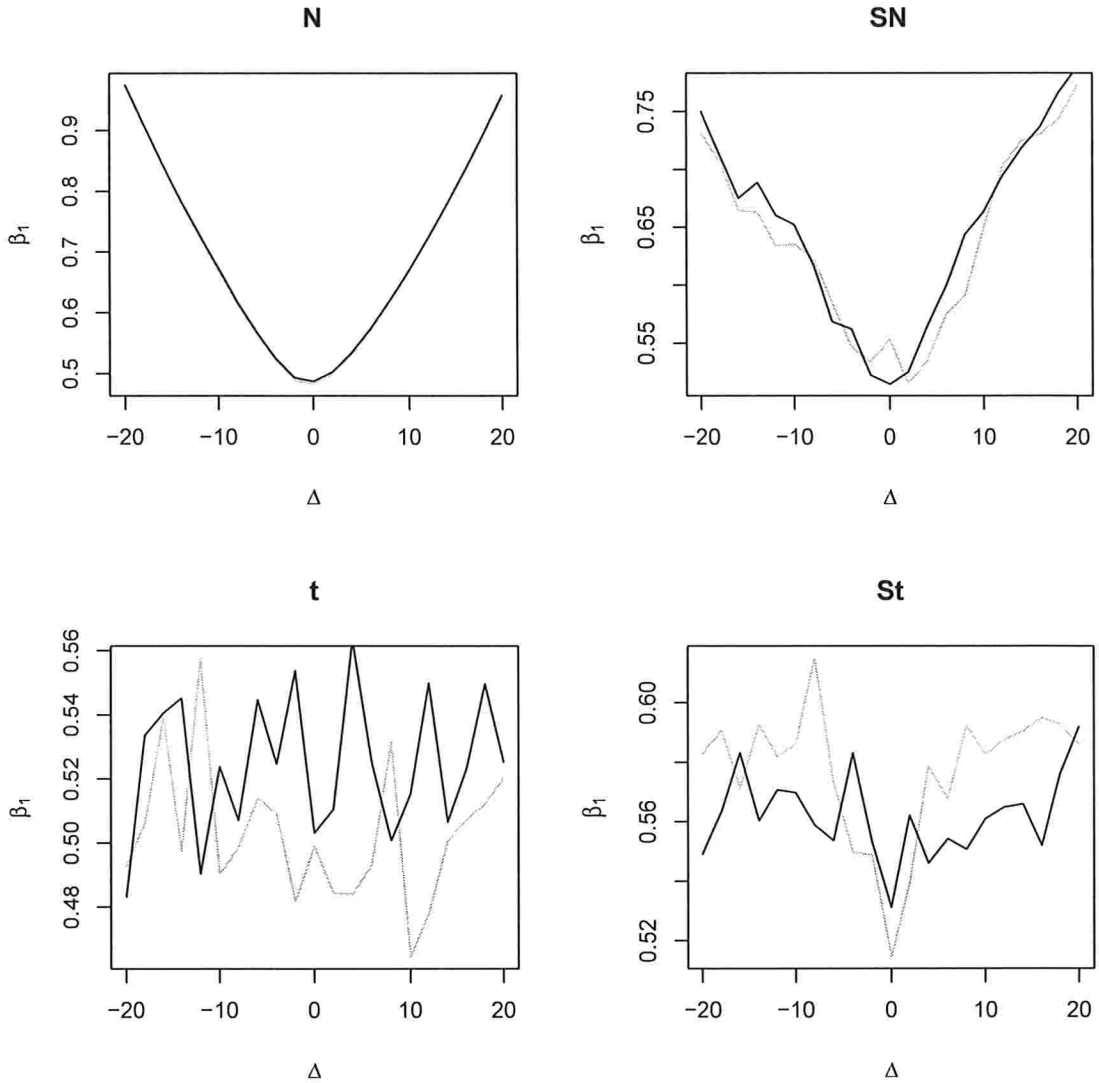

Figura 6.4: Amplitude do intervalo HPD de 95% para β_1 para o modelo dependente (linha preta) e para o modelo dependente (linha tracejada) sob as distribuições t-assimétrica, t, normal assimétrica e normal para diferentes contaminações Δ de uma observação

Carlo desta divergência é dada por

$$\widehat{D}(\Delta) = \log \frac{\{\prod_{s=1}^{B} 1/\vartheta(\boldsymbol{\theta}^s)\}^{1/B}}{\{B^{-1}\sum_{s=1}^{B} \vartheta(\boldsymbol{\theta}^s)\}^{-1}}$$

sendo $\vartheta(\boldsymbol{\theta}^s) = f(y_{ij} + \Delta \mid \boldsymbol{\theta}^s)/f(y_{ij} \mid \boldsymbol{\theta}^s)$ e $\{\boldsymbol{\theta}^s\}_{s=1}^{B}$ uma amostra da distribuição *a posteriori* de $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y})$. Na figura 6.5 apresentamos esta distancia para as quatro distribuições e para os modelos independente e dependente e , notamos que a influência de um valor *outlier* é ilimitada para os

modelos SN e N mas limitada para os modelos t e St. Também notamos que a influência é mais forte no caso dependente.

Também, podemos observar que para contaminações pequenas ($|\Delta| < 2$), todos os modelos tem a mesma curva de influência, isto ocorre devido a que a observação contaminada não esta suficientemente afastada do valor original para ser considerada como um valor *outlier*. Assim, todos os modelos tem a mesma eficiência para pequenas contaminações.

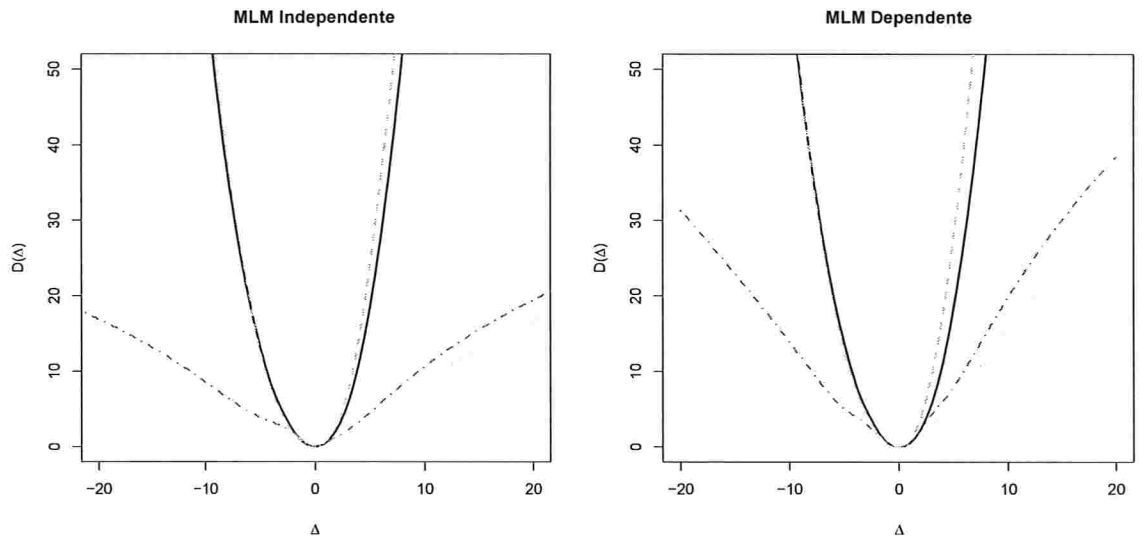

Figura 6.5: Distancia de Kullback–Leibler entre $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y})$ e $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}(\Delta))$ para os modelos: N (linha solida), SN (linha tracejada), t (linha pontilhada) e St (linha ponto-traço)

Capítulo 7

Modelo com Erros das Variáveis de Intercepto Nulo

7.1 Introdução

Em modelos de regressão com erros nas variáveis, usualmente, as covariáveis associadas com a variável resposta não são diretamente observáveis ou são observadas com erro. Nesse cenário, uma suposição usual é que a covariável é medida com erro para obter uma estimação mais precisa dos parâmetros do modelo. Modelos que regressão que consideram erros de medida são usualmente conhecidos como modelos com erros nas variáveis. Estes modelos tem sido estudado amplamente na literatura, por exemplo, Fuller (1987) em regressão linear, e Carroll, Ruppert e Stefanski (1995) para modelos de regressão não lineares. Frequentemente, modelos de intercepto nulo são utilizados para o análise de dados de ensaios clínicos onde a unidade experimental é mensuradas em dois o mais dispositivos. Aoki, Bolfarine, Achcar e Leão-Pinto (2003) estenderam estes modelos para o análise de dados longitudinais. No entanto, em todos estes modelos é assumido que a distribuição dos erros aleatório assim como a das covariáveis não observadas é normal. A suposição de normalidade pode ser muito restritiva e apresentar de falta de robustez a presença de valores *outlier*. Labra, Aoki e Bolfarine (2005) estudaram a influência local no modelo de intercepto nulo e recomendaram a utilização de uma distribuição robusta para a uma estimação mais acurada dos parâmetros do modelo. Arellano-Valle, Ozan, Bolfarine e Lachos (2005) utilizaram a distribuição normal assimétrica no modelo com erros nas variáveis linear. Recentemente Aoki, Pinto, Achcar e Bolfarine (2006) mostraram as vantagens de de utilizar modelos com misturas de normais no contexto do modelo multivariado com erros nas variáveis.

