

Modelo de Regressão com Erros Normais Assimétricos: Uma Abordagem Bayesiana

Luiz Antonio de Freitas

Orientador: Prof. Dr. Josemar Rodrigues

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos

Dezembro de 2005

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

F866mr

Freitas, Luiz Antonio de.

Modelo de regressão com erros normais assimétricos:
uma abordagem bayesiana / Luiz Antonio de Freitas. -- São
Carlos : UFSCar, 2006.

111 p.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2005.

1. Análise de regressão. 2. Inferência bayesiana I. Título.

CDD: 519.536 (20ª)

à Lara.

*"Se voce tem uma maçã e eu tenho outra;
e nós trocamos as maçãs, então cada um terá sua maçã.
Mas se voce tem uma idéia e eu tenho outra, e nós as trocamos;
então cada um terá duas idéias"*

(George Bernard Shaw)

Agradecimentos

Agradeço,

Ao destino que me tirou da pobreza e me fez professor, me fez também menino.

À Luiz Gomes de Freitas, in memoriam, e Ulisses de Oliveira, in memoriam, com os quais eu me identificava.

Ao professor Dr. Josemar Rodrigues pela orientação e pelas idéias durante todo este trabalho.

Aos professores Dr. José Roberto Zorzatto e Dr. Leandro Sauer, que acreditaram em minha capacidade.

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

Aos colegas, professores e funcionários do Departamento de Estatística da UFSCar, pela amizade.

Aos meus amigos Velho, Pedro, Tereza e Lito, dos quais tirei lição de vida.

À Edina, Maria da glória e Juberti.

À Lara, por me fazer feliz.

Resumo

A análise estatística para o estudo de dados contínuos tem sido desenvolvida em grande parte com base no modelo normal, que se destaca sobretudo no contexto de modelos lineares e não lineares. No estudo da regressão linear univariada, mesmo aceitando como razoável a suposição de simetria para os erros, em muitas situações práticas essa suposição de normalidade pode nos levar a inferências pouco apropriadas sobre os parâmetros de interesse.

Na literatura clássica é suposto que os erros seguem uma distribuição normal com média zero e variância desconhecida. Neste trabalho, pretendemos flexibilizar essa suposição de normalidade, dispondo de novas classes de distribuições assimétricas. A primeira classe considerada é a proposta por Azzalini [3], que inclui a distribuição normal como um caso particular. A segunda é a dos modelos flexíveis de Ma & Genton [22] e a terceira é a classe de distribuições normais assimétricas estendidas, proposta por Arellano-Valle & Azzalini [2]. Estas duas últimas foram incluídas como propostas alternativas ao modelo de Azzalini [3]. Com esse objetivo, métodos de inferência baseados na abordagem bayesiana, serão desenvolvidos para estimar os parâmetros envolvidos no modelo de regressão linear simples com erros assimétricos.

Palavras-chave: Inferência Bayesiana, MCMC, Distribuição normal assimétrica, Modelo flexível, Parâmetro de assimetria, Distribuição normal assimétrica estendida, Seleção de modelos, Modelo com efeito aleatório, DIC, BIC.

Abstract

The statistical analysis of the continuous data set has been developed in most cases for normal models, specially, in the linear and non linear models context. In the simple linear regression models, even accepting as reasonable the normal error assumption, usually it can take misleading inferences for the parameter of interest.

In the classical literature is supposed that the errors follow the normal distribution with zero mean and unknown variance. In this work, we intend to make the normality assumption flexible using a new class of asymmetric distributions. The first class considered was studied by Azzalini [3], which includes the normal distributions as a particular case. The second flexible class was introduced by Ma & Genton [22] and the third one is the extended normal distributions proposed by Arellano-Valle [2]. These two last ones had been enclosed as alternative proposals to the model of Azzalini [3]. With this purpose, inferences procedures based on Bayesian approaches, will be developed to estimate the parameters involved in the simple linear regression models with asymmetrical errors.

Sumário

1	Introdução	1
2	Classe de distribuições normais assimétricas	3
2.1	Lema e representação estocástica	3
2.2	Definição e propriedades	5
2.3	Extensões para o caso univariado	11
2.3.1	Parâmetros de locação e escala	11
3	Construções de uma distribuição normal assimétrica	14
3.1	Construção por condicionamento	14
3.2	Representação de Henze[18]	16
3.3	Estatísticas de ordem	17
3.4	Versão de Cartinhour[13]	18
3.4.1	Exemplos	23
3.5	Versão de Arellano-Valle e Azzalini[2]	28
3.6	Interpretação bayesiana	29
3.7	Uma classe flexível de distribuições assimétricas	31
3.7.1	Algumas ordens	33
4	Inferência e discussão dos métodos	36
4.1	Inferência pelo método dos momentos	37
4.2	Inferência pelo método de máxima verossimilhança	41
4.3	Inferência bayesiana	43
4.3.1	Representação de Henze[18]	44
4.3.2	Exemplos	44

5	Modelo de regressão com erro normal assimétrico	48
5.1	Modelos	49
5.1.1	Modelo linear simples com erro normal	50
5.1.2	Modelo linear simples com erro normal assimétrico	51
5.2	Análise bayesiana	55
5.2.1	Modelo (5.1)	55
5.2.2	Modelo (5.3)	58
5.2.3	Método dos uns	61
5.2.4	Simulação	62
6	Modelos alternativos para o erro	68
6.1	Modelo flexível de Ma & Genton (2004)	68
6.1.1	Distribuição a posteriori de $\alpha, \beta, \sigma^2, \mu, \beta_1$ e β_3	69
6.2	Modelo com efeito aleatório da observação	70
6.2.1	Distribuição a posteriori de $\alpha, \beta, \sigma_\varepsilon^2, \sigma_\gamma^2, \delta$ e λ	73
6.3	Modelo com efeito aleatório de variável binária	74
6.3.1	Distribuição a posteriori de $\alpha, \beta, \sigma^2, \gamma_1, \gamma_2$ e λ	75
7	Aplicação	77
7.1	Estimação dos parâmetros	79
7.2	Comparação de modelos	84
8	Considerações finais	91
	Referências Bibliográficas	94
A	Conjuntos de dados	99
A.1	Renda Pessoal e investimento pessoal (Zelner[35])	99
A.2	Dados usados na aplicação do capítulo 7	100
B	Programas em R	102
B.1	Programa para geração do gráfico de uma distribuição $SN(\lambda)$ dada por (2.3)	102
B.2	Geração de amostra de regressão com erro normal assimétrico	102

C	Programas em WinBUGS	104
C.1	Programa 1 para o Modelo 1	104
C.2	Programa 2 para o Modelo 1	105
C.3	Programa para o Modelo 2	106
C.4	Programa para o Modelo 3	107
C.5	Programa Para o Modelo 4	108
C.6	Programa para o Modelo 5	109
D	CrITÉrios de seleÇo de modelos	110
E	Lista de distribuices	111

Capítulo 1

Introdução

Na prática a intuição leva a especular que assimetria é resposta à influência de fatores que definem a população. Entre estes fatores cita-se o fato de a população ter limites inferior e superior, muito comum na prática. Como exemplo, se a população tiver um limite inferior bem definido e o seu limite superior é não definido, a relação de proximidade relativa da média com esse limite inferior decide a intensidade da sua assimetria à direita. Por aproximação relativa entenda-se a distância dada em unidade padrão. Para ilustrar este exemplo cita-se a idade dos alunos de primeiro ano de um curso de graduação. A idade mínima destes alunos é bem definida, digamos 16 anos, enquanto que a idade máxima não é bem definida. Neste caso, um histograma acusa assimetria à direita e a intensidade dessa assimetria parece estar ligada ao fato de a média estar mais próxima ou não do limite inferior. Se coletarmos uma mostra de tamanho n de uma população com idades entre zero e 200 anos, a assimetria do histograma da amostra de idades que porventura venha ocorrer será mais por influência de outros fatores aqui não considerados e menos pela influência destes limites.

A distribuição normal assimétrica foi formalmente introduzida por Azzalini [3], que propos uma nova classe de distribuições que inclui a normal como um membro próprio. Nessa proposta, que é para o caso univariado, Azzalini chega a generalizar para o caso multivariado. Henze [18] obtém a densidade proposta por Azzalini via transformação de variáveis, onde a variável em questão é obtida como uma combinação linear de variáveis aleatórias independentes, e esse método permite a obtenção de seus momentos de ordem

ímpar, além de ter grande utilidade na simulação desta variável. É fácil de ser simulada em linguagem **R** [8]. Em 1996, Azzalini & Dalla Valle [6] trabalham com o caso multivariado, com ênfase para o caso bivariado e apresentam a construção da densidade via condicionamento e via transformação de variáveis. Em 1999, Azzalini & Capitanio [4] publicam artigo abordando o caso multivariado, usando a construção por condicionamento e, baseados nesta construção, determinam a distribuição marginal de um subvetor de um vetor com distribuição normal assimétrica multivariada. Ainda baseados na construção por condicionamento determinam também a distribuição de transformações lineares e das formas quadráticas de vetores com distribuições assimétricas, estudam condições de independência entre vetores, entre formas quadráticas e entre vetores e formas quadráticas. Estendem a abordagem para distribuições com parâmetros de locação e escala.

Aqui trabalharemos apenas com o caso univariado. Portanto, todas as referências à distribuição normal assimétrica serão para o caso univariado, a menos de menção em contrário. No capítulo 2 serão estudados as definições e teoremas introdutórios. As formas de construção da densidade são apresentadas no capítulo 3, bem como os modelos de distribuições normais assimétricas flexíveis de ordem k , de Ma e Genton [22] e a apresentação de alguns exemplos. No capítulo 4 é discutida a estimação dos parâmetros envolvidos e são propostos métodos de inferência baseados nas abordagens bayesiana e clássica. Um grande problema enfrentado por todos os pesquisadores é a estimação do parâmetro de assimetria. Há propostas bayesianas e não bayesianas, e todas ainda apresentam problemas, que são levantados naquele capítulo, mas a abordagem bayesiana tem mostrado melhores resultados. Ainda no capítulo 4 é apresentado os resultados do trabalho de Branco & Rodrigues [12].

No capítulo 5 são introduzidos dois modelos de regressão linear simples, que se diferem pela distribuição dos erros e a inferência bayesiana dos parâmetros de interesse. No capítulo 6 são apresentados mais três modelos de regressão linear simples como alternativa aos modelos do capítulo 5, e no capítulo 7 é feita a aplicação dos modelos apresentados nos dois capítulos anteriores. No capítulo 8 são feitas propostas futuras como continuidade das pesquisas.

Capítulo 2

Classe de distribuições normais assimétricas

2.1 Lema e representação estocástica

Lema 1. Sejam f_0 uma função densidade de probabilidade simétrica em torno de 0 e G uma função de distribuição acumulada, tal que G' é simétrica em torno de 0. Então

$$f(z) = 2f_0(z)G\{w(z)\}, \quad -\infty < z < \infty, \quad (2.1)$$

é uma função densidade para alguma função $w(\cdot)$ ímpar.

Demonstração. Sejam Y e X variáveis aleatórias independentes, com densidades f_0 e G' respectivamente, e considere-se uma função $W = w(Y)$ ímpar, ou seja $w(-y) = -w(y)$ para todo $y \in \mathbb{R}$. Então, dado y temos que

$$\begin{aligned} P\{W \leq w(y)\} &= P\{y_0 \in \mathbb{R}; w(y_0) \leq w(y)\} \\ &= P\{-y_0; w(y_0) \leq w(y)\} \\ &= P\{-y_0; -w(y_0) \geq -w(y)\} \\ &= P\{-y_0; w(-y_0) \geq -w(y)\} \\ &= P\{y_1 \in \mathbb{R}; w(y_1) \geq -w(y)\} \end{aligned}$$

$$= P \{W \geq -w(y)\}.$$

Portanto $W = w(Y)$ tem densidade simétrica em torno de 0. Considere-se ainda a função $Z = X - w(Y)$. Temos que, dado z ,

$$\begin{aligned} P[X - w(Y) \leq z] &= P \{(x_0, y_0) \in \mathbb{R}^2; x_0 - w(y_0) \leq z\} \\ &= P \{(-x_0, -y_0); x_0 - w(y_0) \leq z\} \\ &= P \{(-x_0, -y_0); -x_0 + w(y_0) \geq -z\} \\ &= P \{(-x_0, -y_0); -x_0 - w(-y_0) \geq -z\} \\ &= P \{(x_1, y_1) \in \mathbb{R}^2; x_1 - w(y_1) \geq -z\} \\ &= P[X - w(Y) \geq -z]. \end{aligned}$$

Portanto, Z também tem densidade simétrica em torno de 0. Para concluir o lema temos que

$$1/2 = P[X - w(Y) \leq 0] = E_Y \{P[X - w(Y) \leq 0|Y]\} = \int_{\mathbb{R}} G\{w(z)\} f_0(z) dz \blacksquare$$

Uma versão simplificada deste lema é vista em Azzalini [3], onde se exige que G seja absolutamente contínua. Nessa versão temos que $w(z) = \lambda z$ um a um. Em Ma & Genton [22] é desenvolvida uma classe mais flexível de funções assimétricas com multimodalidade, o que é visto em mais detalhes no capítulo 3.

Representação estocástica

Sejam Y e X variáveis aleatórias independentes como definidas na demonstração do lema 1. Defina-se a variável aleatória Z , dada por

$$Z = \begin{cases} Y, & \text{se } G(X) < G(w(Y)), \\ -Y, & \text{caso contrário.} \end{cases} \quad (2.2)$$

Então, Z tem densidade dada por (2.1) [7].

2.2 Definição e propriedades

Definição 1. Uma variável aleatória Z tem distribuição normal assimétrica padrão com parâmetro de assimetria λ se sua densidade é dada por

$$f(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \quad (2.3)$$

onde λ é uma constante em \mathbb{R} , e $\phi(\cdot)$ e $\Phi(\cdot)$ denotam a função densidade de probabilidade (fdp) e a função de distribuição acumulada (fda), respectivamente, da distribuição normal padrão $N(0,1)$. Para denotar que Z tem distribuição normal assimétrica padrão com parâmetro de assimetria λ escreve-se

$$Z \sim SN(\lambda).$$

Esta definição é uma versão bastante simplificada do lema 1, onde $w(z) = \lambda z$. A Figura 2.1 mostra o comportamento de $f(z)$ para três valores diferentes de λ e o programa em R [8] para o traçado do gráfico é dado no apêndice B, programa B.1. É fácil notar que a normal padrão está incluída como um caso particular desta definição, quando $\lambda = 0$. Ou seja, uma família de densidades normais assimétricas padrão inclui a distribuição normal padrão como um membro próprio e, nessa família, o que define a assimetria da densidade, à esquerda ou à direita, bem como a sua intensidade, é o valor de λ . O comportamento da distribuição normal assimétrica padrão, de um modo geral, é esboçado através das propriedades a serem vistas nesta seção, todas elas decorrentes das definições em (2.3), e em (2.8). Para que estas propriedades sejam desenvolvidas deveremos ter em mãos a construção da função de distribuição acumulada da distribuição normal assimétrica padrão, que é dada a seguir.

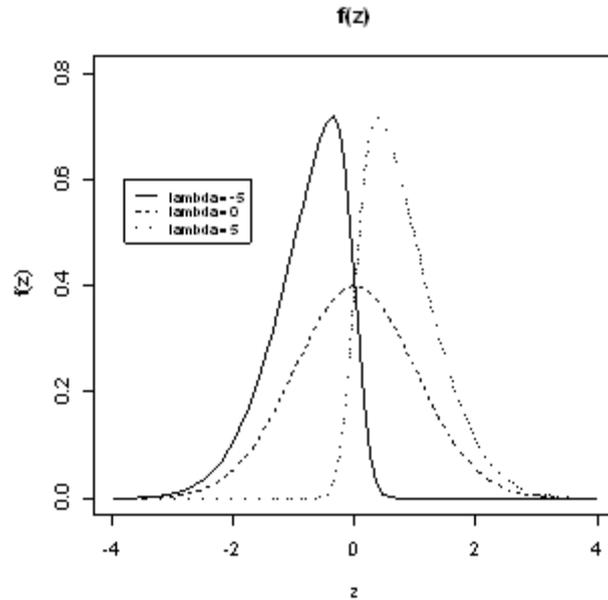


Figura 2.1 : Função densidade da distribuição normal assimétrica para três diferentes valores λ .