7.2 Modelo com Erros nas variáveis com intercepto nulo

O modelo com erro nas variáveis multivariado de intercepto nulo, pode ser escrito como (ver Lachos, Bolfarine, Arellano-Valle e Montenegro, 2007)

$$\mathbf{X}_i = \mathbf{x}_i + \delta_i, \tag{7.1}$$

$$\mathbf{y}_{k_i} = \mathbf{x}_i\beta_{k_i} + \epsilon_{k_i}, \tag{7.2}$$

sendo, $i = 1, \ldots, p$, $k = 1, \ldots, m$, $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})^\top$ é um vetor de covariáveis observadas, $\mathbf{y}_{k_i} = (y_{k_{i1}}, \ldots, y_{k_{in_i}})^\top$ é o vetor das variáveis resposta, $\mathbf{x}_i = (x_{i1}, \ldots, x_{in_i})^\top$ é o vetor das covariáveis

não observadas (variáveis latentes), β_{k_i} é o parâmetro de regressão, e $\boldsymbol{\epsilon}_{k_i} = (\epsilon_{k_{i1}}, \ldots, \epsilon_{k_{in_i}})^\top$, $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{in_i})^\top$ são os vetores de efeitos aleatórios.

Sob normalidade, o modelo linear especificado em (7.1)-(7.2) pode ser expresso como,

$$\mathbf{z}_{ij} = \boldsymbol{\beta}_{0i} x_{ij} + \boldsymbol{\zeta}_{ij}, \tag{7.3}$$

$$\begin{pmatrix} x_{ij} \\ \boldsymbol{\zeta}_{ij} \end{pmatrix} \sim N_{m+2}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}_i) \right), \tag{7.4}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$, onde $D(\phi_x, \boldsymbol{\phi}_i)$ denota uma matriz diagonal com elementos ϕ_x e $\boldsymbol{\phi}_i = (\sigma_u^2, \sigma_{e_i}^2, \ldots, \sigma_{e_i}^2)^\top$, $\mathbf{z}_{ij} = (X_{ij}, y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$ é um vetor de observações; o vetor de parâmetros de regressão é dado por $\boldsymbol{\beta}_{0i} = (1, \boldsymbol{\beta}_i^\top)^\top$, com $\boldsymbol{\beta}_i = (\beta_{1_i}, \ldots, \beta_{m_i})^\top$ e os efeitos aleatórios $\boldsymbol{\zeta}_{ij} = (\delta_{ij}, \epsilon_{1_{ij}}, \ldots, \epsilon_{m_{ij}})^\top$ são todos de dimensão $(m+1) \times 1$.

Neste trabalho, substituímos a suposição de normalidade em (7.4) pela distribuição t- assimétrica tipo I

$$\begin{pmatrix} x_{ij} \\ \boldsymbol{\zeta}_{ij} \end{pmatrix} \overset{\text{ind}}{\sim} \text{St-I}_{m+2}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}_i), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix}, \nu \right), \tag{7.5}$$

e nos referiremos a este modelo como St-MEM. Reescrevemos (7.5) como uma mistura de uma normal assimétrica

$$\begin{pmatrix} x_{ij} \\ \boldsymbol{\zeta}_{ij} \end{pmatrix} | w_{ij} \sim \text{SN-I}_{m+2}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, w_{ij}^{-1} D(\phi_x, \boldsymbol{\phi}_i), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix} \right) \text{ and } w_{ij} \sim Gama\,(\nu/2, \nu/2)\,, \tag{7.6}$$

$i = 1, \ldots, p$, $j = 1, \ldots, n_i$. Como a distribuição normal assimétrica em (7.6) tem a matriz de escala bloco diagonal, pode ser mostrado que condicional a w_{ij}; os $\boldsymbol{\zeta}_{ij}$ e x_{ij} são independentes e que marginalmente,

$$\boldsymbol{\zeta}_{ij} \overset{\text{ind}}{\sim} t_{m+1}(\mathbf{0}, \mathrm{D}(\boldsymbol{\phi}_i), H) \text{ and } x_{ij} \overset{\text{iid}}{\sim} StI_1(\mu_x, \phi_x, \lambda_x, H). \tag{7.7}$$

Desde que para cada $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, $\boldsymbol{\zeta}_{ij}$ e x_{ij} são indexados pelo mesmo fator w_{ij}, em geral eles não são independentes. A independência corresponde ao caso onde $w_{ij} = 1$ ($i = 1, \ldots, p, j = 1, \ldots, n_i$). Logo o modelo se reduz a caso normal assimétrico estudado em Lachos, Bolfarine, Arellano-Valle e Montenegro (2007). No entanto, pode se mostrar que $\boldsymbol{\zeta}_{ij}$ and x_{ij} são não correlacionados, isto é, $Cov(\boldsymbol{\zeta}_{ij} x_{ij}) = E[\boldsymbol{\zeta}_{ij} x_{ij} | U_{ij}] = 0$.

O parâmetro de assimetria λ_x induz assimetria na variável latente x_{ij} e consequentemente, nas quantidades observadas $\mathbf{z}_{ij}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. Assim pode ser mostrado que $\mathbf{z}_{ij}$ tem distribuição marginal t assimétrica. A distribuição marginal para a resposta $\mathbf{z}_{ij}$, é dada por

$$f(\mathbf{z}_{ij}|\boldsymbol{\theta}) = 2t_{m+1}(\mathbf{z}_{ij}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) T_1\left(\bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Sigma}_i^{-1/2} (\mathbf{z}_{ij} - \boldsymbol{\mu}_i) \left(\frac{\nu + n}{q(\mathbf{z}_{ij}) + \nu} \right)^{1/2}, \nu + 1 \right), \tag{7.8}$$

i.e., $\mathbf{z}_{ij} \overset{\text{ind}}{\sim} \text{SN-I}_{m+1}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i, H)$, sendo $\boldsymbol{\mu}_i = \boldsymbol{\beta}_{0i}\mu_x$, $\boldsymbol{\Sigma}_i = \mathrm{D}(\boldsymbol{\phi}_i) + \phi_x \boldsymbol{\beta}_{0i}\boldsymbol{\beta}_{0i}^\top$, $\bar{\boldsymbol{\lambda}}_i = \dfrac{\lambda_x \phi_x \boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\beta}_{0i}}{\sqrt{\phi_x + \lambda_x^2 \Lambda_i}}$, onde $\Lambda_i = \dfrac{\phi_x}{1 + \phi_x \boldsymbol{\beta}_{0i}^\top \mathrm{D}^{-1}(\boldsymbol{\phi})\boldsymbol{\beta}_{0i}}$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$ e $q(\mathbf{z}_{ij}) = (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i (\mathbf{z}_{ij} - \boldsymbol{\mu}_i)$.

7.3 Distribuição *a Priori*

Para completar a especificação do modelo precisamos de distribuições *a priori* para todos os parâmetros desconhecidos $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top, \sigma_{e_1}^2, \ldots, \sigma_{e_p}^2, \sigma_u^2, \mu_x, \upsilon_x^2, \tau_x, \boldsymbol{\nu}^\top)^\top$. Consideramos a seguinte distribuição *a priori*

$$\pi(\boldsymbol{\theta}) = \left(\prod_{i=1}^{p}\prod_{k=1}^{m} \pi(\beta_{k_i})\right)\left(\prod_{i=1}^{p} \pi(\sigma_{e_i}^2)\right) \pi(\sigma_u^2)\pi(\mu_x)\pi(\upsilon_x^2)\pi(\tau_x)\pi(\boldsymbol{\nu}), \tag{7.9}$$

onde $\pi(\mu_x) \sim N(\varsigma_x, \sigma_{\mu_x}^2)$, $\pi(\eta_x) \sim t(\mu_\eta, \sigma_\eta^2 \tau_x^2, \nu_\eta)$, $\pi(\tau_x^2) \sim IG(\frac{a}{2}, \frac{b}{2})$, $\pi(\beta_{k_i}) \sim N(a_{k_i}, b_{k_i})I(0,1)$, $k = 1, \ldots, m$, $i = 1, \ldots, p$, $\pi(\sigma_{e_i}^2) \sim IG(\frac{a_i}{2}, \frac{b_i}{2})$, $i = 1, \ldots, p$, onde $N(a,b)I(0,1)$ denota uma distribuição normal truncada no intervalo $[0,1]$ este truncamento sera necessário para a aplicação, $IG(a,b)$ denota a distribuição inversa gama com parâmetro de forma $a > 0$ e escala $b > 0$, com função de densidade de probabilidade dada por

$$\pi(x|a,b) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{x}\right)^{a+1} \exp\left\{-\frac{b}{x}\right\}.$$

A distribuição *a priori* de $\boldsymbol{\nu}$, é assumida ser $\pi(\boldsymbol{\nu}) = \nu^{-2} I(\nu > 3)$.

7.4 O Modelo Hierárquico e o Algoritmo MCMC

Reescrevemos o modelo em forma hierárquica considerando $(\mathbf{x}, \mathbf{u}, \mathbf{w})$ como variáveis latentes, onde $\mathbf{x} = (\mathbf{x}_1^\top, \ldots, \mathbf{x}_p^\top)^\top$, $\mathbf{u} = (\mathbf{u}_1^\top, \ldots, \mathbf{u}_p^\top)^\top$ e $\mathbf{w} = (\mathbf{w}_1^\top, \ldots, \mathbf{w}_p^\top)^\top$, com $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$ e $\mathbf{w}_i = (w_{i1}, \ldots, w_{in_i})^\top$, $i = 1, \ldots, p$. Deste modo, sob a representação estocástica dada na proposição 2.5 , com $\lambda_x = \eta_x/\tau_x$ and $\phi_x = \eta_x^2 + \tau_x^2$, segue que

$$\mathbf{z}_{ij} \mid x_{ij}, w_{ij} \overset{\text{ind}}{\sim} N_{m+1}(\boldsymbol{\beta}_{0i} x_{ij}, w_{ij}^{-1}\mathrm{D}(\boldsymbol{\phi}_i)), \tag{7.10}$$

$$x_{ij} \mid w_{ij}, u_{ij} \overset{\text{ind}}{\sim} N_1(\mu_x + \eta_x u_{ij}, w_{ij}^{-1}\tau_x), \tag{7.11}$$

$$u_{ij} \overset{\text{iid}}{\sim} \mathrm{N}_1(0, w_{ij}^{-1}) I(u_{ij} > 0), \tag{7.12}$$

$$w_{ij} \overset{\text{ind}}{\sim} Gama(\nu/2, \nu/2), \tag{7.13}$$

A distribuição de $(\mathbf{z}, \mathbf{x}, \mathbf{t}, \mathbf{u})$ é dada por

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}, \mathbf{w}, \mathbf{w}) \propto\ & (\upsilon_x^2)^{-n/2} \prod_{i=1}^{p} (|D(\boldsymbol{\phi}_i)|)^{-n_i/2} \prod_{i=1}^{p} \prod_{j=1}^{n_i} (w_{ij}^{m/2+1} g(w_{ij}|\nu/2, \nu/2)) \times \\ & \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} w_{ij} (\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i} x_{ij})^\top D^{-1}(\boldsymbol{\phi}_i)(\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i} x_{ij}) \right. \\ & \left. -\frac{1}{2\tau_x^2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} w_{ij} (x_{ij} - \mu_x - \eta_x u_{ij})^2 - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} w_{ij} u_{ij}^2 \right\}, \end{aligned}$$

onde $n = \sum_{i=1}^{p} n_i$ e $g(.\ |, a, b)$ representa a densidade de uma distribuição gama com parâmetro de forma a e de escala b.

O produto da função de verossimilhança aumentada (7.14) e da distribuição *a priori*, fornece a distribuição *a posteriori* para $\boldsymbol{\theta}$ que é dada por,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{t}, \mathbf{x}, \mathbf{u}|\mathbf{z}) \propto\ & (\tau_x^2)^{-n/2} (\sigma_u^2)^{-n/2} \prod_{i=1}^{p} (\sigma_{e_i}^2)^{-n_i/2} \prod_{i=1}^{p} \prod_{j=1}^{n_i} (w_{ij}^{m/2+1} g(w_{ij}|\nu/2, \nu/2)) \times \\ & \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \frac{w_{ij} (\mathbf{y}_{ij} - \boldsymbol{\beta}_i x_{ij})^\top (\mathbf{y}_{ij} - \boldsymbol{\beta}_i x_{ij})}{\sigma_{e_i}^2} - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} \frac{w_{ij} (X_{ij} - x_{ij})^2}{\sigma_u^2} \right. \\ & \left. -\frac{1}{2\tau_x^2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} w_{ij} (x_{ij} - \mu_x - \eta_x t_{ij})^2 - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n_i} u_{ij} u_{ij}^2 \right\} \pi(\boldsymbol{\theta}), \end{aligned} \tag{7.14}$$

onde $\mathbf{y}_{ij} = (y_{1_{ij}}, \ldots, y_{m_{ij}})^\top$. A distribuição (7.14) é intratável analiticamente. Métodos MCMC, como o amostrador de Gibbs, podem ser utlizados para se obter amostras da distribuição *a posteriori*. Para completar a especificação do amostrador de Gibbs, precisamos das distribuições condicionais completas as quais são dadas por

(i) $\quad u_{ij}|\boldsymbol{\theta}, \mathbf{z}, \mathbf{x}, \mathbf{w} \sim N\left(\frac{\eta_x (x_{ij} - \mu_x)}{\eta_x^2 + \upsilon_x^2}, u_{ij}^{-1} \frac{\tau_x^2}{\eta_x^2 + \tau_x^2} \right) I(u_{ij} > 0),$

(ii) $\quad x_{ij}|\boldsymbol{\theta}, \mathbf{z}, \mathbf{u}, \mathbf{w} \sim N\left(\mu_{x_{ij}}, w_{ij}^{-1} \sigma_{x_{ij}}^2\right)$, sendo

$$\mu_{x_{ij}} = \frac{\mu_x + \eta_x u_{ij} + \tau_x^2 \boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i) \mathbf{z}_{ij}}{1 + \tau_x^2 \boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i) \boldsymbol{\beta}_{0i}} \text{ and } \sigma_{x_{ij}}^2 = \frac{\tau_x^2}{1 + \tau_x^2 \boldsymbol{\beta}_{0i}^\top D^{-1}(\boldsymbol{\phi}_i) \boldsymbol{\beta}_{0i}};$$

(iii) $\quad \beta_{k_i}|\boldsymbol{\theta}_{(-\beta_{k_i})}, \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{w} \sim N\left(\mu_{\beta_{k_i}}, \frac{b_{k_i} \sigma_{e_i}^2}{b_{k_i} \sum_{j=1}^{n_i} w_{ij} x_{ij}^2 + \sigma_{e_i}^2} \right) I(0, 1)$, sendo

$$\mu_{\beta_{k_i}} = \frac{b_{k_i} \sum_{j=1}^{n_i} w_{ij} x_{ij} y_{k_{ij}} + a_{k_i} \sigma_{e_i}^2}{b_{k_i} \sum_{j=1}^{n_i} w_{ij} x_{ij}^2 + \sigma_{e_i}^2};$$

(iv) $\sigma^2_{\epsilon_i}|\boldsymbol{\theta}_{(-\sigma^2_{\epsilon_i})}, \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{w} \sim IG\left(\frac{n_i + a_i}{2}, b_{\sigma^2_{\epsilon_i}}\right)$, sendo

$$b_{\sigma^2_{\epsilon_i}} = \frac{1}{2}\left(\sum_{j=1}^{n_i} w_{ij}(\mathbf{y}_{ij} - \boldsymbol{\beta}_i x_{ij})^\top(\mathbf{y}_{ij} - \boldsymbol{\beta}_i x_{ij}) + b_i\right);$$

(v) $\sigma^2_\delta|\boldsymbol{\theta}_{(-\sigma^2_\delta)}, \mathbf{z}, \mathbf{x}, \mathbf{t}, \mathbf{u} \sim IG\left(\frac{n + a_u}{2}, \frac{1}{2}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_{ij}(X_{ij} - x_{ij})^2 + b_u\right)\right)$;

(vi) $\mu_x|\boldsymbol{\theta}_{(-\mu_x)}, \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{w} \sim N\left(\mu_{\mu_x}, \frac{\tau_x^2\sigma^2_{\mu_x}}{\sigma^2_{\mu_x}\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_{ij} + \tau_x^2}\right)$, sendo

$$\mu_{\mu_x} = \frac{\sigma^2_{\mu_x}\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_{ij}(x_{ij} - \eta_x u_{ij}) + \tau_x^2\varsigma_x}{\sigma^2_{\mu_x}\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_{ij} + \tau_x^2};$$

(vii) $\tau_x^2|\boldsymbol{\theta}_{(-\tau_x^2)}, \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{w} \sim IG\left(\frac{n + a}{2}, b_{\upsilon_x^2}\right)$, sendo

$$b_{\tau_x^2} = \frac{1}{2}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} w_{ij}(x_{ij} - \mu_x - \eta_x u_{ij})^2 + b\right);$$

$(viii)$ $\eta_x|\boldsymbol{\theta}_{(-\eta_x)}, \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{w} \sim N\left(\frac{\kappa_2\sigma_\eta^2 + \tau_x^2\mu_\eta}{\kappa_1\sigma_\eta^2 + \tau_x^2}, \frac{\tau_x^2\sigma_\eta^2}{\kappa_1\sigma_\eta^2 + \tau_x^2}\right)$, com

$\kappa_1 = \sum_{i=1}^{p}\sum_{j=1}^{n_i} w_{ij}u_{ij}^2$, $\kappa_2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i} u_{ij}(x_{ij} - \mu_x)$.