Função de distribuição

Denotaremos por $F(z; \lambda)$ a função de distribuição acumulada da densidade dada pela definição (2.3), isto é

$$F(z; \lambda) = 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t) \phi(u) du dt.$$

A região de integração é denotada por A na Figura 2.2 para valores positivos de z e λ . Recorremos à função $T(h, a)$, estudada por Owen [24] e usada por Azzalini [3] para a construção da função de distribuição acumulada da densidade normal assimétrica padrão. Para h e a positivos, $T(h, a)$ fornece a integral da densidade normal padrão bivariada na região $x \geq h$, $0 \leq y \leq ax$, no plano (x, y) , ou seja,

$$T(h, a) = \int_h^{\infty} \int_0^{ax} \phi(x) \phi(y) dy dx.$$

Então, com o auxílio da Figura 2.2 é possível verificar que, para valores positivos de z e

λ ,

$$\begin{aligned} F(z; \lambda) &= 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t) \phi(u) du dt \\ &= 2 \left[\int_{-\infty}^z \int_{-\infty}^0 \phi(t) \phi(u) du dt - \int_z^{\infty} \int_0^{\lambda t} \phi(t) \phi(u) du dt \right] \\ &= 2 \int_{-\infty}^z \int_{-\infty}^0 \phi(t) \phi(u) du dt - 2 \int_z^{\infty} \int_0^{\lambda t} \phi(t) \phi(u) du dt. \end{aligned}$$

Portanto,

$$F(z; \lambda) = \Phi(z) - 2T(z, \lambda), \quad (2.4)$$

onde $T(z, \lambda)$ é a integral de $\phi(t) \phi(u)$ sobre a região B da Figura 1.2.

Se $h_2 > h_1 > 0$, temos

$$T(h_2, a) = \int_{h_2}^{\infty} \int_0^{ax} \phi(x) \phi(y) dy dx < \int_{h_1}^{\infty} \int_0^{ax} \phi(x) \phi(y) dy dx = T(h_1, a)$$

Portanto $T(h, a)$ é decrescente em h . Temos ainda

$$-T(h, a) = - \int_h^{\infty} \int_0^{ax} \phi(x) \phi(y) dy dx = \int_h^{\infty} \int_0^{-ax} \phi(x) \phi(y) dy dx = T(h, -a), \quad (2.5)$$

$$T(-h, a) = \int_{-\infty}^{-h} \int_0^{ax} \phi(x) \phi(y) dy dx = \int_h^{\infty} \int_0^{ax} \phi(x) \phi(y) dy dx = T(h, a), \quad (2.6)$$

e

$$2T(h, 1) = \Phi(h) \Phi(-h). \quad (2.7)$$

Portanto, pelas propriedades de $T(h, a)$ e verificando a Figura 2.2, temos que (2.4) vale também para valores negativos de z e λ , isto é, (2.4) é a expressão geral da função de distribuição acumulada da densidade dada por (2.3).

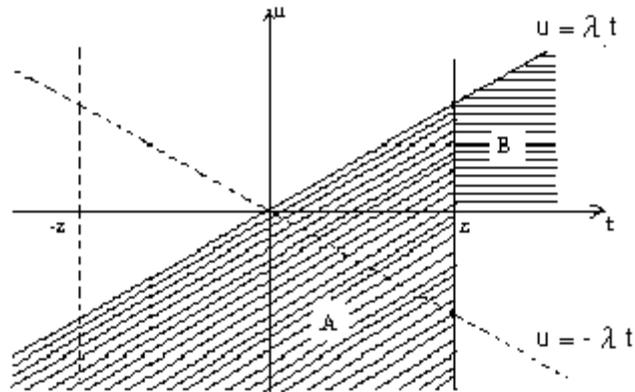


Figura 2.2 : Representação da função $T(h, a)$.

Propriedade 1. A densidade $SN(0)$ é a $N(0, 1)$.

Demonstração. Para $\lambda = 0$ temos que

$$f(z) = 2\phi(z)\Phi(0) = 2\phi(z)\frac{1}{2} = \phi(z) \blacksquare$$

Ou seja, a distribuição $N(0, 1)$ é um caso particular da distribuição $SN(\lambda)$ quando $\lambda = 0$.

Propriedade 2. Se $Z \sim SN(\lambda)$ então $Y = |Z| \sim HN(0, 1)$, onde $HN(0, 1)$ representa a distribuição denominada half-normal com densidade dada por $f(y) = 2\phi(y)I_{\{y>0\}}(y)$.

Demonstração. A variável aleatória Y é dada por

$$Y = \begin{cases} Z, & \text{se } Z \geq 0 \\ -Z, & \text{se } Z < 0. \end{cases}$$

Para $y < 0$ temos $P(Y \leq y) = 0$. Para $y > 0$, por (2.4) e (2.6) temos que

$$\begin{aligned} P[Y \leq y] &= P[-y \leq Z < 0] + [0 < Z \leq y] \\ &= \{F(0, \lambda) - F(-y, \lambda)\} + \{F(y, \lambda) - F(0, \lambda)\} \\ &= \{[\Phi(0) - 2T(0, \lambda)] - [\Phi(-y) - 2T(-y, \lambda)]\} \end{aligned}$$

$$\begin{aligned}
& + \{[\Phi(y) - 2T(y, \lambda)] - [\Phi(0) - 2T(0, \lambda)]\} \\
& = \Phi(0) - 2T(0, \lambda) - \Phi(-y) + 2T(-y, \lambda) \\
& \quad + \Phi(y) - 2T(y, \lambda) - \Phi(0) + 2T(0, \lambda) \\
& = \Phi(y) - \Phi(-y) + 2T(y, \lambda) - 2T(y, \lambda) \\
& = \Phi(y) - \Phi(-y).
\end{aligned}$$

Derivando $P[Y \leq y]$ em relação à y ($y > 0$) temos

$$\frac{\partial P[Y \leq y]}{\partial y} = \phi(y) - (-1)\phi(-y) = \phi(y) + \phi(-y) = \phi(y) + \phi(y) = 2\phi(y).$$

Para concluir basta definir que

$$f_Y(0) = \phi(0) \blacksquare$$

Propriedade 3. Se $Z \sim SN(\lambda)$ então, quando $\lambda \rightarrow \infty$, $Z \xrightarrow{D} HN(0, 1)$.

Demonstração. Se $Z \sim SN(\lambda)$, então sua densidade é dada por

$$f(z) = 2\phi(z)\Phi(\lambda z).$$

Quando $Z < 0$ temos que $\Phi(\lambda z) \rightarrow 0$ e, portanto $f(z) \rightarrow 0$.

Para $Z = 0$ temos $f(0) \rightarrow \phi(0) \forall \lambda$. Para $Z > 0$ temos que $\Phi(\lambda z) \rightarrow 1$, quando $\lambda \rightarrow \infty$, o que resulta em $f(z) \rightarrow 2\phi(z)$, quando $\lambda \rightarrow \infty$ ■

Propriedade 4. Se $Z \sim SN(\lambda)$, então $-Z \sim SN(-\lambda)$.

Demonstração. $f(-z) = 2\phi(-z)\Phi[\lambda(-z)] = 2\phi(z)\Phi(-\lambda z)$ ■

Propriedade 5. Se $Z \sim SN(\lambda)$ então a fdp de Z é unimodal e $\log f(z)$ é uma função côncava [3].

Propriedade 6. $1 - F(-z, \lambda) = F(z, -\lambda)$.

Demonstração. Por (2.4), (2.6) e (2.5) temos que

$$1 - F(-z; \lambda) = 1 - [\Phi(-z) - 2T(-z, \lambda)]$$

$$\begin{aligned}
&= 1 - [\Phi(-z) - 2T(z, \lambda)] \\
&= 1 - \Phi(-z) + 2T(z, \lambda) \\
&= 1 - \Phi(-z) - 2[-T(z, \lambda)] \\
&= \Phi(z) - 2T(z, -\lambda) \\
&= F(z; -\lambda) \blacksquare
\end{aligned}$$

Propriedade 7. $F(z, 1) = \{\Phi(z)\}^2$.

Demonstração. Por (2.4) e (2.7) temos

$$\begin{aligned}
F(z; 1) &= \Phi(z) - 2T(z, 1) \\
&= \Phi(z) - \Phi(z)\Phi(-z) \\
&= \Phi(z) [1 - \Phi(-z)] \\
&= \Phi(z)\Phi(z) \\
&= [\Phi(z)]^2 \blacksquare
\end{aligned}$$

Propriedade 8. $\sup_z |\Phi(z) - F(z, \lambda)| = \pi^{-1} \arctan |\lambda|$ [3].

Propriedade 9. Se $Z \sim SN(\lambda)$ então $Z^2 \sim \chi_1^2$.

Demonstração. Seja $Y = Z^2$. Ao transformar variáveis, isso implica $Z = \pm Y^{1/2}$ e $\frac{\partial z}{\partial y} = \frac{1}{2\sqrt{y}}$. Então, pela propriedade 4 temos

$$\begin{aligned}
f_z(y) &= \frac{1}{2\sqrt{y}} 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \Phi\left(\lambda y^{\frac{1}{2}}\right) + \frac{1}{2\sqrt{y}} 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \Phi\left(-\lambda y^{\frac{1}{2}}\right) \\
&= \frac{1}{\sqrt{\pi}} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y} \left[\Phi\left(\lambda y^{\frac{1}{2}}\right) + \Phi\left(-\lambda y^{\frac{1}{2}}\right)\right] \\
&= \frac{1}{\Gamma\left(\frac{1}{2}\right)} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y} \left[\Phi\left(\lambda y^{\frac{1}{2}}\right) + 1 - \Phi\left(\lambda y^{\frac{1}{2}}\right)\right]
\end{aligned}$$

$$= \frac{1}{\Gamma\left(\frac{1}{2}\right)} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y} \blacksquare$$

2.3 Extensões para o caso univariado

Apresentaremos nesta seção algumas versões estendidas de distribuições normais assimétricas para o caso univariado. Estas versões são obtidas através da inclusão de parâmetros de locação e escala, vistas também nas seções 3.4 e 3.5. O caso apresentado nas duas primeiras seções deste capítulo pode ser chamado de distribuição normal assimétrica padrão em referência à normal padrão. Distribuições com parâmetros de locação e escala podem ser vistas em Azzalini [3], Azzalini & Capitanio [4], Azzalini & Capitanio [5] e Azzalini & Dalla-Valle [6].

2.3.1 Parâmetros de locação e escala

Definição 2. Uma variável aleatória Y tem distribuição normal assimétrica com parâmetro de assimetria λ , parâmetro de posição μ e parâmetro de escala σ , com $\sigma > 0$, se sua densidade é dada por

$$f(y) = 2\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right), \quad y \in \mathbb{R}, \quad (2.8)$$

onde λ , $\phi(\cdot)$ e $\Phi(\cdot)$ denotam a função densidade de probabilidade (fdp) e a função de distribuição acumulada (fda), respectivamente, da distribuição normal padrão $N(0, 1)$. Como notação escrevemos $Y \sim SN(\mu, \sigma^2, \lambda)$.

Se $Z \sim SN(\lambda)$ e $Y = \mu + \sigma Z$, então $Y \sim SN(\mu, \sigma^2, \lambda)$. Ou seja, qualquer combinação linear de uma variável aleatória normal assimétrica também terá distribuição normal assimétrica

Demonstração. A densidade de Z é dada por

$$f(z) = 2\phi(z)\Phi(\lambda z).$$

Escrevendo Z em função de Y temos que

$$\frac{Y - \mu}{\sigma} = Z \implies \left| \frac{\partial Z}{\partial Y} \right| = \frac{1}{\sigma}.$$

Pontanto, a densidade de Y é dada por

$$f(y) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\frac{Y - \mu}{\sigma}\right) \blacksquare$$

Nas propriedades a seguir consideramos $Y \sim SN(\mu, \sigma^2, \lambda)$.

Propriedade 10. $Y_2 = a + bY \sim SN(a + b\mu, b^2\sigma^2, \lambda)$ para a e b reais.

Demonstração. A densidade de Y é dada por

$$f(y) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right).$$

Escrevendo Y em função de Y_2 temos que

$$\frac{Y_2 - a}{b} = Y \implies \left| \frac{\partial Y}{\partial Y_2} \right| = \frac{1}{b}.$$

Logo, a densidade de Y_2 é dada por

$$f(y_2) = \frac{2}{b\sigma} \phi\left(\frac{y_2 - (a + b\mu)}{b\sigma}\right) \Phi\left(\lambda \frac{y_2 - (a + b\mu)}{b\sigma}\right) \blacksquare$$

Propriedade 11. A função geradora demomentos de Y é dada por

$$M(t) = 2 \exp\left(\frac{(t - \mu)^2}{2\sigma^2}\right) \Phi\left(\delta \frac{t - \mu}{\sigma}\right),$$

onde $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$ [9].

Propriedade 12. A média e a variância de Y são

$$E(Y) = \mu + \sigma \delta \sqrt{\frac{2}{\pi}}$$

$$Var(Y) = \sigma^2 \left[1 - \frac{2}{\pi} \delta^2 \right].$$

Esta propriedade 12 é demonstrada na seção 4.1 do capítulo 4.

Propriedade 13. O terceiro momento e o coeficiente de assimetria são dados, respectivamente, por

$$E(Y^3) = \mu^3 + 3\mu^3\sigma\delta\sqrt{\frac{2}{\pi}} + 3\mu\sigma^2 + 3\sigma^3\delta\sqrt{\frac{2}{\pi}} - \sigma^3\delta^3\sqrt{\frac{2}{\pi}},$$

$$\gamma_1 = \sqrt{\frac{2}{\pi}} \delta^3 \left[\frac{4}{\pi} - 1 \right] \left[1 - \frac{2}{\pi} \delta^2 \right]^{-3/2},$$

este último assumindo valores no intervalo $(-0.99527, 0.99527)$ [9]. Mais detalhes podem ser encontrados em Pewsey [26] e Azzalini [3]. No capítulo 4 temos uma abordagem um pouco mais detalhada sobre a função geradora de momentos e sobre as estimativas dos parâmetros de interesse pelo método dos momentos.

Capítulo 3

Construções de uma distribuição normal assimétrica

Neste capítulo são abordadas algumas formas de se obter a classe de distribuições caracterizada pela densidade dada em (2.3) e (2.8). Nos três primeiros casos a construção será feita a partir de uma distribuição normal bivariada. No quarto caso, versão de Cart-inhour [13], parte-se de uma distribuição normal multivariada, e a partir desta forma de construção são apresentados e desenvolvidos alguns exemplos. No último caso é considerada uma versão de Arellano-Valle & Azzalini [2]. A representação estocástica, vista no capítulo 2, também é uma maneira de se construir a distribuição normal assimétrica. Ainda neste capítulo é feita uma interpretação bayesiana para a distribuição normal assimétrica. Por último é apresentado o modelo flexível de Ma & Genton [22].

3.1 Construção por condicionamento

Considere-se o vetor aleatório bivariado $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $X \sim N_2(0, \Omega)$, onde

$$\Omega = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}, \quad \text{com } |\delta| < 1.$$

Defina-se $Z \stackrel{D}{=} X_2 | [X_1 > 0]$. Então $Z \sim SN(\lambda)$, onde $\lambda = \frac{\delta}{\sqrt{1-\delta^2}}$.

Demonstração.

$$f_Z(z) = f_{X_2|X_1>0}(x_2) = \frac{f_{X_2}(x_2) P(X_1 > 0|X_2 = x_2)}{P(X_1 > 0)}.$$

Temos que

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{1-\delta^2} & \frac{-\delta}{1-\delta^2} \\ \frac{-\delta}{1-\delta^2} & \frac{1}{1-\delta^2} \end{bmatrix}.$$

Portanto, a densidade conjunta de (X_1, X_2) é dada por

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{1}{2\pi |\Omega|^{1/2}} \exp \left[-\frac{1}{2} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \Omega^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right] \\ &= \frac{1}{2\pi |\Omega|^{1/2}} \exp \left[-\frac{1}{2} \frac{1}{(1-\delta^2)} (x_1^2 - 2x_1x_2\delta + x_2^2) \right] \\ &= \frac{1}{2\pi |\Omega|^{1/2}} \exp \left[-\frac{1}{2} \frac{1}{(1-\delta^2)} (x_1^2 - 2x_1x_2\delta + x_2^2 - x_2^2\delta^2 + x_2^2\delta^2) \right] \\ &= \frac{1}{2\pi |\Omega|^{1/2}} \exp \left[-\frac{1}{2} \frac{1}{(1-\delta^2)} [(x_1^2 - 2x_1x_2\delta + x_2^2\delta^2) + (1-\delta^2)x_2^2] \right] \\ &= \frac{1}{2\pi |\Omega|^{1/2}} \exp \left[-\frac{1}{2} \frac{1}{(1-\delta^2)} (x_1^2 - 2x_1x_2\delta + x_2^2\delta^2) - \frac{1}{2} x_2^2 \right] \\ &= \frac{1}{\sqrt{2\pi} (1-\delta^2)} \exp \left[-\frac{1}{2} \frac{1}{(1-\delta^2)} (x_1 - x_2\delta)^2 \right] \times \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} x_2^2 \right) \\ &= f_{X_1|X_2}(x_1|x_2) \times f_{X_2}(x_2). \end{aligned}$$

Temos então que

$$X_2 \sim N(0, 1), \quad X_1 \sim N(0, 1) \quad \text{e}$$

$$X_1|[X_2 = x_2] \sim N(\delta x_2, 1 - \delta^2).$$

Logo,

$$\begin{aligned} P(X_1 > 0 | X_2 = x_2) &= P\left\{ \frac{X_1 - \delta x_2}{\sqrt{(1 - \delta^2)}} > \frac{-\delta x_2}{\sqrt{(1 - \delta^2)}} \right\} \\ &= P\left\{ Z < \frac{\delta x_2}{\sqrt{(1 - \delta^2)}} \right\} \\ &= \Phi\left\{ \frac{\delta x_2}{\sqrt{(1 - \delta^2)}} \right\}. \end{aligned}$$

Temos ainda que $P(X_1 > 0) = \frac{1}{2}$, o que completa a demonstração, ou seja

$$f_Z(z) = 2\phi(z)\Phi\left(\frac{\delta z}{\sqrt{1 - \delta^2}}\right) \blacksquare \quad (3.1)$$

3.2 Representação de Henze[18]

A representação estocástica de Henze [18] consiste em construir a distribuição normal assimétrica como uma combinação linear de variáveis aleatórias independentes, e este método facilita em muito os trabalhos de simulação, uma vez que, a partir desta representação, pode-se implementar algoritmos computacionais para a geração de amostras.