(ix) $\lambda_x = \eta_x/\tau_x$ and $\sigma_x^2 = \eta_x^2 + \tau_x^2$.

(x) $u_{ij}|\boldsymbol{\theta}, \mathbf{z}, \mathbf{x}, \mathbf{t} \sim Gamma(\frac{m+\nu+2}{2}; \frac{1}{2}(\nu + A_{ij} + B_{ij} + t_{ij}^2))$,
com $A_{ij} = (\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i}x_{ij})^\top D^{-1}(\boldsymbol{\phi}_i)(\mathbf{z}_{ij} - \boldsymbol{\beta}_{0i}x_{ij})$ e $B_{ij} = \frac{(x_{ij} - \mu_x - \tau_x t_{ij})^2}{\upsilon_x^2}$

(xi) $\nu|\boldsymbol{\theta}_{(-\eta_x)}, \mathbf{z}, \mathbf{x}, \mathbf{u}, \mathbf{w} \propto \pi_1(\nu) \times Gama(\frac{n\nu}{2} + 1, \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(u_{ij} - \log u_{ij}))I_{(3,\infty)}$, sendo $\pi_1(\nu) = (2^{\nu/2}\Gamma(\nu/2))^{-n}$.

7.5 Aplicação

Nesta seção ilustramos o modelo com um conjunto de dados reais de um ensaio clínico odontológico (Hadgu e Koch, 1999). Este conjunto de dados tem sido previamente analisado por Aoki et al. (2003) onde 109 voluntários adultos de sexo masculino e feminino com placa dental preexistente foram alocados aleatoriamente para o enxágue bucal de controle (C) ou qualquer um dos enxagues bucais experimentais denotados por A e B. O objetivo principal deste estudo foi comparar a eficiência em inibir o desenvolvimento de placa dental de A e B com o controle. Os sujeitos foram avaliados no índice de placa dental no *baseline* (x). Desde que esta covariável em medida em forma

imprecisa, uma forma de analisar estes dados é assumir um modelo com erros nas variáveis. Não incluímos interceptos no modelo porque é razoável assumir que se no início do estudo o índice de placa dental é zero então implica um valor de zero nos valores depois do estudo, isto é, assumimos que a placa dental não deve se incrementar depois da utilização de enxagues bucais. Assim consideramos o modelo descrito em (7.1)-(7.2) para modelar este dados. Outro aspecto deste conjunto de dados é que as quantidades observadas não são simétricas e que o índice de placa dental pode precisar de uma transformação para obter uma melhor a aproximação ao modelo normal (ver figura 7.1). No entanto, transformação de variáveis pode ser não recomendado neste cenário.

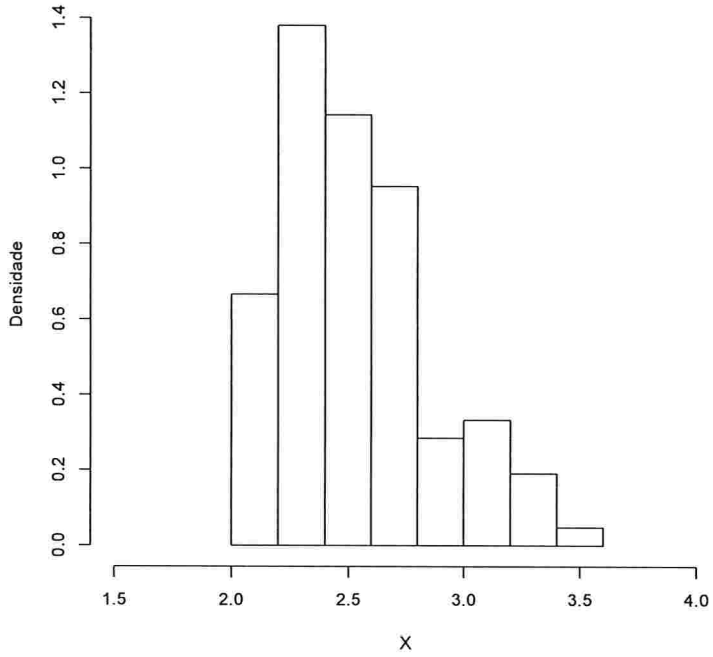

Figura 7.1: Histograma do índice de placa dental no *baseline*

A variável resposta de interesse neste contexto é verdadeiro índice de placa dental. Deste modo, $y_{1_{ij}}$ and $y_{2_{ij}}$ representam respectivamente, o índice de placa observado do enxágue i $(= 1, 2, 3)$ para o sujeito j $(= 1, \cdots, n_i)$ depois de três e seis meses do começo do estudo. Assim, $i = 1, 2, 3$ representam o enxágue control, o enxágue experimental A e o enxágue experimental B, respectivamente e $k = 1, 2$ representam três e seis meses depois do *baseline*, respectivamente. Seja, X_{ij} o índice de placa observado no *baseline* e x_{ij} o índice de placa real que é não observado. Deste modo, baseados nestas observações, consideramos o seguinte modelo para o índice de placa

observado:

$$\begin{aligned} \text{baseline observado } X_{ij} &= \text{verdadeiro baseline não observado } x_{ij} + \delta_{ij}, & (7.15)\\ y_{k_{ij}} &= \beta_{k_i} \times \text{ baseline não observado } x_{ij} + \epsilon_{k_{ij}}. & (7.16) \end{aligned}$$

Uma suposição usual é $\delta_{ij} \overset{\text{iid}}{\sim} N_1(0, \sigma^2)$, $\epsilon_{kij} \overset{\text{ind}}{\sim} N_1(0, \sigma^2_{e_i})$, δ_{ij} and ϵ_{kij} são não correlacionados e independentes das quantidades não observadas $x_{ij} \overset{\text{iid}}{\sim} N_1(\mu_x, \sigma^2_x)$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$ $k = 1, \ldots, m$. Apesar da suposição de normalidade ser razoável em muitas situações, esta pode não ser apropriada neste caso porque os dado apresentam um afastamento da suposição de normalidade. Neste cenário consideramos a distribuição t-assimétrica que permite acomodar dados que apresentam assimetria e curtose.

Utilizamos o amostrador de Gibbs (Gelfand and Smith, 1990) para obter amostras da distribuição, para cada modelo geramos 120000 amostras, assumi-se um *burn-in* de 20000. Tomamos valores de 10 em 10 para reduzir a autocorrelação entre as cadeias e obtermos uma amostra de Monte Carlo de tamanho 10000 aproximadamente independente.

Vários modelos estatísticos foram considerados com diferentes distribuições para a covariável não observada e os erros aleatórios. Estes modelos são:

Modelo 1: Normal para a variável latente e os erros (N).

Modelo 2: Normal assimétrica para a variável latente e normal para os erros (SN).

Modelo 3: t assimétrica para a variável latente e t para os erros (St).

Na tabela 7.1 apresentamos as medidas de comparação dos modelos. Notemos que os modelos assimétricos são melhores que o modelo normal em todos os critérios apresentados. O pseudo fator de Bayes (pBF) é calculado com referência ao modelo normal. O melhor modelo segundo estes critérios é o modelo normal assimétrico.

Tabela 7.1: Critérios de comparação de modelos.

Critério	N	SN	St
EAIC	443.97	429.80	435.14
EBIC	475.82	464.30	472.30
pBF	1	3283.94	396.39

Na tabela 7.2, apresentamos a média *a posteriori*, desvio padrão e o intervalo HPD de 95% de probabilidade para os modelo SN e St. Na figura 7.2 é mostrado o boxplot para o parâmetro de assimetria nos modelos SN e St. Notamos que o intervalo de credibilidade não inclue o zero para os dois modelos o que confirma a assimetria positiva presente nos dados.

O objetivo do experimento foi comparar a eficiência dos dois enxagues bucais experimentais, A e B, com o enxágue controle. Deste modo, estamos interessados em comparar os parâmetros de inclinação β_{k_2} e β_{k_3} com respeito a β_{k_1}, $k = 1, 2$. Uma forma de testar esta hipotese ($\beta_{k_1} = \beta_{k_i}$, $i = 2, 3$, $k = 1, 2$), é se consideramos os estimados obtidos pelo amostrador de Gibbs de $\beta_{k_1} - \beta_{k_2}$ e $\beta_{k_1} - \beta_{k_3}$, $k = 1, 2$, e verificamos se o intervalo de credibilidade contem o valor de zero. Uma outra

Tabela 7.2: Média *a posteriori*, desvio padrão e intervalo HPD de 95% para os modelos SN e St.

	SN				St			
Parameter	mean	D.P.	2.5%	97.5%	mean	D.P.	2.5%	97.5%
β_{1_1}	0.7026	0.0351	0.6336	0.7717	0.6999	0.0362	0.6288	0.7706
β_{1_2}	0.5365	0.0456	0.4469	0.6260	0.5379	0.0455	0.4485	0.6273
β_{1_3}	0.5086	0.0327	0.4444	0.5731	0.5069	0.0335	0.4413	0.5727
β_{2_1}	0.6864	0.0352	0.6170	0.7556	0.6878	0.0356	0.6177	0.7577
β_{2_2}	0.5025	0.0455	0.4131	0.5923	0.4943	0.0458	0.4044	0.5845
β_{2_3}	0.4138	0.0327	0.3494	0.4779	0.4165	0.0332	0.3512	0.4815
$\sigma^2_{e_1}$	0.2919	0.0516	0.2076	0.4085	0.2777	0.0527	0.1911	0.3968
$\sigma^2_{e_2}$	0.4554	0.0838	0.3198	0.6471	0.4152	0.0835	0.2788	0.6051
$\sigma^2_{e_3}$	0.2381	0.0416	0.1699	0.3321	0.2302	0.0429	0.1600	0.3270
σ^2_u	0.0082	0.0046	0.0024	0.0200	0.0083	0.0047	0.0025	0.0203
μ_x	2.1357	0.0435	2.0591	2.2226	2.1478	0.0440	2.0695	2.2425
ϕ_x	0.2733	0.0524	0.1825	0.3867	0.2423	0.0523	0.1510	0.3567
λ_x	5.1212	1.9449	2.1986	9.7594	4.6240	1.8344	1.8817	8.9878
ν					26.4126	12.8229	9.7079	58.5798

questão de interesse é se os enxagues bucais continuam a reduzir o índice de placa depois de três meses o que equivale a comparar os parâmetros de inclinação β_{1_i} and β_{2_i} $i = 1, 2, 3$. Na tabela 7.3 apresentamos os valores da média e o intervalo HPD de 95% para as quetões de interesse descritas, assim como, a comparação dos enxagues bucais A and B (β_{k_2} and β_{k_3}).

Depois de analisar estes resultados concluimos que ambos enxagues bucais experimentais são mais eficientes que o controle para reduzir a placa dental depois de três meses desde o *baseline* e também depois de seis meses, no entanto, não existem diferenças significativas entre os dois enxagues experimentais A e B. O único enxágue que tem efeito duradouro é o enxágue B.

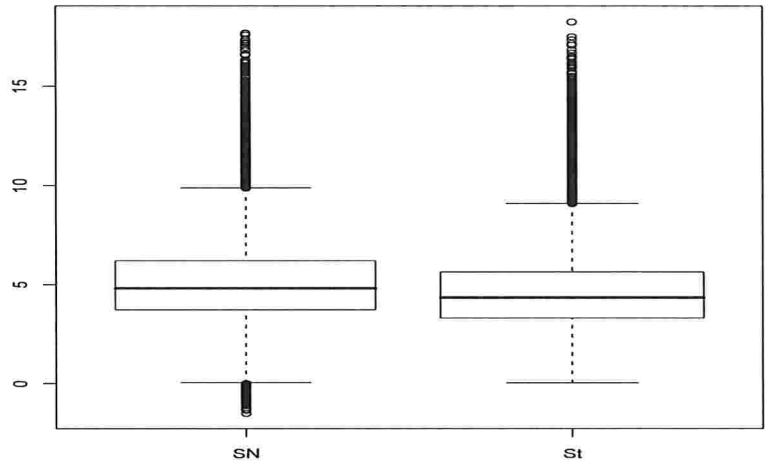

Figura 7.2: Box plot de λ_x.

Tabela 7.3: Média *a posteriori* e HPD de 95% para o Modelo 2

$\beta_{1_1} - \beta_{1_2}$	$\beta_{1_1} - \beta_{1_3}$	$\beta_{1_2} - \beta_{1_3}$	$\beta_{2_1} - \beta_{2_2}$	$\beta_{2_1} - \beta_{2_3}$
0.1649	0.1925	0.0276	0.1830	0.2713
(0.0539 , 0.2810)	(0.0976 , 0.2861)	(-0.0816 , 0.1376)	(0.0714 , 0.2964)	(0.1770 , 0.3639)
$\beta_{2_2} - \beta_{2_3}$	$\beta_{1_1} - \beta_{2_1}$	$\beta_{1_2} - \beta_{2_2}$	$\beta_{1_3} - \beta_{2_3}$	
0.0884	0.0162	0.0343	0.0950	
(-0.0206 ,0.1952)	(-0.0767 , 0.1134)	(-0.0912 , 0.1604)	(0.0057 , 0.1875)	

7.6 Estudo de Robustez

A robustez do modelo *St* pode ser estudada considerando a influência de uma observação *outlier* na distribuição dos parâmetros. Para simplificar, consideremos somente o modelo para o enxágue control ($i = 1$).

$$\text{baseline observado } X_{1j} = \text{verdadeiro baseline não observado } x_{1j} + \delta_{1j} \tag{7.17}$$

$$y_{k_{1j}} = \beta_{k_1} \times \text{verdadeiro baseline não observado } x_{1j} + \epsilon_{k_{1j}}. \tag{7.18}$$

Estudamos a influência de uma mudança de Δ unidades em uma observação na média *a posteriori* e no intervalo HPD de 95% de β_{k_1}, $k = 1, 2$. Substituímos $y_{k_{1j}}$ por $y_{k_{1j}}(\Delta) = y_{k_{1j}} + \Delta$. Deste modo, contaminamos a observação $y_{1_{1j}}$, isto é, o índice de placa observado depois de três meses, e variamos Δ entre -10 and 10.

Nas figuras 3 e 4, apresentamos a média *a posteriori* e HPD de 95% de probabilidade para β_{1_1} e β_{2_1} respectivamentes, para os modelos N, SN, St. Como esperado, β_{1_1} é mais afetado que β_{2_1}, desde que perturbamos $y_{1_{1j}}$. Notemos também que o modelo St é menos afetado pelas variações de Δ que os modelos N e SN. Nos modelos N e SN, uma única observação *outlier* tem considerável impacto na amplitude dos intervalo HPD para ambosβ_{1_1} and β_{2_1}, e na média *a posteriori* de β_{1_1}.

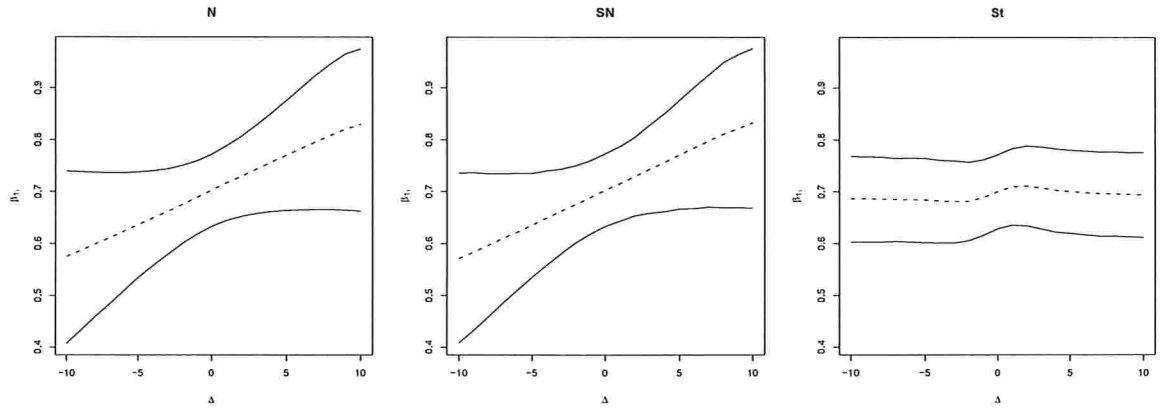

Figura 7.3: Média *a posteriori* (linha tracejada) e HPD de 95% de probabilidade (linha sólida) para β_{1_1} dos modelos N, SN, St para diferentes contaminações Δ de uma observação