Lema. Sejam $X_0 \sim N(0, 1)$ e $X_1 \sim N(0, 1)$ independentes, $\delta \in (-1, 1)$ e $Z = \delta |X_0| + \sqrt{1 - \delta^2} X_1$. Então

$$Z \sim SN(\lambda), \text{ onde } \lambda = \frac{\delta}{\sqrt{1 - \delta^2}}. \quad (3.2)$$

Demonstração.

Seja $W = |X_0|$. Então $W \sim HN(0, 1)$, pela propriedade 2. Temos

$$\begin{aligned} |X_0| &= W \text{ e} \\ X_1 &= \frac{Z - \delta W}{\sqrt{1 - \delta^2}}. \end{aligned}$$

O jacobiano da transformação é dado por

$$J = \left| \begin{array}{cc} 1 & 0 \\ \frac{-\delta}{\sqrt{1-\delta^2}} & \frac{1}{\sqrt{1-\delta^2}} \end{array} \right| = \frac{1}{\sqrt{1-\delta^2}}.$$

Deste modo,

$$\begin{aligned} f_Z(z) &= \int_0^\infty \frac{1}{\sqrt{1-\delta^2}} f_{|X_0|}(w) f_{X_1}\left(\frac{z-\delta w}{\sqrt{1-\delta^2}}\right) dw \\ &= \int_0^\infty \frac{2}{\sqrt{1-\delta^2}} \phi(w) \phi\left(\frac{z-\delta w}{\sqrt{1-\delta^2}}\right) dw \\ &= \frac{2}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{1-\delta^2}\sqrt{2\pi}} e^{-\frac{(w-\delta z)^2}{2(1-\delta^2)}} e^{-\frac{(z^2-\delta^2 z^2)}{2(1-\delta^2)}} dw \\ &= \frac{2e^{-z^2/2}}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{1-\delta^2}\sqrt{2\pi}} e^{-\frac{(w-\delta z)^2}{2(1-\delta^2)}} dw \\ &= 2\phi(z) \int_{\frac{-\delta}{\sqrt{1-\delta^2}}}^\infty \phi(w) dw = 2\phi(z) \Phi\left(\frac{\delta z}{\sqrt{1-\delta^2}}\right) \blacksquare \end{aligned}$$

3.3 Estatísticas de ordem

Este método de construção pode ser apresentado através do seguinte exemplo encontrado em Roberts [27]. Ao estudar o efeito de determinada doença em uma amostra de irmãos gêmeos considerou-se apenas a idade do primeiro gêmeo a adquirir a doença. Se considerarmos X_1 e X_2 as idades dos irmãos ao adquirirem a doença, podemos supor que conjuntamente estas variáveis tenham distribuição normal bivariada. Como esperava-se que os valores de X_1 e X_2 estivessem bastante próximos registrou-se apenas a variável $\min(X_1, X_2)$. Esta variável tem distribuição assimétrica. Considere-se o vetor

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\mathbf{0}, \mathbf{\Omega})$$

onde $\mathbf{\Omega} = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}$.

Seja $W = \min(X_1, X_2)$. Então,

$$\begin{aligned} F_W(w) &= F_{X_1, X_2}(w, w) = \int_{-\infty}^w \int_{-\infty}^w f_{X_1, X_2}(w, w) dx_1 dx_2 \\ &= \int_{-\infty}^w \int_{-\infty}^w \frac{1}{2\pi (1 - \delta^2)^{1/2}} \exp \left\{ -\frac{x_1^2 + x_2^2 - 2\delta x_1 x_2}{2(1 - \delta^2)} \right\} dx_1 dx_2. \end{aligned}$$

Fazendo-se $x_1^2 - 2\delta x_1 x_2 = (x_1 - \delta x_2)^2 - \delta^2 x_2^2$, temos que

$$\begin{aligned} F_W(w) &= \int_{-\infty}^w \frac{1}{\sqrt{2\pi}} \exp(-x_2^2/2) \int_{-\infty}^w \frac{1}{\sqrt{2\pi(1 - \delta^2)}} \exp \left\{ -\frac{(x_1 - \delta x_2)^2}{2(1 - \delta^2)} \right\} dx_1 dx_2 \\ &= \int_{-\infty}^w \frac{1}{\sqrt{2\pi}} \exp(-x_2^2/2) \Phi \left(\frac{w - \delta x_2}{\sqrt{1 - \delta^2}} \right) dx_2. \end{aligned}$$

Derivando em relação a w em ambos os lados temos que

$$f_W(w) = 2\phi(w) \Phi \left(\frac{1 - \delta}{\sqrt{1 - \delta^2}} w \right),$$

ou

$$f_W(w) = 2\phi(w) \Phi(\lambda w),$$

com $\lambda = \frac{1 - \delta}{\sqrt{1 - \delta^2}}$.

3.4 Versão de Cartinhour[13]

Na sua versão de construção da distribuição normal assimétrica Cartinhour [13] parte de uma distribuição normal multivariada, na qual todas as componentes do vetor aleatório são truncadas em um intervalo fechado. O problema consiste em determinar a distribuição de uma das componentes sob influência do truncamento nas demais.

Seja $\underline{x} = (x_1, x_2, \dots, x_{n-1}, x_n)$ um vetor com densidade normal multivariada truncada,

isto é,

$$f(\underline{x}) = \begin{cases} \frac{1}{P\sqrt{(2\pi)^n|C|}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'A(\underline{x}-\underline{\mu})} , & \underline{x} \in R \\ 0, & \text{caso contrário,} \end{cases} \quad (3.3)$$

onde $A = C^{-1}$ e R é um retângulo em \mathbb{R}^n , isto é,

$$R = \{(x_1, \dots, x_n) \in \mathbb{R}^n : b_i \leq x_i \leq a_i, i = 1, \dots, n\} \quad (3.4)$$

e P é uma constante de normalização dada por

$$P = \int_R \frac{1}{\sqrt{(2\pi)^n|C|}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'A(\underline{x}-\underline{\mu})} d\underline{x} . \quad (3.5)$$

Observe-se que $A = C^{-1}$, onde C é a matriz de covariâncias e $\underline{\mu}$ o vetor de médias \underline{x} .

Representaremos a densidade marginal de x_n por $f_n(x_n)$. Para obtê-la, integraremos a densidade multivariada em relação às outras variáveis, isto é,

$$f_n(x_n) = \begin{cases} \int_{b_{n-1}}^{a_{n-1}} \int_{b_{n-2}}^{a_{n-2}} \dots \int_{b_1}^{a_1} f(\underline{x}) dx_1 dx_2 \dots dx_{n-1}, & b_n \leq x_n \leq a_n \\ 0, & \text{caso contrário.} \end{cases} \quad (3.6)$$

Neste caso, se optarmos por uma mudança de variáveis, que resulta em média 0 e uma matriz de covariância diagonal, podemos complicar os limites de integração. Não devemos nos esquecer de que estamos integrando sobre um retângulo, trazendo-nos desvantagem na busca de nossos objetivos. Portanto, outra estratégia é a partição do vetor $\underline{\mu}$ e da matriz C tal que permita que a densidade marginal $f_n(x_n)$ seja expressa como

$$f_n(x_n) = \begin{cases} g(x_n) \int_{b_{n-1}}^{a_{n-1}} \dots \int_{b_1}^{a_1} h(\underline{x}_1, x_n) d\underline{x}_1 , & b_n \leq x_n \leq a_n \\ 0, & \text{caso contrário,} \end{cases} \quad (3.7)$$

onde $\underline{x}_1 = (x_1, x_2, \dots, x_{n-1})'$ e $h(\underline{x}_1, x_n)$ é a densidade de uma distribuição normal multivariada de dimensão $(n-1)$ e parametrizada por x_n . Em outras palavras, o vetor de médias da densidade é função de x_n .

Da equação (3.3) temos a forma quadrática $\underline{y}'Ay$, onde $\underline{y} = (\underline{x} - \underline{\mu})'$, que será particionada para ser melhor manipulada. O primeiro passo é observar que

$$\begin{aligned} \underline{y}'Ay &= \begin{pmatrix} \underline{y}'_1 & y_n \end{pmatrix} \begin{bmatrix} A_1 & \underline{a} \\ \underline{a}' & a_{nn} \end{bmatrix} \begin{pmatrix} \underline{y}_1 \\ y_n \end{pmatrix} \\ &= \underline{y}'_1 A_1 \underline{y}_1 + 2\underline{a}' \underline{y}_1 y_n + a_{nn} y_n^2, \end{aligned} \quad (3.8)$$

onde $\underline{y}_1 = \underline{x}_1 - \underline{\mu}_1$ e $y_n = x_n - \mu_n$.

A expressão (3.8) pode ser desenvolvida da seguinte forma:

$$\begin{aligned} \underline{y}'Ay &= \underline{y}'_1 A_1 \underline{y}_1 + 2\underline{a}' \underline{y}_1 y_n + a_{nn} y_n^2 = \\ &= \underline{y}'_1 A_1 \underline{y}_1 + y_n \underline{y}'_1 \underline{a} + y_n \underline{a}' \underline{y}_1 + y_n^2 \underline{a}' A_1^{-1} \underline{a} - y_n^2 \underline{a}' A_1^{-1} \underline{a} + y_n^2 a_{nn} \\ &= \underline{y}'_1 A_1 \underline{y}_1 + y_n \underline{y}'_1 A_1 A_1^{-1} \underline{a} + y_n \underline{a}' A_1^{-1} A_1 \underline{y}_1 \\ &\quad + y_n^2 \underline{a}' A_1^{-1} A_1 A_1^{-1} \underline{a} - y_n^2 \underline{a}' A_1^{-1} \underline{a} + y_n^2 a_{nn} \\ &= \underline{y}'_1 A_1 (\underline{y}_1 + y_n A_1^{-1} \underline{a}) + \underline{a}' A_1^{-1} y_n A_1 (\underline{y}_1 + y_n A_1^{-1} \underline{a}) - y_n^2 [\underline{a}' A_1^{-1} \underline{a} - a_{nn}]. \end{aligned}$$

Temos então,

$$\underline{y}'Ay = \left(\underline{y}_1 + y_n A_1^{-1} \underline{a} \right)' A_1 \left(\underline{y}_1 + y_n A_1^{-1} \underline{a} \right) - y_n^2 \left(\underline{a}' A_1^{-1} \underline{a} - a_{nn} \right). \quad (3.9)$$

Esta expressão pode ainda ser simplificada particionando a matriz de covariâncias C e lembrando que $A = C^{-1}$. Logo,

$$\begin{bmatrix} A_1 & \underline{a} \\ \underline{a}' & a_{nn} \end{bmatrix} \begin{bmatrix} C_1 & \underline{c} \\ \underline{c}' & c_{nn} \end{bmatrix} = \begin{bmatrix} I & \underline{0} \\ \underline{0} & 1 \end{bmatrix}. \quad (3.10)$$

Multiplicando a linha 1 da matriz A pela coluna 2 da matriz C , obtemos que

$$\underline{a} = \frac{-A_1 \underline{c}}{c_{nn}}, \quad (3.11)$$

de onde temos $A_1^{-1} \underline{a} c_{nn} = -\underline{c}$, $\underline{a}' = \frac{-\underline{c}' A_1}{c_{nn}}$ e $c_{nn} \underline{a}' A_1^{-1} = -\underline{c}'$.

Similarmente, multiplicando a linha 2 pela coluna 2 temos que

$$\underline{a}' \underline{c} + a_{nn} c_{nn} = 1. \quad (3.12)$$

Substituindo (3.11) em (3.12) obtemos

$$\begin{aligned} \frac{-\underline{c}' A_1 \underline{c}}{c_{nn}} + a_{nn} c_{nn} &= 1 \\ \Rightarrow \frac{-c_{nn} \underline{a}' A_1^{-1} A_1 A_1^{-1} \underline{a} c_{nn}}{c_{nn}} + a_{nn} c_{nn} &= 1 \\ \Rightarrow -\underline{a}' A_1^{-1} \underline{a} c_{nn} + a_{nn} c_{nn} &= 1 \\ \Rightarrow a_{nn} - \underline{a}' A_1^{-1} \underline{a} &= \frac{1}{c_{nn}}. \end{aligned} \quad (3.13)$$

As equações (3.11) e (3.13) substituídas em (3.9) e levando-se em conta que $\underline{y} = (\underline{x} - \underline{\mu})$, resultam em

$$\begin{aligned} (\underline{x} - \underline{\mu})' A (\underline{x} - \underline{\mu}) &= \left(\underline{y}_1 + y_n A_1^{-1} \underline{a} \right)' A_1 \left(\underline{y}_1 + y_n A_1^{-1} \underline{a} \right) + \frac{y_n^2}{c_{nn}} \\ &= \left[\underline{x}_1 - \left(\underline{\mu}_1 + \frac{(x_n - \mu_n) \underline{c}}{c_{nn}} \right) \right]' A_1 \left[\underline{x}_1 - \left(\underline{\mu}_1 + \frac{(x_n - \mu_n) \underline{c}}{c_{nn}} \right) \right] \\ &\quad + \frac{(x_n - \mu_n)^2}{c_{nn}}, \end{aligned} \quad (3.14)$$

onde \underline{x} e $\underline{\mu}$ são particionados da forma

$$\underline{x} = \begin{pmatrix} \underline{x}_1 \\ x_n \end{pmatrix} \quad \text{e} \quad \underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \mu_n \end{bmatrix}. \quad (3.15)$$

Para completar a demonstração, temos que

$$|A| = |A_1| \left| a_{nn} - \underline{a}' A_1^{-1} \underline{a} \right|. \quad (3.16)$$

Usando (3.15) com (3.13) e o fato de que $1/|A_1^{-1}| = |A_1|$, temos

$$(2\pi)^n |C| = (2\pi)^{n-1} |A_1^{-1}| (2\pi) c_{nn}. \quad (3.17)$$

Portanto, a densidade marginal de x_n pode ser expressa como

$$f_n(x_n) = \begin{cases} \frac{1}{P} \frac{1}{\sqrt{2\pi c_{nn}}} e^{-\frac{1}{2} \frac{(x_n - \mu_n)^2}{c_{nn}}} \\ \times \int_{b_{n-1}}^{a_{n-1}} \dots \int_{b_1}^{a_1} \frac{1}{\sqrt{(2\pi)^{n-1} |A_1^{-1}|}} e^{-\frac{1}{2} (\underline{x}_1 - \underline{m}(x_n))' A_1 (\underline{x}_1 - \underline{m}(x_n))} dx_1 \dots dx_{n-1}, \\ b_n \leq x_n \leq a_n \\ 0, \quad \text{caso contrário,} \end{cases} \quad (3.18)$$

onde

$$\underline{m}(x_n) = \underline{\mu}_1 + \frac{(x_n - \mu_n) \underline{c}}{c_{nn}}. \quad (3.19)$$

A equação (3.18) tem a forma da equação (3.7), onde queríamos chegar. Portanto, $f_n(x_n)$ pode ser obtida para um dado x_n . É fácil notar que se não ocorrer truncamento, a equação (3.18) se reduz à normal univariada com média μ_n e variância c_{nn} . A equação (3.18) pode ser expressa como

$$f_n(x_n) = \begin{cases} \frac{S(x_n)}{\sqrt{2\pi c_{nn}}} e^{-\frac{1}{2} \frac{(x_n - \mu_n)^2}{c_{nn}}}, & b \leq x_n \leq a_n \\ 0, & \text{caso contrário,} \end{cases} \quad (3.20)$$

onde $S(x_n)$ pode ser escrita como

$$S(x_n) = \frac{1}{P} \int_{b_{n-1}}^{a_{n-1}} \dots \int_{b_1}^{a_1} \frac{1}{\sqrt{(2\pi)^{n-1} |A_1^{-1}|}} e^{-\frac{1}{2} [\underline{x}_1 - \underline{m}(x_n)]' A_1 [\underline{x}_1 - \underline{m}(x_n)]} dx_1 \dots dx_{n-1}, \quad b_n \leq x_n \leq a_n \quad (3.21)$$

e vista como uma função de assimetria. Isto é, (3.21) provoca uma distorção na simetria da densidade normal com média μ_n e variância c_{nn} . Em geral, quanto mais truncamento houver, ou, de outra forma, quanto mais intenso o efeito do truncamento, mais assimétrica será $f_n(x_n)$. O efeito do truncamento dependerá também da correlação entre a variável X_n e as demais variáveis.