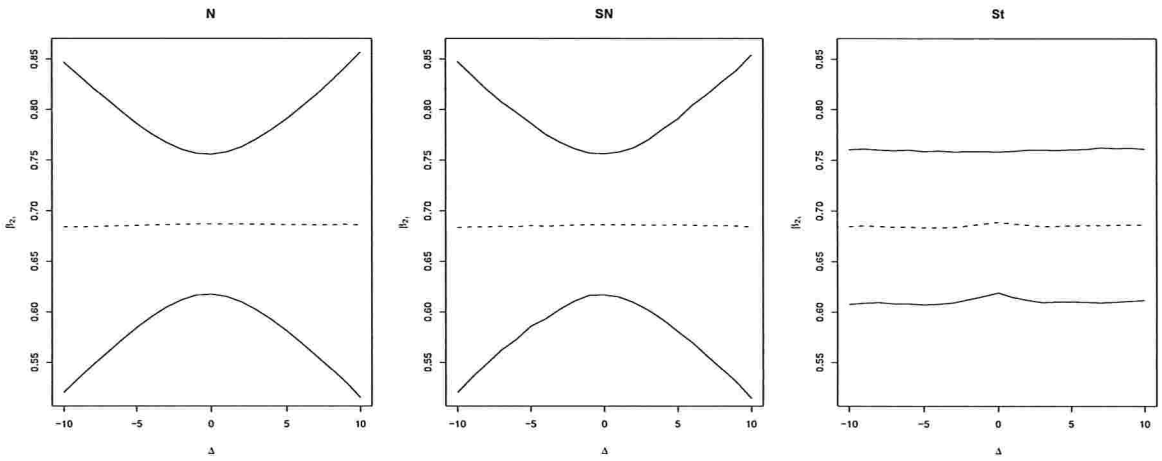

Figura 7.4: Média *a posteriori* (linha tracejada) e HPD de 95% de probabilidade (linha sólida) para β_{2_1} dos modelos N, SN, St para diferentes contaminações Δ de uma observação

Capítulo 8

Problemas de inferência em modelos t-assimétricos multivariados

O trabalho de Chib, Osiewalski e Steel (1991) discute a impossibilidade de atualização da distribuição à posteriori dos graus de liberdades em modelos de regressão com erros t multivariados. Tal resultado, coloca em questão o uso do modelo de regressão dependente t multivariado. Neste capítulo estendemos esse resultado ao considerar um modelo de regressão não-linear com erros com distribuição t-assimétrica multivariada. Sob condições bastante brandas impostas a distribuição *a priori* provamos que distribuição *a posteriori* dos graus de liberdade não é atualizada pela informação amostral.

8.1 Modelo

O modelo de regressão não linear é representado por

$$y = h(X, \beta) + \epsilon \tag{8.1}$$

sendo, X matriz $n \times r$, β um vetor de coeficientes de regressão, $h(X, \beta)$ uma função vetorial de (X, β), e

$$\epsilon \sim StF\left(0, \frac{1}{\tau}V(X, \eta), \mathbf{\Delta}, \nu\right) \tag{8.2}$$

Consideraremos as seguintes suposições:

1. Seja $\omega = (\beta, \mathbf{\Delta}, \tau, \eta, \nu)$ e $\vartheta = (\beta, \mathbf{\Delta}, \eta)$, onde $\beta \in B \subseteq \Re^k$, $\eta \in H \subseteq \Re^q$, $\vartheta \in \Theta \subseteq B \times H$, $\tau \in \Re^+$, $\mathbf{\Delta}$ matriz $n \times m$ tal que $\|\mathbf{\Delta} a\| < 1$ $\forall a \in \Re^m$ e $\nu \in \Re^+$.

2. $h(X, \beta) : n \times 1$ é uma função conhecida de X e β.

3. $V(X, \eta) : n \times n$ é uma matriz definida positiva, e função conhecida de X e η.

4. X é uma matriz aleatória tal que a densidade conjunta de $p(X, \omega)$ se fatora como $p(X)p(\omega)$, sendo $p(\omega)$ a densidade *a priori* para os parâmetros. Também assumimos que $p(X)$ é própria.

Logo, $\boldsymbol{y} \sim StF\left(h(X,\beta), \frac{1}{\tau}V(X,\eta), \boldsymbol{\Delta}, \nu\right)$ e da proposição 2.1 temos que

$$\begin{aligned} \boldsymbol{y} \mid X, z, \omega &\sim SNF\left(h(X,\beta), \frac{1}{\tau z}V(X,\eta), \boldsymbol{\Delta}\right) \\ z \mid X, \omega = z, \nu &\sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \end{aligned} \tag{8.3}$$

Notamos que a densidade de mistura, $p(z \mid X, \omega)$, é independente de X e de todos os parâmetros com exceção de ν. Assim podemos escrever o seguinte modelo bayesiano conjunto (dado X)

$$p(\boldsymbol{y}, z, \omega \mid X) = p(\boldsymbol{y} \mid X, z, , \omega)p(z \mid \nu)p(\omega).$$

8.2 A Distribuição *à Posteriori*

Pelo teorema de Bayes, a distribuição *a posteriori* de ν é dada por

$$p(\nu \mid \boldsymbol{y}, X) = \int p(\omega \mid \boldsymbol{y}, X) d\tau d\vartheta$$

sendo $p(\omega \mid \boldsymbol{y}, X) = \int p(\boldsymbol{y}, z, \omega \mid X) / \int p(\boldsymbol{y}, z, \omega \mid X) d\omega dz$ a distribuição *a posteriori* de ω. Consideramos a seguinte transformação,

$$(\tau, z) \rightarrow (\kappa, z), \quad \kappa = \tau z \tag{8.4}$$

com Jacobiano igual a $1/z$. Notamos que (8.3) e (8.4) nos permitem concluir que $\boldsymbol{y}$ e (z, ν) são independentes condicional a X, ϑ e κ:

$$\boldsymbol{y} \perp\!\!\!\perp (z, \nu) \mid X, \vartheta, \kappa. \tag{8.5}$$

Da suposição 4 e de (8.3) temos que

$$X \perp\!\!\!\perp (z, \omega). \tag{8.6}$$

Logo, de (8.4) e (8.6),

$$X \perp\!\!\!\perp (z, \vartheta, \kappa, \nu). \tag{8.7}$$

Utilizando as propriedades da independência condicional podemos inferir de (8.5) e (8.7) que $\boldsymbol{y} \perp\!\!\!\perp \nu \mid X, \vartheta, \kappa$ e $X \perp\!\!\!\perp \nu \mid \vartheta, \kappa$, assim podemos escrever

$$\nu \perp\!\!\!\perp (\boldsymbol{y}, X) \mid \vartheta, \kappa. \tag{8.8}$$

Esta ultima equação indica claramente que se ν e (ϑ, κ) são independentes, então $(X, \boldsymbol{y})$ e ν são independentes incondicionalmente, e a distribuição *a posteriori* marginal de ν é idêntica a sua distribuição marginal *a priori*. Isto é, neste caso os dados não podem modificar a nossa opinião *a priori* com respeito a ν. Este resultado é dado no seguinte teorema,

Teorema 8.1 *Sob as suposições* $1-4$ *com distribuição a priori própria para* ω, *se*

$$\nu \perp\!\!\!\perp (\vartheta, \kappa)$$

onde κ *é definida em (8.4), então*

$$\nu \perp\!\!\!\perp (\boldsymbol{y}, X).$$

Capítulo 9

Considerações Finais

Nestes trabalho foi desenvolvida a distribuição t-assimétrica fundamental que inclui a distribuição t assimétrica tipo I (Azzalini e Dalla Valle, 1996) e a distribuição t assimétrica tipo II (Sahu et al., 2003); e baseados nas propriedades desta distribuição analisamos alguns modelos lineares do ponto de vista bayesiano.

No capítulo 2, descrevemos a distribuição t-assimétrica fundamental e desenvolvemos algumas de suas propriedades baseados no resultados de Arellano-Valle e Genton (2005). Também são estudados como casos especiais desta distribuição como a distribuição assimétrica tipo I e a distribuição assimétrica tipo II. Finalmente, analisamos a interpretação do parâmetro de assimetria nestes casos especiais. Concluímos que as parametrizações apresentadas por Azzalini e Capitanio (1999) e Branco e Dey (2001) não fornecem informação sob a assimetria das marginais.

No capítulo 3, apresentamos o modelo de regressão multivariado considerando a distribuição t-assimétrica tipo II, desenvolvemos o algoritmo de Gibbs para implementar a inferência bayesiana neste modelo. Estudamos vários critérios de informação e mostramos que os critérios EAIC, EBIC e pBF podem ser utilizados para detectar possíveis afastamentos da suposição de normalidade e de simetria, destes critérios os melhores resultados foram apresentados por o EAIC e o EBIC. Concluímos que, o DIC em geral tende a favorecer modelos mais complexos.

No capítulo 4, consideramos uma nova classe de modelos lineares mistos, considerando que os erros e os efeitos aleatórios seguem em forma conjunta uma distribuição t-assimétrica tipo II. Como os erros e os efeitos aleatórios não são mais independentes chamamos de este modelo de linear misto t-assimétrico dependente. Estudamos as propriedades deste modelo e implementamos o algoritmo de Gibbs. Na aplicação utilizamos os dados de colesterol de Framingham.

No capítulo 5, estendemos o modelo linear misto considerando que os erros e os efeitos aleatórios são independentes e seguem uma distribuição t-assimétrica tipo I e denominamos este modelo como modelo de linear misto t-assimétrico independente. Estudamos as propriedades deste modelo e implementamos o algoritmo de Gibbs. Realizamos uma aplicação do modelo para um conjunto de dados de toxicologia, fizemos um estudo de sensibilidade que em geral mostrou que os efeitos fixos são insensíveis a escolha da distribuição *a priori* dos parâmetros de forma: assimetria e graus de liberdade. No entanto, a distribuição *a posteriori* destes parâmetros pode ser influenciada pela escolha da distribuição *a priori*.

No capítulo 6, apresentamos um estudo de robustez considerando um conjunto de dados dentários, onde comparamos os modelos dependente e independente para o modelo linear misto. Consideramos a influência de um valor *outlier* na distribuição *a posteriori* dos efeitos fixos e também na distribuição *a posteriori* de todos os parâmetros. Concluímos que os modelos normal e normal assimétrico são mais afetados pela presencia de um valor outlier que os modelos t e t-assimétrico. Também observamos que o modelo dependente é mais afetado que o modelo independente.

No capítulo 7, estendemos um modelo com erros nas variáveis de intercepto nulo considerando a distribuição t-assimétrica tipo I conjunta para os erros e as covariáveis não observadas, a assimetria é introduzida somente nas covariáveis. Estudamos as propriedades deste modelo e propomos um algoritmo de Gibbs para implementar a inferência bayesiana neste modelo. Ilustramos o modelo com uma aplicação para um conjunto de dados de um estudo odontológico.

No capítulo 8, estendemos o resultado de não atualização da distribuição *a posteriori* dos graus de liberdade num modelo de regressão não linear quando os erros conjuntamente tem distribuição t-Student (Chib et al., 1991) para a distribuição t-assimétrica fundamental, deste modo este resultado é valido para a distribuição t-assimétrica tipo I (Azzalini e Capitanio, 1999) e a tipo II (Sahu et al., 2003).

No apêndice B, estudamos a distribuição condicional *a posteriori* dos graus de liberdade e mostramos que é log-concava o que nos permite utilizar o algoritmo de rejeição adaptativa (Gilks e Wild, 1992) para gerar amostras desta distribuição.

Na literatura recente, outros autores também tem considerado a distribuição t-assimétrica para o modelo linear misto, assim temos os trabalhos de Jara, Quintana e San Martín (2008) no enfoque bayesiano e Zhou e He (2008) no enfoque clássico.

9.1 Perspectiva Futuras

Salientamos que os resultados apresentados neste trabalho podem ser facilmente estendidos para uma classe mais geral de distribuições, a de misturas na escala de uma distribuição normal assimétrica. Se consideramos a representação da distribuição t-assimétrica dada na proposição 2.1 bastaria considerar alguma outra distribuição para w, assim poderíamos definir esta classe como

$$\begin{aligned} \boldsymbol{Y} \mid w &\sim SNF_n\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Omega}}{w}, \frac{\boldsymbol{D}}{w^{1/2}}\right) \\ w &\sim f_w. \end{aligned} \tag{9.1}$$

Uma extensão para a distribuição t-assimétrica tipo I considerando o modelo de mistura na escala de uma normal assimétrica foi proposto como um exemplo em Branco e Dey (2001), algumas propriedades são estudadas em Kim (2008) e tem algumas aplicações desta classe na literatura recente, por exemplo, o modelo analisado no capítulo 7 é estudado nesta classe mais geral em Ghosh, Bayes e Lachos (2009).

Este estudo da distribuição t assimétrica fundamental nos oferece várias possibilidades para futuras pesquisas. Podemos estudar o modelo mais amplo de distribuições assimétricas proposto

em (9.1). Também consideramos aplicações desta classe de distribuições no modelo linear misto generalizado relaxando a suposição de normalidade dos efeitos aleatórios.

Apêndice A

CPO e Pseudo-Fator de Bayes

Seja $\boldsymbol{Y}$ um conjunto de observações $(y_r, r = 1, ..., n)$, $\boldsymbol{Y}_{(r)}$ denota todos os elementos de Y com exceção de y_r. A ordenada preditiva condicional CPO (*Conditional Predictive Ordinate*) é o valor da densidade condicional preditiva de validação cruzada avaliada no valor observado $y_{r,obs}$,

$$CPO_r = f(y_{r,obs} \mid Y_{(r),obs}).$$

Essa medida foi proposta inicialmente por Geisser e Eddy (1979) e estudada posteriormente por Gelfand (1996). O gráfico dos valores da CPO versus as observações, por exemplo, pode ser usado como diagnóstico para detectar *outliers*, já que um valor alto de CPO indica que a observação está bem ajustada, e um valor baixo,o contrario.

Para o cálculo do CPO considera-se a seguinte estimativa de Monte-Carlo,

$$\widehat{CPO}_r = \frac{1}{\sum_{j=1}^{M} f(y_{r,obs} \mid \boldsymbol{Y}_{(r),obs}, \theta_j)}$$

sendo $(\theta_1, \ldots, \theta_M)$ uma amostra da distribuição *a posteriori*.

Podemos utilizar como uma medida informal para a seleção de modelos o produto dos CPO, assim para cada modelo calculamos,

$$\prod_{r=1}^{n} CPO_r,$$

e escolhemos o modelo com o maior valor dessa medida. O pseudo-fator de Bayes para dois modelos competitivos M_1 e M_2 é definido por

$$PFB = \prod_{r=1}^{n} \frac{f(y_{r,obs} \mid \boldsymbol{Y}_{(r),obs}, M_1)}{f(y_{r,obs} \mid \boldsymbol{Y}_{(r),obs}, M_2)} = \prod_{r=1}^{n} \frac{CPO_r^{M_1}}{CPO_r^{M_2}},$$

valores altos do pBF indicam que os dados favorecem ao modelo M_1.

Apêndice B

Distribuição condicional dos graus de liberdade

Temos que a distribuição condicional *a posteriori* dos graus de liberdade ν para os modelos estudados neste trabalho tem a forma:

$$f(\nu \mid .) \propto \frac{1}{\Gamma(\nu/2)^m} \nu^{-2} \left(\frac{\nu}{2}\right)^{m\frac{\nu}{2}} \exp\left(-\sum_{i=1}^{m}(w_i - \log w_i)\frac{\nu}{2}\right). \tag{B.1}$$

sendo $w_i > 0$, $\forall i = 1, ..., m$.

O seguinte resultado será útil para a prova da proposição B.1

Lema B.1 *Seja $\psi^{(1)}(x)$ a função trigamma, isto é, a segunda derivada de $\log\Gamma(x)$, então temos que*

$$\psi^{(1)}(x) > \frac{1}{x} + \frac{1}{2x^2}$$

Prova : Ver o teorema 9 de Alzer (1997) para $n = 0$ e $k = 1$.

Proposição B.1 *A distribuição condicional a posteriori dos graus de liberdade ν dada em (B.1) é log-concava para $m > 4$.*

Prova : Então temos que a segunda derivada do logaritmo de $f(\nu \mid .)$ é dada por

$$\frac{\partial^2 f}{\partial \nu^2} = -\frac{m}{4}\psi^{(1)}\left(\frac{\nu}{2}\right) + \frac{2}{\nu^2} + \frac{m}{2v}$$

Logo, pelo lema B.1 temos que

$$\frac{\partial^2 f}{\partial \nu^2} < -\frac{1}{\nu^2}\left(\frac{m}{2} - 2\right).$$

O que prova a proposição B.1.

Este resultado é importante porque nos permite utilizar o algoritmo de rejeição adaptativa de Gilks e Wild (1992) para gerar amostras da distribuição condicional completa *a posteriori* dos graus de liberdade. Watanabe (2001) obtiveram um resultado similar para o modelo t-Student considerando uma distribuição *a priori* gamma para os graus de liberdade, no entanto, a forma de provar a log-concavidade é diferente a abordagem apresentada nesta seção.