A função $S(x_n)$ faz o mesmo papel da função $G\{w(z)\}$ em (2.1). Ou melhor, a densidade (2.1) é um caso particular da densidade (3.20). E mais, o método de construção por condicionamento é um caso particular desta versão de Cartinhour [13]. Esta versão é pouco usual na literatura, mas a sua importância está na sua generalização, permitindo trabalhar com qualquer tipo de condicionamento e permite trabalhar de forma direta com casos em que a distribuição normal assimétrica tenha parâmetro de locação diferente de 0 e parâmetro de escala diferente de 1. A seguir a apresentação de três exemplos nos quais os resultados são obtidos utilizando a proposta de Cartinhour [13].

3.4.1 Exemplos

Exemplo 1: Consideremos o vetor

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right),$$

onde a região R é dada por

$$R = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq \infty, -\infty \leq x_2 \leq \infty\}$$

$$\text{e } C = A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Da equação (3.5) temos que

$$P = \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^2}} e^{-\frac{1}{2}(x_1-x_2)^2} dx dy = \frac{1}{2}.$$

Podemos escrever (3.19) como

$$m(x_2) = 0 + \left(\frac{x_2 - 0}{1} \right) = x_2 .$$

Substituindo o valor de P e o resultado de $m(x_2)$ em (3.21) temos

$$S(x_2) = 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} dx_1 = 1 .$$

Portanto, a densidade da marginal de X_2 é dada por

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2} , \quad -\infty < x_2 < \infty ,$$

ou seja, $X_2 \sim N(0, 1)$. Neste caso nota-se que a condição de que $X_1 > 0$ não afeta a densidade marginal de X_2 pelo fato de as duas marginais serem independentes.

Exemplo 2: Seja o vetor $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} , \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix} \right)$, $\delta \neq 0$, com a mesma região R do Exemplo 1. Temos

$$C = \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix} \quad \text{e} \quad A = \begin{bmatrix} \frac{1}{1-\delta^2} & \frac{-\delta}{1-\delta^2} \\ \frac{-\delta}{1-\delta^2} & \frac{1}{1-\delta^2} \end{bmatrix} .$$

Neste caso, da equação (3.5) temos que $P = \frac{1}{2}$, e de (3.19) temos que

$$m(x_2) = x_2 \delta .$$

A função de assimetria é dada por

$$S(x_2) = 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi(1-\delta^2)}} e^{-\frac{1}{2(1-\delta)}(x_1-x_2\delta)^2} dx_1 , \quad -\infty < x_2 < \infty .$$

Logo, a distribuição marginal de X_2 é dada por

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2} \times 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi(1-\delta^2)}} e^{-\frac{1}{2(1-\delta^2)}(x_1-x_2\delta)^2} dx_1.$$

Neste exemplo podemos verificar que, devido à correlação, o truncamento de X_1 é responsável por um comprometimento na simetria de $f_{X_2}(x_2)$. Neste e no exemplo 1, onde X_1 é truncada na média, podemos escrever $f_2(x_2)$ de outra forma:

$$\begin{aligned} f_2(x_2) &= 2\phi_z(x_2)P(X \geq 0|X_2 = x_2) \\ &= 2\phi_z(x_2)P\left(\frac{X_1 - x_2\delta}{\sqrt{1-\delta^2}} \geq \frac{-x_2\delta}{\sqrt{1-\delta^2}}\right) \\ &= 2\phi_z(x_2)P\left(z \leq \frac{\delta x_2}{\sqrt{1-\delta^2}}\right). \end{aligned}$$

Logo, temos que

$$f_2(x_2) = 2\phi(x_2)\Phi(\lambda x_2),$$

onde $\lambda = \frac{\delta}{\sqrt{1-\delta^2}}$ é o parâmetro de assimetria da distribuição. Este caso se enquadra perfeitamente na definição em (2.3) e este resultado pode ser obtido também pelo método do condicionamento ou pela representação de Henze [18]. No exemplo 1 temos que $\lambda = 0$.

Exemplo 3: Consideramos agora uma situação mais geral. Suponhamos que

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right),$$

com $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ elemento da região R dos exemplos 1 e 2.

As expressões de C e A são dadas por

$$C = \begin{bmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \quad \text{e} \quad A = \begin{bmatrix} \frac{1}{\sigma_1^2(1-\delta^2)} & \frac{-\delta}{\sigma_1\sigma_2(1-\delta^2)} \\ \frac{-\delta}{\sigma_1\sigma_2(1-\delta^2)} & \frac{1}{\sigma_2^2(1-\delta^2)} \end{bmatrix}.$$

Da equação (3.5) temos

$$\begin{aligned} P &= \int_0^\infty \int_{-\infty}^\infty f(x_1, x_2) dx_1 dx_2 = P(X_1 > 0) = P(x_1 - \mu_1 > -\mu_1) \\ &= P\left(\frac{x_1 - \mu_1}{\sigma_1} < \frac{\mu_1}{\sigma_1}\right) = \Phi_z\left(\frac{\mu_1}{\sigma_1}\right). \end{aligned}$$

De (3.21) obtemos

$$S(x_2) = \frac{1}{\Phi\left(\frac{\mu_1}{\sigma_1}\right)} \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_1^2(1-\delta^2)}} e^{-\frac{1}{2\sigma_1^2(1-\delta^2)}\left[x_1 - \left(\mu_1 + \frac{(x_2 - \mu_2)}{\sigma_2}\delta\sigma_1\right)\right]^2} dx_1,$$

$$-\infty < x_2 < \infty.$$

Conseqüentemente, temos que

$$\begin{aligned} f_2(x_2) &= \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2} \frac{1}{\Phi\left(\frac{\mu_1}{\sigma_1}\right)} \\ &\quad \times \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_1^2(1-\delta^2)}} e^{-\frac{1}{2\sigma_1^2(1-\delta^2)}\left[x_1 - \left(\mu_1 + \frac{(x_2 - \mu_2)}{\sigma_2}\delta\sigma_1\right)\right]^2} dx_1 \\ &= \frac{1}{\sigma_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2} \frac{1}{\Phi\left(\frac{\mu_1}{\sigma_1}\right)} \\ &\quad \times \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}} e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1 - \left(\mu_1 + \frac{(x_2 - \mu_2)}{\sigma_2}\rho\sigma_1\right)\right]^2} dx_1 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma_2} \phi \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \frac{1}{\Phi \left(\frac{\mu_1}{\sigma_1} \right)} \\
&\quad \times P \left(\frac{X_1 - \left(\mu_1 + \frac{(x_2 - \mu_2)}{\sigma_2} \delta \sigma_1 \right)}{\sigma_2 \sqrt{1 - \delta^2}} > - \frac{\mu_1 + \frac{(x_2 - \mu_2)}{\sigma_2} \delta \sigma_1}{\sigma_2 \sqrt{1 - \delta^2}} \right) \\
&= \frac{1}{\sigma_2} \phi \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \frac{1}{\Phi \left(\frac{\mu_1}{\sigma_1} \right)} \times P \left(Z \leq \frac{\mu_1 \sigma_2 + x_2 \delta \sigma_1 - \mu_2 \delta \sigma_1}{\sigma_2 \sqrt{1 - \delta^2}} \right) \\
&= \frac{1}{\sigma_2} \phi \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \frac{1}{\Phi \left(\frac{\mu_1}{\sigma_1} \right)} \times \Phi \left(\frac{\mu_1 \sigma_2 + x_2 \delta \sigma_1 - \mu_2 \delta \sigma_1}{\sigma_2^2 \sqrt{1 - \delta^2}} \right) \\
&= \frac{1}{\sigma_2} \phi \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \frac{1}{\Phi \left(\frac{\mu_1}{\sigma_1} \right)} \times \Phi \left(\frac{\left[x_2 - \mu_2 + \frac{\mu_1 \sigma_2}{\delta \sigma_1} \right] \delta \sigma_1}{\sigma_2^2 \sqrt{1 - \delta^2}} \right) \\
&= \frac{1}{\sigma_2} \phi \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \frac{1}{\Phi_z \left(\frac{\mu_1}{\sigma_1} \right)} \times \Phi_z \left(\frac{\left[x_2 - \mu_2 + \frac{\mu_1 \sigma_2}{\delta \sigma_1} \right] \delta \sigma_1}{\sigma_2^2 \sqrt{1 - \delta^2}} \right).
\end{aligned}$$

Finalmente, a densidade de $X_2|[X_1 > 0]$ pode ser escrita como

$$f_{X_2|X_1>0}(x_2) = \frac{1}{\sigma_2} \phi \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \frac{1}{\Phi \left(\frac{\mu_1}{\sigma_1} \right)} \times \Phi \left(\frac{\delta}{\sqrt{1 - \delta^2}} \frac{\sigma_1}{\sigma_2} \left[\frac{x_2 - \mu_2}{\sigma_2} + \frac{\mu_1}{\delta \sigma_1} \right] \right).$$

Este é um caso mais geral de distribuição normal assimétrica e a densidade acima pode ser obtida pelo método de construção por condicionamento. Uma forma mais específica de se obter esta densidade também é vista na seção 3.5, pelo método de construção de Cartinhour [13]. A abordagem de Cartinhour é pouco usual, mas permite calcular a densidade de X_2 seja qual for a condição ou o truncamento imposto à X_1 .

3.5 Versão de Arellano-Valle e Azzalini[2]

A versão de Arellano-Valle e Azzalini refere-se ao caso multivariado. Em nosso trabalho enfocamos o caso univariado.

Considere-se o vetor aleatório bivariado $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $X \sim N_2(0, \Omega)$, onde

$$\Omega = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}, \quad \text{com } |\delta| < 1.$$

Defina-se $Z \stackrel{D}{=} X_2 | [X_1 + \tau > 0]$ ou $X_2 | [X_1 > -\tau]$. Então, Z tem densidade dada por

$$f_Z(z) = \frac{1}{\Phi(\tau)} \phi(z) \Phi \left(\frac{\tau}{\sqrt{1-\delta^2}} + \frac{\delta}{\sqrt{1-\delta^2}} x_2 \right).$$

Demonstração.

$$f_Z(z) = f_{X_2 | [X_1 > -\tau]}(x_2) = \frac{f_{X_2}(x_2) P(X_1 > -\tau | X_2 = x_2)}{P(X_1 > -\tau)}.$$

Da seção 3.1 temos que a densidade conjunta de (X_1, X_2) é dada por

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{1}{\sqrt{2\pi(1-\delta^2)}} \exp \left[-\frac{1}{2(1-\delta^2)} (x_1 - x_2\delta)^2 \right] \times \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} x_2^2 \right) \\ &= f_{X_1 | X_2}(x_1 | x_2) \times f_{X_2}(x_2). \end{aligned}$$

Temos então que

$$X_2 \sim N(0, 1), \quad X_1 \sim N(0, 1) \quad \text{e}$$

$$X_1 | X_2 = x_2 \sim N(\delta x_2, 1 - \delta^2).$$

Logo,

$$P(X_1 > -\tau | X_2 = x_2) = P \left[\frac{x_1 - \delta x_2}{\sqrt{1-\delta^2}} > \frac{-\tau - \delta x_2}{\sqrt{1-\delta^2}} \right]$$

$$\begin{aligned}
&= P \left[Z < \frac{\tau + \delta x_2}{\sqrt{1 - \delta^2}} \right] \\
&= \Phi \left[\frac{\tau + \delta x_2}{\sqrt{1 - \delta^2}} \right].
\end{aligned}$$

Temos ainda que $P(X_1 > -\tau) = P(X_1 < \tau) = \Phi(\tau)$. Portanto,

$$\begin{aligned}
f_Z(z) &= \frac{1}{\Phi(\tau)} f_{X_2}(z) \Phi \left(\frac{\tau}{\sqrt{1 - \delta^2}} + \frac{\delta}{\sqrt{1 - \delta^2}} x_2 \right). \\
&= \frac{1}{\Phi(\tau)} \phi(z) \Phi \left(\frac{\tau}{\sqrt{1 - \delta^2}} + \frac{\delta}{\sqrt{1 - \delta^2}} x_2 \right) \blacksquare
\end{aligned}$$

Veremos a seguir uma generalização deste caso onde Ω agora é matriz de covariâncias. Há uma certa semelhança entre a versão de Arellano Valle e Azzalini [2] e a versão de Cartinhour [13].

Considere-se o vetor

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

e seja $Y = X_2 | [X_1 + \tau > 0]$. A densidade de Y é dada por

$$f(y) = \frac{1}{\Phi\left(\frac{\tau}{\sigma_1}\right)} \frac{1}{\sigma_2} \phi\left(\frac{y}{\sigma_2}\right) \Phi\left(\frac{\tau}{\sigma_2\sqrt{1 - \delta^2}} + \frac{\delta}{\sqrt{1 - \delta^2}} \frac{x_2}{\sigma_2}\right).$$

3.6 Interpretação bayesiana

A distribuição normal assimétrica surge pela primeira vez em um contexto bayesiano no trabalho de O'Hagan & Leonard [23]. Ao deduzirem a distribuição a priori marginal para a média μ de uma população normal, estes autores obtêm como resultado a distribuição normal assimétrica. Apresentamos aqui uma versão do trabalho de O'Hagan & Leonard.

Considere-se

$$\mu | \theta \sim N(\theta, \sigma^2),$$

$$\theta \sim N(0, \tau^2),$$

sob a condição $\theta > 0$. A densidade conjunta de (μ, θ) é dada por

$$\begin{aligned} f(\mu, \theta) &= f(\mu | \theta) f(\theta) \\ &= \frac{1}{2\pi\sigma\tau} \exp \left\{ -\frac{1}{2\sigma^2} (\mu - \theta)^2 - \frac{1}{2\tau^2} \theta^2 \right\} \\ &= \frac{1}{2\pi\sigma\tau} \exp \left\{ -\frac{1}{2} \left[\frac{1}{\sigma^2} (\mu^2 - 2\mu\theta + \theta^2) + \frac{1}{\tau^2} \theta^2 \right] \right\} \\ &= \frac{1}{2\pi\sigma\tau} \exp \left\{ -\frac{1}{2} \left[\frac{1}{\sigma^2} \mu^2 - \frac{2\mu\theta}{\sigma^2} + \theta^2 \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right) \right] \right\} \\ &= \frac{1}{2\pi\sigma\tau} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \mu & \theta \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma^2} & -\frac{1}{\sigma^2} \\ -\frac{1}{\sigma^2} & \frac{1}{\sigma^2} + \frac{1}{\tau^2} \end{pmatrix} \begin{pmatrix} \mu \\ \theta \end{pmatrix} \right\}. \end{aligned}$$

Portanto,

$$\begin{pmatrix} \mu \\ \theta \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \right],$$

onde

$$\Omega^{-1} = \begin{pmatrix} \frac{1}{\sigma^2} & -\frac{1}{\sigma^2} \\ -\frac{1}{\sigma^2} & \frac{1}{\sigma^2} + \frac{1}{\tau^2} \end{pmatrix}.$$

Temos que

$$\Omega = \frac{1}{\Delta} \begin{pmatrix} \frac{1}{\sigma^2} + \frac{1}{\tau^2} & \frac{1}{\sigma^2} \\ \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \end{pmatrix},$$

onde

$$\begin{aligned} \Delta &= \frac{1}{\sigma^2} \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right) - \frac{1}{\sigma^4} \\ &= \frac{1}{\sigma^2} \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} - \frac{1}{\sigma^2} \right) = \frac{1}{\sigma^2 \tau^2}. \end{aligned}$$

Portanto,

$$\Omega = \begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 \end{pmatrix}$$

e

$$\begin{pmatrix} \mu \\ \theta \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 \end{pmatrix} \right].$$

Fazendo

$$\delta = \frac{\tau}{\sqrt{\sigma^2 + \tau^2}},$$

pelo método de condicionamento e pela definição em (2.8) temos que

$$\lambda = \frac{\tau}{\sigma}$$

e

$$\mu | [\theta > 0] \sim SN(0, \sigma^2 + \tau^2, \lambda).$$

Para um dado valor de σ^2 , quanto maior o valor de τ^2 , mais credibilidade está se dando ao fato de que $\theta > 0$ e maior será o valor de λ . Por outro lado, fixando-se o valor de τ^2 , quanto menor o valor de σ^2 maior será o efeito da influência de $\theta > 0$ na assimetria da distribuição a condicional de μ .

3.7 Uma classe flexível de distribuições assimétricas

Nesta seção é considerada uma classe mais flexível de densidades que captam assimetria e multimodalidade, conforme proposta de Ma & Genton[22].

Um resultado de Wang *et al.* [33] confirma que uma função densidade $g(x)$ admite, para algum ξ uma única representação *skew-simétrica*, onde a expressão *skew-simétrica* representa uma classe de densidades (que nesta seção será denotada por SS), dada por

$$g(x) = 2f(x - \xi) \pi(x - \xi), \quad (3.22)$$

onde $f : \mathbb{R} \rightarrow \mathbb{R}_+$ é uma função densidade simétrica e $\pi : \mathbb{R} \rightarrow [0, 1]$ é uma função de assimetria satisfazendo a relação $\pi(-x) = 1 - \pi(x)$. Reciprocamente, uma função $g(\cdot)$ definida por (3.22) é uma densidade. Por simetria tem-se que $f(x) = f(-x)$. O Lema 1 e a definição (2.3) são casos particulares dessa definição, quando $\xi = 0$, para $\pi(x) = \Phi(\lambda x)$, onde $f(\cdot)$ e $\Phi(\cdot)$ representam a densidade e a função de distribuição

acumulada da distribuição $N(0, 1)$, respectivamente. Também as extensões para o caso univariado se enquadram nesta definição.