Referências Bibliográficas

Aigner, D. J., Lovell, C. A. K. e Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model, *Journal of Econometrics* **12**: 12–37.

Alzer, H. (1997). On some inequalities for the gamma and psi functions, *Mathematics of Computation* **66**: 373–389.

Aoki, R., Bolfarine, H., Achcar, J. e Leão-Pinto, D. (2003). Bayesian analysis of a multivariate null intercept error-in-variables regression model, *Journal of Biopharmaceutical Statistics* **13**: 763–771.

Aoki, R., Pinto, J., Achcar, J. e Bolfarine, H. (2006). Mixture of normal distributions in multivariate null intercept measurement error model, *Journal Of Biopharmaceutical Statistics* **16**: 785–802.

Arellano-Valle, R. e Azzalini, A. (2006). On the unification opf the families of skew-normal distributions, *Scandinavian Journal fo Statistics* **33**: 561–574.

Arellano-Valle, R. B., Bolfarine, H. e Lachos, V. (2005). Skew-normal linear mixed models, *Journal of Data Science* **3**: 415–438.

Arellano-Valle, R. B., Bolfarine, H. e Lachos, V. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics* **6**: 663–682.

Arellano-Valle, R. B., Branco, M. D. e Genton, M. G. (2006). A unified view on skewed distributions arising form selections, *The Canadian Journal of Statistics* **34**: 581–601.

Arellano-Valle, R. B. e Genton, M. G. (2005). On fundamental skew distributions, *Journal of Multivariate Analysis* **96**: 93–116.

Arellano-Valle, R., Ozan, S., Bolfarine, H. e Lachos, V. (2005). Skew-normal measurement error models, *Journal of Multivariate Analysis* **96**: 93–116.

Azzalini, A. (1985). A class of distributions wich includes the normal ones, *Scandinavian Journal of Statistics* **12**: 171–178.

Azzalini, A. e Arellano-Valle, R. B. (2008). The centred parametrization for the multivariate skew-normal distribution, *Journal of Multivariate Analysis . In press.*

Azzalini, A. e Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution, *Journal of the Royal Statistical Society: Series B* **61(3)**: 579–602.

Azzalini, A. e Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society Ser. B* **65**: 367–389.

Azzalini, A. e Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika* **83**: 715–726.

Bayes, C. L. e Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution, *Brazilian Journal of Probability and Statistics* **21**: 141–163.

Branco, M. D., Bolfarine, H., Iglesias, P. e Arellano-Valle, R. B. (2000). Bayesian analysis of the calibration problem under elliptical distributions, *Journal of Statistical Planning and Inference* **90**: 69–85.

Branco, M. D. e Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis* **79**: 99–114.

Branco, M. D. e Dey, D. K. (2002). Regression model under skew elliptical error distribution, *The Journal of Mathematical Sciences* **1**: 151–169.

Brooks, S. P. (2002). Discussion on the paper by spiegelhalter, d. j., best, n. g., carlin, b. p., and van der linde, a., *Journal Royal Statistical Society, Series B* **64**: 616–618.

Carroll, R., Ruppert, D. e Stefanski, L. (1995). *Measurement Error in NonLinear Models*, Chapman & Hall.

Chib, S., Osiewalski, J. e Steel, M. F. J. (1991). Posterior inference on the degrees of freedom parameter in multivariate-t regression models, *Economics Letters* **37**: 391–397.

Dempster, A. P., Patel, C. M., Selwyn, M. R. e Roth, A. J. (1984). Statistical and computational aspects of mixed model analysis, *Applied Statistics* **33**(2): 203–214.

Fernández, C. e Steel, M. F. J. (1998). On bayesian modelling of fat tails and skewness, *Journal of The American Statistical Association* **93**: 359–371.

Fuller, W. (1987). *Measurement error models*, New York: Wiley.

Geisser, S. e Eddy, W. F. (1979). A predictive approach to model selection, *Journal of the American Statistical Association* **74**: 153–160.

Gelfand, A. E. (1996). Model determination using sampling-based methods, *in* W. Gilks, S. Richardson e D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, 1st edn, Chapman Hall, London, chapter 9, pp. 145–161.

Gelman, A., Carlin, J. e Stern, H. (2003). *Bayesian Data Analysis*, Chapman and Hall.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall/CRC, Boca Raton, FL.

Ghosh, P., Bayes, C. L. e Lachos, V. H. (2009). A robust bayesian approach to null intercept measurement error model with application to dental data, *Computational Statistics and Data Analysis* **53**: 1066–1079.

Gilks, W. R. e Wild, P. (1992). Adaptive rejection sampling for gibbs sampling, *Applied Statistics* **41**: 337–348.

Hadgu, A. e Koch, G. (1999). Application of generalized estimating equations to a dental randomized clinical trial, *Journal of Biopharmaceutical Statistics* **9**: 61–178.

Jara, A., Quintana, F. A. e San Martín, E. (2008). Linear mixed models with skew-elliptical distributions: A bayesian approach, *Computational Statistics and Data Analysis* **52**: 5033–5045.

Jones, M. C. e Faddy, M. J. (2003). A skew extension of the t-distribution, with applications, *Journal of the Royal Statistical Society: Series B* **65**(1): 159–174.

Kelker, D. (1970). Distribution theory of special distributions and location-scale parameter, *Sankhya* **32**: 319–430.

Kim, H. (2008). A note on scale mixtures of skew normal distribution, *Statistics and Probability Letters.* **78**: 1694–1701.

Labra, F., Aoki, R. e Bolfarine, H. (2005). Local influence in null intercept measurement error regression under a student-t model, *Journal Applied Statistics* **32**: 723–740.

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. e Montenegro, L. C. (2007). Likelihood based inference for multivariate skew-normal regression models, *Communications in Statistics: Theory and Methods* **36**: 1769–1786.

Laird, N. L. e Ware, J. H. (1982). Random-effects model for longitudinal data, *Biometrics* **38**: 963–974.

Lange, K. L., Little, R. J. A. e Taylor, J. M. G. (1989). Robust statistical modeling using the t-distribution, *Journal of the American Statistical Association* **84**: 881–896.

Liseo, B. e Loperfido, N. (2004). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* to appear.

Liu, C. (1995). Missing data imputation using the multivariate t distribution, *Journal of Multivariate Analysis* **53**: 139–158.

Ma, Y. e Genton, M. (2004). A flexible class of skew-symmetric distributions, *Scandinavian Journal of Statistics* **31**: 459–468.

Ma, Y., Genton, M. e Davidian, M. (2004). Linear mixed effects models with flexible generalized skew-elliptical random effects, *in* M. Genton (ed.), *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*, Chapman & Hall/CRC, Boca Raton, FL, pp. 339–358.

O'Hagan, A. e Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints, *Biometrika* **63**: 201–202.

Osiewalski, J. e Steel, M. F. J. (1993). Robust bayesian inference in elliptical regression models, *Journal of Econometrics* **57**: 345–363.

Peng, F. e Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures, *The Canadian Journal of Statistics* **23**: 199–213.

Pinheiro, J. C., Liu, C. e Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution, *Journal of Computational and Graphical Statistics* **10**: 249–276.

Potthoff, R. F. e Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**: 313–326.

Roberts, C. (1966). A correlation model useful in the study of twins, *Journal of the American Statistical Association* **61**: 1184–1190.

Rosa, G. J. M., Gianola, D. e Padovani, C. R. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC, *Journal of Applied Statistics* **31**(7): 855–873.

Rosa, G. J. M., Padovani, C. R. e Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and bayesian MCMC implementation, *Biometrical Journal* **45**(5): 573–590.

Sahu, S. K., Dey, D. K. e Branco, M. D. (2003). A new class of multivariate skew distributions with aplications to bayesian regression models, *The Canadian Journal of Statistics* **31(2)**: 129–150.

Sartori, N. (2003). Bias prevention of maximum likelihood estimates: skew normal and skew t distributions, *Technical report*, Dip. Scienze Statistiche, Università di Padova, Italy.

Spiegelhalter, D. J., Best, N., Carlin, B. e Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Series B* **64**: 583–640.

Spiegelhalter, D. J., Thomas, A., Best, N. G. e Lunn, D. (2004). *WinBUGS User Manual Version 1.4.1.* http://www.mrc-bsu.cam.ac.uk/bugs.

Watanabe, T. (2001). On sampling the degree-of-freedom of student's-t disturbances, *Statistics and Probability Letters* **52**: 177–181.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms, *Journal of the American Statistical Association* **71**: 400–405.

Zhang, D. e Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data, *Biometrics* **57**: 795–802.

Zhou, T. e He, X. (2008). Three-step estimation in linear mixed models with skew-t distributions, *Journal of Statistical Planning and Inference* **138**: 1542–1555.