Da família de densidades $f(\cdot)$ que se enquadram nesta definição será considerada neste trabalho somente a classe C_0 , formada pelas funções f tais que $f(x) \rightarrow 0$ quando $|x| \rightarrow \infty$, da qual pertence a distribuição $N(\mu, \sigma^2)$. Destas restrições a que nos interessa é a distribuição $N(\mu, \sigma^2)$. Quanto à classe de funções $\pi(\cdot)$ de assimetria, esta forma um conjunto de funções

$$\pi(x) = H(w(x)),$$

onde $H(\cdot)$ é a função de distribuição acumulada de uma variável aleatória contínua simétrica em torno de zero, que aqui é restrita à classe $\Phi(w(x))$, onde $w : \mathbb{R} \rightarrow \mathbb{R}$ é uma função contínua ímpar, ou seja, $w(-x) = -w(x)$.

Um polinômio W_k de ordem k , em \mathbb{R} , é definido como uma combinação linear

$$W_k(x) = \sum_{r=0}^k \beta_r x^r, \text{ onde } \beta_r \in \mathbb{R} \text{ com } \beta_k \neq 0.$$

Seja W_k um polinômio de ordem k do tipo

$$\left\{ \begin{array}{l} W_k(x) = \sum_{r=1}^k \beta_r x^r, \text{ onde } \beta_r \in \mathbb{R} \text{ com } \beta_k \neq 0, \\ r = 1, 3, 5, \dots, k, \text{ com } k = 2n - 1, n \in \mathbb{N}. \end{array} \right.$$

Então W_k é chamado de polinômio ímpar e satisfaz $W_k(-x) = -W_k(x)$.

Dados $\mu, \sigma \in \mathbb{R}$ com $\sigma > 0$, seja $P_k(x)$ uma função do tipo

$$\left\{ \begin{array}{l} P_k(x) = \sum_{r=1}^k \beta_r \left(\frac{x-\mu}{\sigma}\right)^r, \text{ onde } \beta_r \in \mathbb{R} \text{ com } \beta_k \neq 0, \\ r = 1, 3, 5, \dots, k, \text{ com } k = 2n - 1, n \in \mathbb{N}. \end{array} \right.$$

Então, $P_k(x)$ pode ser escrita como um polinômio na forma

$$\left\{ \begin{array}{l} P_k(z) = \sum_{r=1}^k \beta_r z^r, \text{ onde } \beta_r \in \mathbb{R} \text{ com } \beta_k \neq 0, \\ r = 1, 3, 5, \dots, k, \quad k = 2n - 1, \quad n \in \mathbb{N}. \end{array} \right.$$

em z , estendido para uma reparametrização de x através da transformação $x = \mu + \sigma z$, onde

$$z = z(x) = \frac{x - \mu}{\sigma},$$

ou seja, a densidade (3.22) se reduz a

$$g(x) = \frac{2}{\sigma} f(z(x)) \pi(P_k[z(x)]).$$

3.7.1 Algumas ordens

Apresentamos nesta subseção a família de distribuições flexíveis de Ma & Genton até ordem 9, dados μ e σ^2 . As funções $\phi(\cdot)$ e $\Phi(\cdot)$ representam a densidade e a função de distribuição acumulada da $N(0, 1)$, respectivamente. Por convenção estes modelos serão chamados de distribuições flexíveis generalizadas *skew-normal* de ordem k , e serão denotados por *FGSN*.

FGSN de ordem 1

$$g(x) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\beta_1 \frac{x - \mu}{\sigma}\right), \beta_1 \in \mathbb{R}.$$

Notação: $X \sim FGSN(\mu, \sigma^2, \beta_1)$ ou $X \sim SN(\mu, \sigma^2, \beta_1)$.

FGSN de Ordem 3

$$g(x) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left[\beta_1 \left(\frac{x - \mu}{\sigma}\right) + \beta_3 \left(\frac{x - \mu}{\sigma}\right)^3\right], \beta_1, \beta_3 \in \mathbb{R}.$$

Notação: $X \sim FGSN(\mu, \sigma^2, \beta_1, \beta_3)$.

FGSN de Ordem 5

$$g(x) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left[\beta_1 \left(\frac{x - \mu}{\sigma}\right) + \beta_3 \left(\frac{x - \mu}{\sigma}\right)^3 + \beta_5 \left(\frac{x - \mu}{\sigma}\right)^5\right],$$

$$\beta_1, \beta_3, \beta_5 \in \mathbb{R}.$$

Notação: $X \sim FGSN(\mu, \sigma^2, \beta_1, \beta_3, \beta_5)$.

FGSN de Ordem 7

$$g(x) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi \left[\beta_1 \left(\frac{x-\mu}{\sigma}\right) + \beta_3 \left(\frac{x-\mu}{\sigma}\right)^3 + \beta_5 \left(\frac{x-\mu}{\sigma}\right)^5 + \beta_7 \left(\frac{x-\mu}{\sigma}\right)^7 \right],$$

$$\beta_1, \beta_3, \beta_5, \beta_7 \in \mathbb{R}.$$

Notação: $X \sim FGSN(\mu, \sigma^2, \beta_1, \beta_3, \beta_5, \beta_7)$.

FGSN de Ordem 9

$$g(x) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi \left[\beta_1 \left(\frac{x-\mu}{\sigma}\right) + \beta_3 \left(\frac{x-\mu}{\sigma}\right)^3 + \beta_5 \left(\frac{x-\mu}{\sigma}\right)^5 + \beta_7 \left(\frac{x-\mu}{\sigma}\right)^7 + \beta_9 \left(\frac{x-\mu}{\sigma}\right)^9 \right],$$

$$\beta_1, \beta_3, \beta_5, \beta_7, \beta_9 \in \mathbb{R}.$$

Notação: $X \sim FGSN(\mu, \sigma^2, \beta_1, \beta_3, \beta_5, \beta_7, \beta_9)$.

Para um modelo FGSN de ordem k ímpar qualquer pode-se propor a notação

$$X \sim FGSN^k(\mu, \sigma^2, \underline{\beta}).$$

A distribuição normal assimétrica definida por Azzalini [3] tem densidade de primeira ordem dada por

$$g(x) = \frac{2}{\sigma} f(z(x)) \pi(P_1[z(x)]).$$

Nesta família a quantidade máxima de modas depende da ordem k . As densidades que mais nos interessam neste trabalho são do tipo

$$g(x) = \frac{2}{\sigma} f(z(x)) \pi(P_3[z(x)]), \quad (3.23)$$

de ordem 3, que tem bimodalidade. Esta densidade ainda é pouco explorada e pouco se sabe sobre ela. Ainda não se tem, por exemplo uma expressão fechada para a média e para a variância. Uma outra questão que se pode levantar é como saber o número máximo de modas da distribuição a partir de um conjunto de dados.

Capítulo 4

Inferência e discussão dos métodos

Neste capítulo são discutidos diferentes métodos de inferência para os parâmetros do modelo normal assimétrico. A inferência nos modelos assimétricos ainda é objeto de investigação e são poucos os resultados obtidos. As abordagens pelo método clássico são as que apresentam mais problemas de identificabilidade e de sub ou super estimação, tanto no método dos momentos como no de máxima verossimilhança. Outros problemas são os de admissibilidade no caso de estimadores de momentos e não convergência no caso de máxima verossimilhança, para particulares resultados amostrais [9]. Tais dificuldades estão relacionadas essencialmente ao fato de que o conjunto de valores admissíveis para o coeficiente de assimetria para o modelo normal assimétrico é limitado.

Branco & Dey [11] usaram a metodologia clássica com o implemento do algoritmo EM para estimar os parâmetros em um modelo de regressão linear simples, supondo que os erros são independentes e identicamente distribuídos com distribuição t -assimétrica. Levantamos aqui a necessidade de se usar métodos bayesianos naquele mesmo trabalho, para que as estimativas possam ser comparadas.

A abordagem bayesiana tem apresentado melhores resultados mesmo quando são especificadas distribuições a priori não informativas. Esta abordagem é de fácil implementação em sistemas como R [8] e WinBUGS [32], por exemplo. Na linguagem BUGS já estão prontas as gerações das cadeias de Markov.

4.1 Inferência pelo método dos momentos

Devido à propriedade 9, os momentos de ordem par centrados na origem de uma variável aleatória $Z \sim SN(\lambda)$ são os mesmos da $N(0, 1)$, enquanto que os de ordem ímpar podem ser obtidos com o uso do resultado devido a Zacks [34], que é enunciado no Lema 2.

Lema 2. Seja $U \sim N(0, 1)$. Então

$$E \{ \Phi(aU + b) \} = \Phi \left(\frac{b}{\sqrt{1 + a^2}} \right).$$

onde a e b são reais.

Este lema é usado no cálculo da função geradora de momentos (*fgm*) de $Z \sim SN(\lambda)$. A *fgm* de Z é dada por

$$\begin{aligned} M_Z(t) &= E(e^{tz}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \Phi(\lambda z) dz \\ &= 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-t)^2 + \frac{t^2}{2}\right) \Phi(\lambda z) dz \\ &= 2 \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-t)^2\right) \Phi(\lambda z) dz. \end{aligned}$$

Da transformação $u = z - t$ temos que $u + t = z$ e

$$\frac{\partial u}{\partial x} = 1 \implies \partial u = \partial x.$$

Segue que

$$M_Z(t) = 2 \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \Phi(\lambda[u+t]) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du$$

e, devido ao Lema 2,

$$M_Z(t) = 2 e^{\frac{t^2}{2}} E[\Phi(\lambda U + \lambda t)]$$

$$\begin{aligned}
&= 2 e^{\frac{t^2}{2}} \Phi\left(\frac{\lambda t}{\sqrt{1+\lambda^2}}\right) \\
&= 2 e^{\frac{t^2}{2}} \Phi(\delta t),
\end{aligned}$$

onde

$$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}.$$

Derivando em relação a t obtemos

$$\frac{\partial M_Z(t)}{\partial t} = 2te^{\frac{t^2}{2}} \Phi(\delta t) + \delta \sqrt{\frac{2}{\pi}} e^{\frac{t^2}{2}} e^{-\frac{(\delta t)^2}{2}}.$$

Temos então que

$$E(Z) = \left. \frac{\partial M_Z(t)}{\partial t} \right|_{t=0} = \delta \sqrt{\frac{2}{\pi}}.$$

Para calcular a variância precisamos obter o segundo momento. Inicialmente obtemos

$$\begin{aligned}
\frac{\partial^2 M_Z(t)}{\partial t^2} &= 2 \left\{ e^{\frac{t^2}{2}} \Phi(\delta t) + t \left[te^{\frac{t^2}{2}} \Phi(\delta t) + \delta^2 t \frac{1}{\sqrt{2\pi}} e^{-\frac{(\delta t)^2}{2}} \right] \right\} \\
&\quad + \delta \sqrt{\frac{2}{\pi}} \left\{ te^{\frac{t^2}{2}} e^{-\frac{(\delta t)^2}{2}} - \delta^2 t e^{\frac{t^2}{2}} e^{-\frac{(\delta t)^2}{2}} \right\},
\end{aligned}$$

e

$$\left. \frac{\partial^2 M_Z(t)}{\partial t^2} \right|_{t=0} = 1.$$

Portanto,

$$\begin{aligned}
\text{Var}(Z) &= E(Z^2) - [E(Z)]^2 \\
&= 1 - \delta^2 \frac{2}{\pi}.
\end{aligned}$$

Para calcular o terceiro momento de Z centrado na origem partimos de

$$\begin{aligned}
\frac{\partial^3 M_Z(t)}{\partial t^3} &= 2t e^{t^2/2} \Phi(\delta t) + 2 e^{t^2/2} \delta \sqrt{\frac{2}{\pi}} e^{-(\delta t)^2/2} \\
&\quad + 4t e^{t^2/2} \Phi(\delta t) + 2t^3 e^{t^2/2} \Phi(\delta t) + 2t^3 e^{t^2/2} \delta \frac{1}{\sqrt{2\pi}} e^{-(\delta t)^2/2}
\end{aligned}$$

$$\begin{aligned}
& + 4\delta^2 t \frac{1}{\sqrt{2\pi}} e^{-\frac{(\delta t)^2}{2} + \frac{1}{2}t^2} + 2\delta^2 t^2 (t - \delta^2 t) e^{-\frac{(\delta t)^2}{2} + \frac{1}{2}t^2} \\
& + \delta \sqrt{\frac{2}{\pi}} e^{-\frac{(\delta t)^2}{2} + \frac{1}{2}t^2} + \delta \sqrt{\frac{2}{\pi}} t (t - \delta^2 t) e^{-\frac{(\delta t)^2}{2} + \frac{1}{2}t^2} \\
& - \delta^3 \sqrt{\frac{2}{\pi}} e^{-\frac{(\delta t)^2}{2} + \frac{1}{2}t^2} - \delta^3 \sqrt{\frac{2}{\pi}} t (t - \delta^2 t) e^{-\frac{(\delta t)^2}{2} + \frac{1}{2}t^2}.
\end{aligned}$$

Logo,

$$\left. \frac{\partial^3 M_Z}{\partial t^3} \right|_{t=0} = 3\delta \sqrt{\frac{2}{\pi}} + \delta^3 \sqrt{\frac{2}{\pi}}.$$

Seja $X = \mu + \sigma Z$, onde $Z \sim SN(\lambda)$. Temos então que

$$X \sim SN(\mu, \sigma^2, \lambda),$$

de modo que

$$E(X) = \mu + \sigma \delta \sqrt{\frac{2}{\pi}} \quad \text{e} \quad Var(X) = \sigma^2 \left(1 - \delta^2 \frac{2}{\pi} \right).$$

É importante observar que quando $\lambda = 0$ temos que

$$E(X) = \mu \quad \text{e} \quad Var(X) = \sigma^2.$$

Portanto neste modelo de densidade μ é um parâmetro de locação e σ^2 é um parâmetro de escala. Note-se que a densidade só é centrada em μ quando $\lambda = 0$. A variância é sempre menor que a variância do respectivo modelo de densidade normal.

O terceiro momento de X centrado na origem é dado por

$$E(X^3) = \mu^3 + 3\mu^3 \sigma \delta \sqrt{\frac{2}{\pi}} + 3\mu \sigma^2 + 3\sigma^3 \delta \sqrt{\frac{2}{\pi}} - \sigma^3 \delta^3 \sqrt{\frac{2}{\pi}}.$$

O método dos momentos consiste em estimar os parâmetros populacionais através de sistemas de equações que envolvem os momentos amostrais. Considere-se $\underline{z} = (z_1, z_2, \dots, z_n)$ amostra aleatória de tamanho n de uma variável aleatória $Z \sim SN(\lambda)$, isto é, z_1, z_2, \dots, z_n é uma seqüência de observações de variáveis aleatórias independentes e

identicamente distribuídas. Os momentos amostrais centrados na origem são dados por

$$M_k = \frac{1}{n} \sum_{i=1}^n z_i^k, \quad k = 1, 2, 3, \dots,$$

e os momentos amostrais centrados na média são dados por

$$m_1 = 0,$$

$$m_k = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^k, \quad k = 2, 3, \dots,$$

onde

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i = M_1,$$

que é o momento amostral de ordem 1 centrado na origem, mais conhecido como média amostral. Por uma questão de comodidade usaremos a notação s^2 para m_2 , estimador de máxima verossimilhança para a variância populacional.

Branco & Arellano-Valle [9] apresentam os resultados que são reproduzidos a seguir.

Sejam $X \sim SN(\mu, \sigma^2, \lambda)$ e Y_1, Y_2, \dots, Y_n uma amostra da variável aleatória padronizada

$$Y = \frac{X - \bar{x}}{s}.$$

Então, $Y \sim SN(\mu_Y, \sigma_Y^2, \lambda)$, onde

$$\mu_Y = \frac{\mu - \bar{x}}{s}$$

$$\sigma_Y^2 = \frac{\sigma^2}{s}.$$

Temos que $\bar{y} = 0$, $s_Y = 1$ e $m_3 = \frac{1}{n} \sum_{i=1}^n y_i^3$. Igualando os momentos amostrais aos momentos populacionais obtemos o seguinte sistema de equações:

$$\begin{cases} \mu_Y + \sigma_Y \delta \sqrt{\frac{2}{\pi}} = 0 \\ \sigma_Y^2 \left[1 - \frac{2}{\pi} \delta^2\right] = 1 \\ \mu_Y^3 + 3\mu_Y^2 \sigma_Y \delta \sqrt{\frac{2}{\pi}} + 3\mu_Y \sigma_Y^2 + 3\sigma_Y^3 \delta \sqrt{\frac{2}{\pi}} - \sigma_Y^3 \delta^3 \sqrt{\frac{2}{\pi}} = m_3. \end{cases}$$

Deste sistema obtemos

$$\hat{\mu}_Y = - \left(\frac{2}{4 - \pi} m_3 \right)^{1/3} s^{-1},$$

$$\hat{\sigma}_Y^2 = 1 + \hat{\mu}_Y^2$$

e

$$\hat{\delta} = - \frac{\hat{\mu}_Y}{\hat{\sigma}_Y} \sqrt{\frac{2}{\pi}}.$$

Portanto, os estimadores de momentos dos parâmetros envolvidos são dados por

$$\hat{\mu} = \bar{x} + s \hat{\mu}_Y,$$

$$\hat{\sigma}^2 = s^2 (1 + \hat{\mu}_Y^2)$$

e

$$\hat{\lambda} = \frac{\hat{\delta}}{\sqrt{1 - \hat{\delta}^2}},$$

este último sob a condição $|\hat{\delta}| < 1$, caso contrário $\hat{\delta}$ é inadmissível e $\hat{\lambda}$ não está definido. Estes estimadores não se comportam bem com a distribuição normal, isto é quando $\delta = 0$, pois o estimador de momentos $\hat{\sigma}^2$ super estima o parâmetro σ^2 e $\hat{\mu}$ super ou sub estima μ , dependendo do sinal de $\hat{\mu}_Y$.

4.2 Inferência pelo método de máxima verossimilhança

Considere-se $\underline{z} = (z_1, z_2, \dots, z_n)$ uma amostra aleatória de tamanho n de uma variável aleatória $Z \sim SN(\lambda)$. A função de verossimilhança é dada por

$$L(\lambda; \underline{z}) = \prod_{i=1}^n 2\phi(z_i) \Phi(\lambda z_i) \propto \prod_{i=1}^n \Phi(\lambda z_i).$$

Se λ é positivo e suficientemente grande, é grande a probabilidade de $z_i > 0, \forall i$. Então, neste caso $L(\lambda; \underline{z})$ é uma função monótona crescente em λ e, portanto, o estimador que maximiza a função de verossimilhança (e.m.v.) será infinito, o que levaria a inferir que

$Z \sim HN(0, 1)$. O mesmo raciocínio pode-se ter no caso de λ suficientemente distante, à esquerda de 0. Liseo [20] apresenta mais detalhes para outras amostras que nos levam ao mesmo problema.

Considere-se agora somente amostras onde o e.m.v. de λ é finito. Para obter o e.m.v. para o parâmetro λ precisamos maximizar o logaritmo da função de verossimilhança, dado por

$$l(\lambda) = \sum_{i=1}^n [\log [2\phi(z_i)] + \log \Phi(\lambda z_i)]$$

A primeira derivada é

$$l'(\lambda) = \sum_{i=1}^n \frac{z_i \phi(\lambda z_i)}{\Phi(\lambda z_i)}.$$

As raízes desta derivada só podem ser obtidas numericamente. A informação de Fisher é dada por

$$I_F(\lambda) = nE \left[\frac{Z^2 \phi^2(\lambda Z)}{\Phi^2(\lambda Z)} \right],$$

de onde obtemos

$$I_F(\lambda) = n \int 2z^2 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz. \quad (4.1)$$

Seja $X = \mu + \sigma Z$, onde $Z \sim SN(\lambda)$. Temos então que $X \sim SN(\mu, \sigma^2, \lambda)$ e a matriz de informação de Fisher é dada por

$$I_F(\mu, \sigma^2, \lambda) = \begin{bmatrix} \frac{1+\lambda^2 a_0}{\sigma^2} & \frac{(E(Z) \frac{1+2\lambda^2}{1+\lambda^2} + \lambda^2 a_1)}{\sigma^2} & \frac{\sqrt{2/\pi} (1+\lambda^2)^{3/2} - \lambda a_1}{\sigma} \\ \frac{(E(Z) \frac{1+2\lambda^2}{1+\lambda^2} + \lambda^2 a_1)}{\sigma^2} & \frac{2+\lambda^2 a_2}{\sigma^2} & -\frac{\lambda a_2}{\sigma} \\ \frac{\sqrt{2/\pi} (1+\lambda^2)^{3/2} - \lambda a_1}{\sigma} & -\frac{\lambda a_2}{\sigma} & a_2 \end{bmatrix},$$

onde $a_k = E_Z \left[Z^k \frac{\phi(\lambda Z)}{\Phi(\lambda Z)} \right]$, $k = 0, 1, 2$.

Um dos aspectos problemáticos é que esta matriz é singular quando $\lambda = 0$, e isto pode impedir que avaliemos a existência de assimetria usando testes baseados nesta matriz, tais como o teste de Wald. É conveniente destacar que este tipo de situação ocorre comumente quando a média é superparametrizada, como é o caso deste modelo normal assimétrico

onde a média depende de μ , σ^2 e λ . Azzalini [3] e Azzalini & Capitanio [4] removem este inconveniente adotando uma reparametrização mais apropriada para o modelo. A maior dificuldade com o método de máxima verossimilhança é que estudos de simulação têm mostrado que o estimador de máxima verossimilhança do parâmetro de assimetria pode não ser limitado, ainda que o verdadeiro valor deste parâmetro seja finito [9]. Uma alternativa é usar o algoritmo EM que requer a obtenção de uma variável latente. Uma outra alternativa é considerar o enfoque bayesiano.

4.3 Inferência bayesiana

A abordagem bayesiana é a que tem proporcionado melhores resultados nas inferências sobre os parâmetros de interesse, principalmente com relação ao parâmetro de assimetria λ . São propostos aqui dois métodos pela abordagem bayesiana que são o uso de MCMC utilizando a representação de Henze [18] e o métodos de uns, este último sugerido no sistema WinBUGS [32] e visto em mais detalhes na seção 5.2. A sigla MCMC é uma abreviação para o método de Monte Carlo com cadeias de Markov. As duas metodologias têm sido bastante úteis e equivalentes e têm sido usadas em todos os casos de simulação para demais inferências dos parâmetros de interesse. É interessante observar que o método dos uns é usado somente em Winbugs, mas evidenciamos a possibilidade de ser testada a sua implementação em R.

Sob a abordagem bayesiana, além do modelo normal assimétrico especificado para o conjunto de dados condicionado ao vetor de parâmetros (μ, σ^2, λ) , deveremos ter também uma distribuição para este vetor de parâmetros. Seja $\underline{\theta} = (\mu, \sigma^2, \lambda)$ e seja $\pi(\theta)$ uma densidade a priori dada para o vetor θ . Neste caso podemos supor uma distribuição a priori $\pi(\theta)$ multivariada para o vetor θ ou considerar independência entre os parâmetros. Nas simulações desta dissertação é assumida a independência entre os parâmetros de interesse. Seja $x = (x_1, x_2, \dots, x_n)$ amostra aleatória da variável aleatória X . Se considerarmos θ multivariado, a distribuição a posteriori conjunta para os parâmetros da variável X é dada mediante aplicação do teorema de Bayes, ou seja,

$$f(\theta|\underline{x}) = \frac{\pi(\theta)L(\theta/\underline{x})}{\int_{\theta} \pi(\theta)L(\theta/\underline{x}) d\theta}.$$

Uma questão relevante aqui é a escolha da distribuição a priori para os parâmetros da distribuição de X . Em princípio, toda inferência bayesiana é baseada no conceito subjetivo de probabilidade [9]. Ver ainda Paulino *et. al.* [25]. Uma distribuição a priori subjetiva para θ deverá descrever a nossa incerteza a respeito deste vetor de parâmetros, baseando-se no nosso conhecimento prévio a respeito do problema [9]. Então, se tivermos evidência de que o parâmetro λ é positivo, podemos atribuir ao mesmo uma densidade marginal a priori que expresse essa positividade, como por exemplo uma exponencial. No caso de σ ou σ^2 que é sempre positivo podemos atribuir uma distribuição a priori *IG* (gama inversa). Ao parâmetro μ também podemos fazer suposições quanto a ser positivo, negativo ou igual a 0, ou ainda pertencente a um intervalo qualquer de variação.

4.3.1 Representação de Henze[18]

A representação de Henze, além de ser útil em simulações de amostras de uma variável aleatória $X \sim SN(\mu, \sigma^2, \lambda)$, tem uma outra utilidade que é a estimação do vetor $\theta = (\mu, \sigma^2, \lambda)$, e isto é devido à propriedade a seguir.

Propriedade 14. Seja $Z \sim SN(\lambda)$ e $T \sim HN(0, 1)$. Então $Z|T = t \sim N(0, 1)$ [18] .

A propriedade 15 a seguir é uma extensão da propriedade anterior.

Propriedade 15. Seja $X \sim SN(\mu, \sigma^2, \lambda)$ e $T \sim HN(0, 1)$. Então,

$$X|T = t \sim N [\mu + \delta\sigma t, \sigma^2 (1 - \delta^2)] \quad (4.2)$$

(Henze [18]).

Estas duas últimas propriedades permitem a estimação dos parâmetros de interesse através de algoritmos computacionais.

4.3.2 Exemplos

Apresentamos nesta subseção dois exemplos nos quais a estimação dos parâmetros é feita mediante métodos bayesianos.

Exemplo 5: Uma opção é considerar os parâmetros independentes e dar a eles as seguintes distribuições a priori:

$$\mu \sim N(a, b^2),$$

$$\sigma^2 \sim IG(u, v) \text{ e}$$

$$\lambda \sim N(\mu_\lambda, \sigma_\lambda^2),$$

tomando $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

A densidade a posteriori para o vetor θ é dada por

$$\begin{aligned} \pi(\mu, \sigma^2, \lambda | \underline{x}, t) &\propto (\sigma^2)^{-(u+\frac{1}{2}+1)} \\ &\times \exp \left[-\frac{1}{2\sigma^2(1-\delta^2)} \sum_{i=1}^n (x_i - \mu - \delta\sigma t)^2 - \frac{1}{b^2}(\mu - a)^2 - \frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2 - \frac{v}{\sigma^2} \right]. \end{aligned}$$

As densidades a posteriori marginais para os três parâmetros são dadas por

$$\pi(\mu | \sigma^2, \lambda, \underline{x}, t) \propto \exp \left[-\frac{1}{2\sigma^2(1-\delta^2)} \sum_{i=1}^n (x_i - \mu - \delta\sigma t)^2 - \frac{1}{b^2}(\mu - a)^2 \right],$$

$$\pi(\sigma^2 | \mu, \lambda, \underline{x}, t) \propto (\sigma^2)^{-(u+\frac{1}{2}+1)} \exp \left[-\frac{1}{2\sigma^2(1-\delta^2)} \sum_{i=1}^n (x_i - \mu - \delta\sigma t)^2 - \frac{v}{\sigma^2} \right]$$

e

$$\pi(\lambda | \mu, \sigma^2, \underline{x}, t) \propto \exp \left[-\frac{1}{2\sigma^2(1-\delta^2)} \sum_{i=1}^n (x_i - \mu - \delta\sigma t)^2 - \frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2 \right].$$

Exemplo 6: Neste caso são consideradas distribuições a priori não informativa para o vetor (μ, σ^2) e t de Student para λ considerando $\lambda \Pi(\mu, \sigma^2)$. Este exemplo é um resumo do trabalho de Branco & Rodriguez [12]. Temos dois casos nos quais a distribuição a priori para o vetor (μ, σ) não é informativa ($\pi(\mu, \sigma) \propto \frac{1}{\sigma}$), e para λ é dada uma distribuição a priori $\pi(\lambda) = t(a, b; d)$, onde $t(a, b; d)$ denota a distribuição t de Student centrada em a , com parâmetro de escala b e com d graus de liberdade. Nos dois casos a seguir a distribuição t de Student para a priori de λ está centrada em 0.

Caso 1

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma},$$

$$\delta \sim U(-1, 1).$$

Temos então que $\lambda \sim t(0, 1/2; 2)$.

Caso 2

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma},$$

$$\lambda \sim t(0, \pi^2/4; 1/2).$$

Esta distribuição a priori para λ é uma aproximação da distribuição de Jeffreys proposta por Liseo & Loperfido [21].

Pela representação estocástica de Henze, propriedade 16, a distribuição de Y é dada por

$$Y_i | \mu, \sigma^2, \lambda, \tau_i \stackrel{ind.}{\sim} N\left(\mu + \frac{\sigma \lambda \tau_i}{\sqrt{1 + \lambda^2}}, \sigma^2 (1 - \delta^2)\right),$$

$$\tau_i \stackrel{iid}{\sim} HN(0, 1),$$

$i = 1, \dots, n$. Para implementar a inferência bayesiana utilizando o sistema WinBUGS Branco & Rodriguez [12] fazem algumas observações e adaptações. Não é possível considerar uma distribuição a priori imprópria do tipo $\pi(\mu, \sigma) \propto \frac{1}{\sigma}$. A distribuição a priori imprópria para o vetor (μ, σ) é aproximada por

$$\mu \sim N(0, 0.001) \quad \text{e} \quad \frac{1}{\sigma} \sim \text{Gama}(0.01, 0.01).$$

Além disso, o WinBUGS não aceita uma distribuição t de Student com número de graus de liberdade menor ou igual a 2. É considerada então a formulação hierárquica equivalente

$$\lambda | w \sim N\left(0, \frac{b}{w}\right) \quad \text{e} \quad w \sim \text{Gama}\left(\frac{d}{2}, \frac{d}{2}\right).$$

Visando a implementação das simulações MCMC com o amostrador de Gibbs efetuamos a reparametrização

$$\eta^2 = \sigma^2 (1 - \delta^2) \quad \text{e} \quad \beta = \sigma \delta$$

de forma que

$$Y_i | \tau_i, \mu, \eta, \beta \sim N(\mu + \beta \tau_i, \eta^2)$$

$$\tau_i \sim HN(0, 1),$$

$i = 1, \dots, n$. Por transformação de variáveis [12] obtemos

$$f(\mu, \beta, \eta, w) \propto \frac{1}{\eta^2} \exp\left(-\frac{1}{2} \frac{\beta^2 w}{\eta^2 b}\right) w^{\frac{d+1}{2}-1} \exp\left(-\frac{d}{2} w\right).$$

As distribuições condicionais utilizadas pelo amostrador de Gibbs são dadas por

$$w | \mu, \beta, \eta, \tau, Y \sim \text{Gama}\left(\frac{d+1}{2}, \frac{1}{2} \left(\frac{\beta^2}{b\eta^2} + d\right)\right),$$

$$\tau_i | \mu, \beta, \eta, w, Y \sim N\left(\frac{(y_i - \mu)\beta}{\eta^2 + \beta^2}, \frac{\eta^2}{\eta^2 + \beta^2}\right), \quad i = 1, \dots, n,$$

$$\mu | \beta, \eta, w, \tau, Y \sim N\left(\frac{\sum_{i=1}^n (y_i - \beta \tau_i)}{n}, \frac{\eta^2}{n}\right),$$

$$\beta | \mu, \eta, w, \tau, Y \sim N\left(\frac{\sum_{i=1}^n (y_i - \mu) \tau_i}{\frac{w}{b} + \sum_{i=1}^n \tau_i^2}, \frac{\eta^2}{\frac{w}{b} + \sum_{i=1}^n \tau_i^2}\right)$$

e

$$\frac{1}{\eta^2} | \mu, \beta, w, \tau, Y \sim \text{Gama}\left(\frac{n+1}{2}, \frac{1}{2} \left(\frac{w\beta^2}{b} + \sum_{i=1}^n (y_i - \mu - \beta \tau_i)^2\right)\right).$$

Capítulo 5

Modelo de regressão com erro normal assimétrico

A estatística para estudo de dados contínuos tem sido desenvolvida em grande parte com base no modelo normal, o qual se destaca sobretudo no contexto de modelos lineares e não lineares. No estudo da regressão linear univariada, mesmo aceitando como razoável a suposição de simetria para os erros, em muitas situações práticas essa suposição de normalidade pode nos levar a inferências pouco apropriadas sobre os parâmetros de interesse. Um exemplo, é quando a distribuição dos erros é simétrica, porém com caudas mais pesadas que a normal. Neste caso, diferentes pesquisadores sugerem utilizar modelos baseados em distribuições simétricas que permitam reduzir a influência de valores extremos, como por exemplo, a distribuição t de Student.

Jeffreys [19] assume erros independentes e identicamente distribuídos, com distribuição t de Student univariada. Em 1976, Zellner [36] publica um trabalho no qual faz análise bayesiana e não bayesiana de um modelo de regressão cuja densidade da variável resposta é multivariada t de Student. Nesse trabalho, Zellner usa um vetor de locação zero, para o vetor de erros, onde os erros são não correlacionados, porém não são independentes, e possuem distribuição marginal t de Student univariada, que inclui as distribuições de Cauchy e normal como casos especiais. Outros trabalhos de análise de regressão bayesiana, onde os erros têm uma distribuição com caudas pesadas podem ser vistos em Geweke [16].

Um outro exemplo surge quando a distribuição dos erros apresenta uma assimetria

que depende de um parâmetro λ . Uma opção é usar o modelo (5.2). Outra opção é o modelo (5.3). Neste trabalho é ajustado um conjunto de dados reais ao modelo (5.3) e desenvolvidos métodos de inferência, baseados na abordagem Bayesiana para estimar os parâmetros envolvidos. Uma terceira opção é usar uma classe de modelos que incluem assimetria e caudas pesadas. Branco & Dey [11] usam a distribuição *t* de Student assimétrica e consideram dois modelos. No primeiro caso é considerada uma distribuição multivariada do vetor de erros, onde as componentes desse vetor são dependentes mas não correlacionadas. No segundo caso os erros são independentes.

Por último podemos listar outras opções que podem ser desenvolvidas dependendo do conjunto de dados. Uma delas é a introdução de modelos lineares com efeitos aleatórios, no qual o efeito aleatório pode inclusive ter uma distribuição mais flexível do que a normal. Uma outra alternativa pode ser encontrada em Ma & Genton [22], que apresentam uma maior flexibilidade para a densidade dos erros do modelo, podendo ser multimodal.

5.1 Modelos

Na estatística o modelo linear geral usado para expressar a relação entre as variáveis Y e x é dado por

$$Y = \mu(x) + \varepsilon ,$$

onde Y e ε são variáveis aleatórias e $\mu(x)$ é uma função de x , definida em um domínio D . Neste trabalho a variável x é considerada não aleatória e, portanto, $\mu(x)$ será também não aleatória e é definida como a parte determinística do modelo. A variável Y é chamada de variável dependente ou variável resposta e a variável x de variável independente ou preditiva. Por exemplo, se Y representa a pressão sangüínea, desconhecida, de um indivíduo e x a sua idade, conhecida, então $\mu(x)$ representa o valor preditivo para a sua pressão sangüínea, a menos de um erro ε . De outra forma, se for medida a pressão sangüínea do indivíduo, obtém-se o valor Y e então a diferença $Y - \mu(x)$ representa o desvio do valor observado com relação ao valor esperado $\mu(x)$. Este desvio é denotado por ε e não é observável, mas a sua distribuição é conhecida, geralmente com esperança zero, e é considerada como parte do modelo.

Quanto à função $\mu(x)$ esta geralmente tem expressão matemática conhecida envolvendo parâmetros desconhecidos, e é de interesse estimá-los. A expressão modelo linear significa que $\mu(x)$ é uma função linear do tipo

$$\mu(x) = \beta_0 + \beta_1 q_1(x) + \beta_2 q_2(x) + \dots + \beta_k q_k(x) ,$$

onde $q_i(x)$ é uma função conhecida de x e não contém parâmetros desconhecidos.

Neste trabalho será considerado apenas o modelo linear dado na sua forma mais simples, ou seja

$$Y = \alpha + \beta x + \varepsilon .$$

Os dois modelos analisados neste capítulo se distinguem apenas pela distribuição do erro. A seguir, a descrição e o desenvolvimento destes dois modelos.

5.1.1 Modelo linear simples com erro normal

Considerando-se uma amostra aleatória de tamanho n , este modelo é dado por

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \quad i = 1, 2, \dots, n, \quad (5.1)$$

onde

$$\varepsilon_i = Y_i - \alpha - \beta x_i ,$$

e sua densidade é dada por

$$f(\varepsilon_i) = \left(\frac{1}{2\pi}\right)^{1/2} \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2} \varepsilon_i^2\right\} .$$

Temos que a densidade conjunta do vetor $\underline{\varepsilon}$ é dada por

$$f(\underline{\varepsilon}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \varepsilon_i^2\right\} .$$

Temos também que

$$Y_i \stackrel{ind.}{\sim} N(\alpha + \beta x_i, \sigma^2),$$

e sua densidade é dada por

$$f(y_i) = \left(\frac{1}{2\pi}\right)^{1/2} \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right\}, \quad i = 1, \dots, n.$$

Logo, a densidade conjunta do vetor \underline{Y} é dada por

$$f(\underline{y}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right\}.$$

Os parâmetros de interesse são α , β e σ^2 . A função de verossimilhança tem expressão

$$L(\alpha, \beta, \sigma^2; \mathcal{D}) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right\}.$$

onde \mathcal{D} denota o conjunto de dados.

Este modelo, como já comentado, é o caso mais simples de regressão linear e os parâmetros de interesse podem ser facilmente estimados tanto pelo método clássico (pelo método de mínimos quadrados), quanto pelo método bayesiano através de MCMC. O problema é que na prática nem sempre as suposições de independência e normalidade para o erro são verificadas, o que podem comprometer as estimativas e demais inferências.

5.1.2 Modelo linear simples com erro normal assimétrico

Como já foi observado no início desta seção o erro não é observável, mas tem distribuição especificada, sendo razoável em diversas aplicações supor que a esperança do mesmo seja zero. São apresentadas aqui duas versões deste modelo, e após uma ligeira análise justifica-se a escolha da versão (5.3) como o modelo a ser adotado.

Modelo não centrado

Considerando-se uma amostra aleatória de tamanho n , este modelo é dado por

$$Y_i = a + \beta x_i + \sigma \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} SN(0, 1, \lambda), \quad i = 1, \dots, n. \quad (5.2)$$

Temos então que

$$Y_i \stackrel{ind.}{\sim} SN(a + \beta x_i, \sigma^2, \lambda),$$

e, por (2.8), a densidade de $y_i|a, \beta, \sigma^2, \lambda$ é

$$f(y_i | a, \beta, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y_i - a - \beta x_i}{\sigma}\right) \Phi\left(\lambda \frac{y_i - a - \beta x_i}{\sigma}\right),$$

$i = 1, \dots, n$.

Neste modelo temos, pela propriedade 12 que

$$E(\varepsilon_i) = \frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}} \quad \text{e} \quad E(Y_i) = a + \beta x_i + \sigma \frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}}.$$

Temos então que $E(\varepsilon_i) = 0$ se, e somente se, $\lambda = 0$, caso em que $\varepsilon_i \sim N(0, \sigma^2)$. Para $a, \beta, \lambda > 0$, o gráfico da reta $y^* = \alpha + \beta x$ é mostrado na Figura 5.1.

Quando se pensa na previsão para Y por este modelo, a tendência natural, e precipitada, é considerar $\hat{y}_i^* = \hat{a} + \hat{\beta}x_i$ como preditor de $Y|x_i$. Neste caso, para \hat{a} e $\hat{\beta}$ não viciados, $E(\hat{y}_i^*) = a + \beta x_i \neq E(Y_i)$. Ou seja, a menos de $\lambda = 0$, este preditor de $Y|x$ subestima ou sobreestima Y , conforme o sinal de λ , com um vício dado por $-\sigma \frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}}$.

Um preditor para $y|x$ é $\hat{y} = \hat{a} + \hat{\beta}x + \hat{\sigma} \frac{\hat{\lambda}}{\sqrt{1 + \hat{\lambda}^2}} \sqrt{\frac{2}{\pi}}$.

O resíduo assimétrico para este modelo é dado por

$$\begin{aligned} E_i &= Y_i - E(Y_i) = a + \beta x_i + \sigma \varepsilon_i - a - \beta x_i - \sigma \frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}} \\ &= \sigma \left(\varepsilon_i - \frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}} \right) = \sigma \varepsilon_i^*, \end{aligned}$$

onde, devido à propriedade 10,

$$\varepsilon_i^* \sim SN\left(-\frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}}, 1, \lambda\right),$$

$i = 1, \dots, n$.

Podemos então escrever Y_i como

$$Y_i = E(Y_i) + E_i = a + \frac{\sigma\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}} + \beta x_i + \sigma \varepsilon_i^* .$$

$i = 1, \dots, n$.

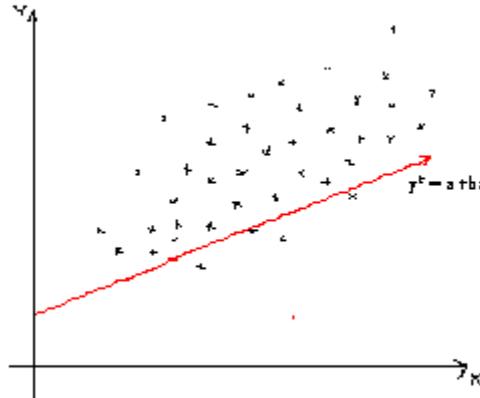


Figura 5.1 : Representação de uma reta de regressão sob o modelo (5.2).

Modelo centrado

Uma outra proposta é o modelo

$$Y_i = \alpha + \beta x_i + \sigma \varepsilon_i, \quad \varepsilon_i \sim SN \left(-\frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}, 1, \lambda \right), \quad i = 1, \dots, n, \quad (5.3)$$

onde $\alpha = a + \frac{\sigma\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}$.

Na representação da Figura 5.2 podem ser visualizadas as retas de regressão correspondentes aos dois modelos, em um mesmo conjunto de dados, com $\lambda > 0$.

Neste novo modelo temos, pela propriedade 12, que

$$E(\varepsilon_i^*) = 0 \quad \text{e} \quad E(Y_i) = \alpha + \beta x_i, \quad i = 1, \dots, n.$$

Portanto, se $\hat{\alpha}$ e $\hat{\beta}$ forem estimadores não viciados, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i$ é um preditor não viciado para $Y|x$, onde $E(\hat{Y}_i) = \alpha + \beta x_i = E(Y_i)$.

Convém observar que os dois modelos são equivalentes, com o mesmo coeficiente angular para a reta de regressão, enquanto que os coeficientes lineares dos dois modelos são tais que $\alpha = a + \sigma \frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}$, mas o modelo (5.3) se mostra mais prático, pois a um profissional de qualquer área não interessa o valor preditivo a menos ou a mais de uma constante, e gostaria de obter a predição centrada.

O modelo a ser adotado como modelo linear simples com erro normal assimétrico com parâmetro de assimetria λ é o modelo centrado e, uma vez adotado este modelo, temos que as funções densidade de ε_i e Y_i são dadas por

$$f(\varepsilon_i) = \frac{2}{\sigma} \phi \left(\frac{\varepsilon_i + \frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}}{\sigma} \right) \Phi \left(\lambda \frac{\varepsilon_i + \frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}}{\sigma} \right),$$

$$f(y_i) = \frac{2}{\sigma} \phi \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right),$$

$i = 1, \dots, n$, onde $\mu = \sigma \frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}$.

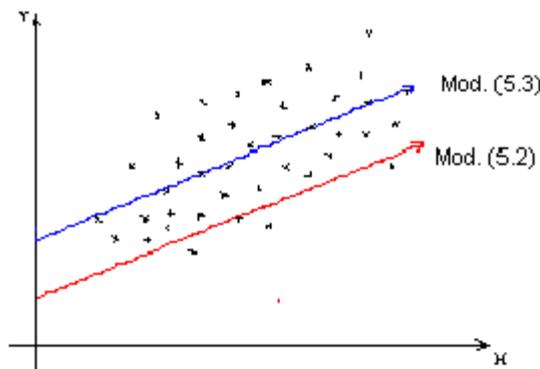


Figura 5.2 : Representação das retas de regressão sob os modelos (5.2) e (5.3).

Então, a função densidade conjunta do vetor \underline{Y} é dada por

$$\begin{aligned} f(\underline{y}) &= \prod_{i=1}^n \frac{2}{\sigma} \phi\left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right) \\ &= \left(\frac{2}{\sigma}\right)^n \prod_{i=1}^n \phi\left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right) \prod_{i=1}^n \Phi\left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right). \end{aligned}$$

Os parâmetros de interesse são α , β , σ^2 e λ , e a função de verossimilhança tem expressão

$$\begin{aligned} L(\alpha, \beta, \sigma^2, \lambda; \mathcal{D}) &\propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2\right\} \\ &\quad \times \prod_{i=1}^n \Phi\left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right). \end{aligned}$$

Neste capítulo são analisados apenas os modelos (5.1) e (5.3). No próximo capítulo são apresentadas mais três propostas de modelo de regressão linear com erros assimétricos.

5.2 Análise bayesiana

5.2.1 Modelo (5.1)

Distribuição a posteriori de α, β e σ^2 considerando independência a priori entre os parâmetros

A função de verossimilhança é dada por

$$L(\alpha, \beta, \sigma^2; \mathcal{D}) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right\}.$$

Supondo independência a priori entre estes parâmetros, são consideradas as seguintes distribuições:

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2) \quad , \quad \beta \sim N(\mu_\beta, \sigma_\beta^2) \quad \text{e} \quad \sigma^2 \sim IG(u, v).$$

A densidade conjunta a posteriori para estes parâmetros é dada por

$$\begin{aligned} \pi(\alpha, \beta, \sigma^2 | \mathcal{D}) &\propto (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right\} \\ &\quad \times \exp\left\{-\frac{1}{2} \left[\frac{1}{\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 + \frac{1}{\sigma_\beta^2} (\beta - \mu_\beta)^2 \right] - \frac{u}{\sigma^2}\right\}, \end{aligned}$$

de onde obtemos as distribuições condicionais requeridas pelo amostrador de Gibbs:

$$\begin{aligned} \alpha | \beta, \sigma^2, \mathcal{D} &\sim N\left(\frac{\sigma_\alpha^2 \sum_{i=1}^n (y_i - \beta x_i) + \mu_\alpha \sigma^2}{n\sigma_\alpha^2 + \sigma^2}, \frac{\sigma^2 \sigma_\alpha^2}{n\sigma_\alpha^2 + \sigma^2}\right), \\ \beta | \alpha, \sigma^2, \mathcal{D} &\sim N\left(\frac{\sigma_\beta^2 \left(\alpha \sum_{i=1}^n x_i + \sum_{i=1}^n x_i y_i\right) + \mu_\beta \sigma^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma^2}, \frac{\sigma^2 \sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma^2}\right) \end{aligned}$$

e

$$\sigma^2 | \alpha, \beta, \mathcal{D} \sim IG\left(u + \frac{n}{2}, \frac{1}{2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 - v\right).$$

Para aplicação destes resultados tomamos um conjunto de dados de Zelner [35], página 64, que relaciona as variáveis renda pessoal e investimento pessoal. Os dados estão no Apêndice A. A renda pessoal é considerada como variável resposta (Y) e o investimento pessoal como variável independente (x), e a este conjunto de dados é ajustado o modelo (5.1). As estimativas bayesianas pelo MCMC (veja programa no apêndice **C1**) foram

$$\hat{\alpha} = 345,70 \quad \text{e} \quad \hat{\beta} = 3,029.$$

enquanto que as estimativas de mínimos quadrados são

$$\hat{\alpha} = 344,70 \quad \text{e} \quad \hat{\beta} = 3,049.$$

Note-se que as estimativas para α e β são equivalentes pelos dois métodos, mas as estimativas pelo método dos mínimos quadrados requer menos tempo, o que torna este último

método o melhor para este caso. Estes resultados serão comparados com os resultados que estão na próxima subseção.

Distribuição a posteriori para α, β e σ^2 , sob distribuição a priori não informativa para os parâmetros.

Considere-se a distribuição a priori

$$\pi(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma}.$$

Assim, a distribuição conjunta a posteriori para os parâmetros é dada por

$$\pi(\alpha, \beta, \sigma^2 | \underline{y}, \underline{x}) \propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right\}.$$

Após operações algébricas vistas em Zelner [35] temos as distribuições a posteriori para os três parâmetros, dadas por

$$\pi(\alpha | \underline{y}, \underline{x}) \propto \left[v + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2 \sum_{i=1}^n \frac{x_i^2}{n}} (\alpha - \hat{\alpha})^2 \right]^{-\frac{v+1}{2}},$$

$$\pi(\beta | \underline{y}, \underline{x}) \propto \left[v + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2} (\beta - \hat{\beta})^2 \right]^{-\frac{v+1}{2}},$$

$$\pi(\sigma^2 | \underline{y}, \underline{x}) \propto \frac{1}{\sigma^{[(n-2)+1]}} \exp \left[-\frac{(n-2) S^2}{2\sigma^2} \right],$$

onde

$$v = n - 2,$$

$$S^2 = v^{-1} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

e

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}_i) (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x}_i)^2}.$$

Em Zellner [35] são apresentadas as estimativas de α e β , através destas distribuições a posteriori, para o exemplo desta subseção e as médias a posteriori para α e β são:

$$\hat{\alpha} = 345,0 \quad \text{e} \quad \hat{\beta} = 3,05.$$

Note-se que os resultados são equivalentes para os três métodos apresentados nesta subseção. No trabalho de Zellner não é apresentada a estimativa de σ^2 , e desta forma não temos como compará-la.

5.2.2 Modelo (5.3)

Distribuição a posteriori de α, β, σ^2 e λ considerando independência a priori entre os parâmetros

Os parâmetros de interesse são α, β, σ^2 e λ , e a função de verossimilhança é dada por

$$L(\alpha, \beta, \sigma^2, \lambda | \mathcal{D}) \propto \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 \right\} \\ \times \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right).$$

onde $\mu = \sigma \frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}$.

Supondo α, β, σ^2 e λ independentes a priori, adotamos as seguintes distribuições:

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2) \quad , \quad \beta \sim N(\mu_\beta, \sigma_\beta^2) \quad , \quad \sigma^2 \sim IG(u, v) \quad \text{e} \quad \lambda \sim N(\mu_\lambda, \sigma_\lambda^2) .$$

A distribuição a posteriori conjunta para α, β, σ^2 e λ é dada por

$$\pi(\alpha, \beta, \sigma^2, \lambda | \mathcal{D}) \propto (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 \right\}$$

$$-\frac{1}{2\sigma_\beta^2} (b - \mu_\beta)^2 - \frac{1}{2\sigma_\lambda^2} (\lambda - \mu_\lambda)^2 - \frac{u}{\sigma^2} \left\} \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right),$$

onde \mathcal{D} denota o conjunto de dados.

As distribuições condicionais requeridas para o algoritmo de Metropolis-Hastings são dadas por

$$\begin{aligned} \pi(\alpha|\beta, \sigma^2, \lambda, \mathcal{D}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 \right\} \\ &\times \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right), \end{aligned}$$

$$\begin{aligned} \pi(\beta|\alpha, \sigma^2, \lambda, \mathcal{D}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2 \right\} \\ &\times \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right), \end{aligned}$$

$$\begin{aligned} \pi(\sigma^2|\alpha, \beta, \lambda, \mathcal{D}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{u}{\sigma^2} \right\} \\ &\times (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) \end{aligned}$$

e

$$\pi(\lambda|\alpha, \beta, \sigma, \mathcal{D}) \propto \exp \left\{ -\frac{1}{2\sigma_\lambda^2} (\lambda - \mu_\lambda)^2 \right\} \times \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right).$$

Distribuição a posteriori para α, β, σ^2 e λ considerando uma distribuição a priori não informativa para o vetor $(\alpha, \beta, \sigma^2)$

Neste caso, a fim de obter as distribuições a posteriori para α, β, σ^2 e λ considerarmos que

$$\lambda \Pi(\alpha, \beta, \sigma^2).$$

Temos três possibilidades de distribuições a priori para λ e $(\alpha, \beta, \sigma^2)$:

- (i) $\pi(\alpha, \beta, \sigma^2)$ arbitrária, $\lambda \sim N(a, b^2)$,
- (ii) $\pi(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}$, $\lambda \sim N(a, b^2)$ e
- (iii) $\pi(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}$, $\pi(\lambda)$ arbitrária.

Sob a representação de Henze em (4.2) temos que

$$Y_i | \alpha, \beta, \sigma^2, \lambda, \tau_i \stackrel{ind.}{\sim} N \left(\frac{\sigma \lambda \tau_i}{\sqrt{1 + \lambda^2}} + \alpha + \beta x_i, \frac{\sigma^2}{1 + \lambda^2} \right),$$

$$\tau_i \sim HN(0, 1) \quad e \quad \tau_i \amalg (\alpha, \beta, \sigma^2, \lambda)$$

$i = 1, \dots, n$.

Seja $w_i = \psi \tau_i$, onde ψ^2 passa a ser o parâmetro de escala dado por

$$\psi = \frac{\sigma^2}{\sqrt{1 + \lambda^2}}.$$

Podemos então reescrever a distribuição de Y_i como

$$Y_i | a, b, \psi^2, \lambda, w_i \sim N(\lambda w_i + \alpha + \beta x_i, \psi^2),$$

$$w_i | \psi^2 \stackrel{iid}{\sim} HN(0, \psi^2) \quad e \quad w_i \amalg (a, b, \lambda).$$

Como quarta possibilidade considere-se as distribuições a priori

$$\pi(\alpha, \beta, \psi^2 | \lambda) \propto \frac{1}{\psi^2} \quad e \quad \lambda \sim N(a, b^2).$$

As distribuições a posteriori para $\alpha, \beta, w, \psi^2, \lambda$, onde $\underline{\beta} = (\alpha, \beta)$, são dadas por

$$\underline{\beta} | \psi^2, \lambda, w, \underline{y} \sim N \left(\widehat{\underline{\beta}}(\underline{y}) - \lambda \widehat{\underline{\beta}}(\underline{w}), \psi^2 (X'X)^{-1} \right),$$

$$\underline{w} | \underline{\beta}, \psi^2, \lambda, \underline{y} \sim TN_n \left(\underline{0}; \frac{\lambda}{1 + \lambda^2} \epsilon, \frac{\psi^2}{1 + \lambda^2} I_n \right),$$

$$\psi^2 | \underline{\beta}, \lambda, \underline{w}, \underline{y} \sim IG \left(\frac{n}{2}, \frac{\|\underline{\epsilon} - \lambda \underline{w}\|^2}{2} \right),$$

$$\lambda | \underline{\beta}, \psi^2, \underline{w}, \underline{y} \sim N \left(\frac{b^2 \underline{w}' \underline{\epsilon} + a \psi^2}{b^2 \|\underline{w}\|^2 + \psi^2}, \frac{b^2 \psi^2}{b^2 \|\underline{w}\|^2 + \psi^2} \right),$$

onde $\|\underline{x}\| = \sqrt{\sum_{i=1}^n x_i}$, $\hat{\underline{\beta}}(\underline{y}) = (X'X)^{-1} X' \underline{y}$ e $\underline{\epsilon} = \underline{y} - X\beta$. Mais detalhes são discutidos em Arellano-Valle et al. [1].

5.2.3 Método dos uns

Usaremos a densidade (2.8) para descrever o método que aqui chamamos de método dos uns.

A densidade de Y_i é dada por

$$f(y_i | \alpha, \beta, \sigma^2, \lambda) = \frac{2}{\sigma} \phi \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) \quad (5.4)$$

e a função de verossimilhança para α , b , σ^2 e λ é dada por

$$L(\alpha, \beta, \sigma^2, \lambda | \underline{x}, \underline{y}) = \prod_{i=1}^n \frac{2}{\sigma} \phi \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) \Phi \left(\lambda \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) \quad (5.5)$$

onde $\mu = -\frac{\lambda}{\sqrt{1+\lambda^2}} \sigma \sqrt{\frac{2}{\pi}}$.

Sejam U_i , $i = 1, 2, \dots, n$, variáveis aleatórias independentes, com distribuição de Bernoulli, de parâmetro p_i , onde

$$p_i = f(y_i | \alpha, \beta, \sigma^2, \lambda) \quad \text{e} \quad 0 < p_i < 1. \quad (5.6)$$

Então, a função de verossimilhança para α , β , σ^2 e λ , em termos das variáveis U_i , é dada por

$$L_{U_i}(\alpha, \beta, \sigma^2, \lambda) = \prod_{i=1}^n p_i^{u_i} (1 - p_i)^{1-u_i}. \quad (5.7)$$

Considere-se $U_1 = 1, U_2 = 1, \dots, U_n = 1$, resultado de uma amostra aleatória das

variáveis U_i , onde tenha ocorrido o valor 1 para $i = 1, \dots, n$. Então, a função de verossimilhança dada por (5.7) se reduz à forma (5.5). Este artifício é aceito pela linguagem BUGS. Então, sob esta condição passemos a considerar a verossimilhança dada por (5.5), para α , b , σ^2 e λ , ou seja,

$$L(\alpha, b, \sigma^2, \lambda | \underline{x}, \underline{y}, \underline{u} = \underline{1}) = \prod_{i=1}^n \frac{2}{\sigma} \phi \left(\frac{y_i - \alpha - bx_i - \mu}{\sigma} \right) \Phi \left(\lambda \frac{y_i - \alpha - bx_i - \mu}{\sigma} \right) \quad (5.8)$$

A partir de (5.8) vale o método bayesiano conhecido. Este método dos uns é usado neste capítulo, em simulação, e no capítulo 7.

5.2.4 Simulação

Efetuamos uma simulação em um modelo de regressão linear simples com erros normais assimétricos (veja programa em **R** no apêndice **B.2**). Os passos são os seguintes:

- i) A variável X assume valores de 1 a 10, com cada valor replicado 100 vezes.
- ii) Os valores dos parâmetros são: $\alpha = 5$, $\beta = 0.6$, $\sigma^2 = 9$, $\mu = 2.140949$ e $\lambda = -2$.
- iii) Os erros ε_i , $i = 1, 2, \dots, 1000$, foram gerados através da implementação da representação de Henze (veja método de construção da distribuição normal assimétrica pela representação de Henze, capítulo 2, com $\varepsilon \sim SN(0, \sigma^2, \lambda)$).
- iv) Os valores da variável Y , variável resposta, foram gerados na forma

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

com valor esperado $E(y_i) = \alpha + \beta x_i$, $i = 1, \dots, n$. Nos gráficos da Figura 5.3 temos as densidades e os histogramas de Y e do erro simulados.

Os parâmetros do modelo, foram estimados pelo MCMC, através do programa **C.1**. Fora considerada independência a priori entre os parâmetros e foram utilizadas as seguintes distribuições:

$$\alpha \sim N(0, 1000), \beta \sim N(0, 1000), \sigma^2 \sim IG(0.01, 0.01) \text{ e } \lambda \sim N(0, 1000)$$

Foram geradas três cadeias com os seguintes valores iniciais:

cadeia 1: $(\tau = 1, \lambda = 2, \alpha = 0, \beta = 0)$,

cadeia 2: $(\tau = 2, \lambda = -1, \alpha = 0.5, \beta = 3)$ e

cadeia 3: $(\tau = 4, \lambda = 1, \alpha = 1, \beta = 2)$.

Foram tomadas 40000 amostras, descartadas as 10000 interações iniciais e dado um salto de tamanho 10. Os resultados encontram-se a seguir. Na tabela 5.1 apresentamos o indicador de convergência \hat{R} (*redução potencial estimada da escala [15]*) que é um indicador de convergência das cadeias. Nas tabelas dos sumários a posteriori do capítulo 7 também é usado o indicador \hat{R} para verificar a convergência das cadeias.

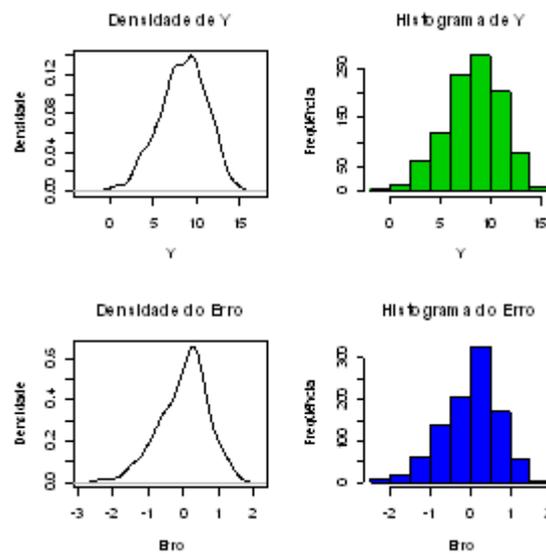


Figura 5.3 : Densidades e histogramas de Y e do erro para os dados simulados.

Tabela 5.1 : Sumário a posteriori para os parâmetros sob o modelo (5.3).

Parâmetros	média	D.P.	2.5%	Mediana	97.5%	\widehat{R}
α	5,001	0,1380	4,717	5,003	5,266	1,001
β	0,6153	0,0215	0,5721	0,6139	0,6568	1,001
δ	-0,8962	0,0219	-0,9316	-0,8986	-0,846	1,002
λ	-2,057	0,2497	-2,564	-2,048	-1,587	1,003
μ	2,169	0,1313	1,900	2,170	2,417	1,001
σ^2	9,199	0,7248	7,856	9,162	10,66	1,001

Como esperado os resultados estão muito próximos aos valores escolhidos na geração dos dados. Os históricos das cadeias de Markov para os parâmetros sob o modelo (5.3) são apresentados nas Figuras 5.4 e 5.5. Note-se que visualmente não existem problemas de convergência nas cadeias geradas. Com o salto igual a 10 foi evitado o problema de correlação nas cadeias geradas, o que é constatado na Figura 5.6. Pelo mesmo motivo, o tamanho do salto usado na geração das cadeias referentes aos resultados do capítulo 6 também é igual a 10.

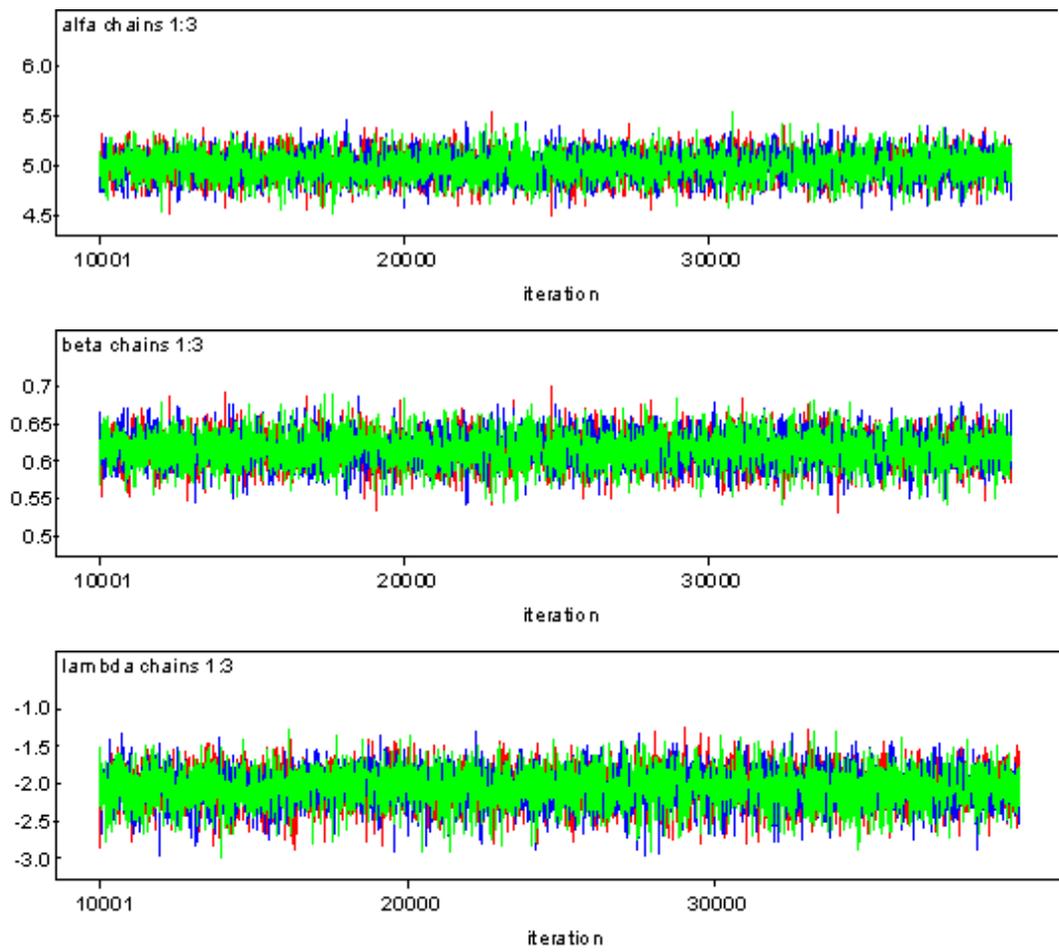


Figura 5.4 : Históricos das cadeias de Markov para α , β e λ .

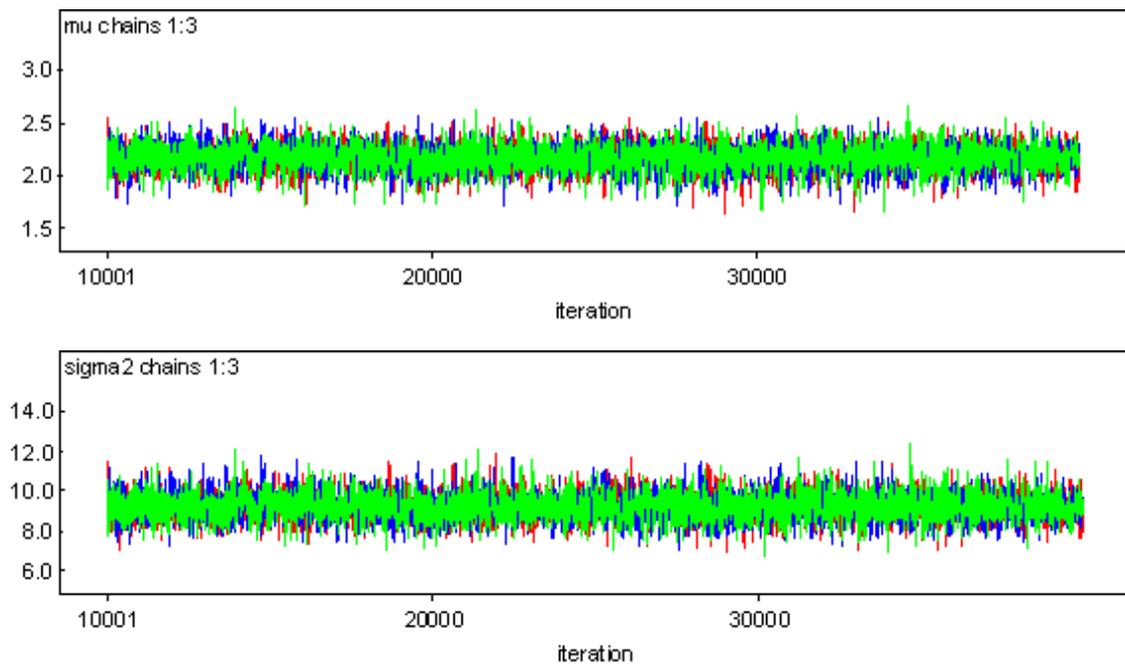


Figura 5.5 : Históricos das cadeias de Markov para μ e σ^2 .

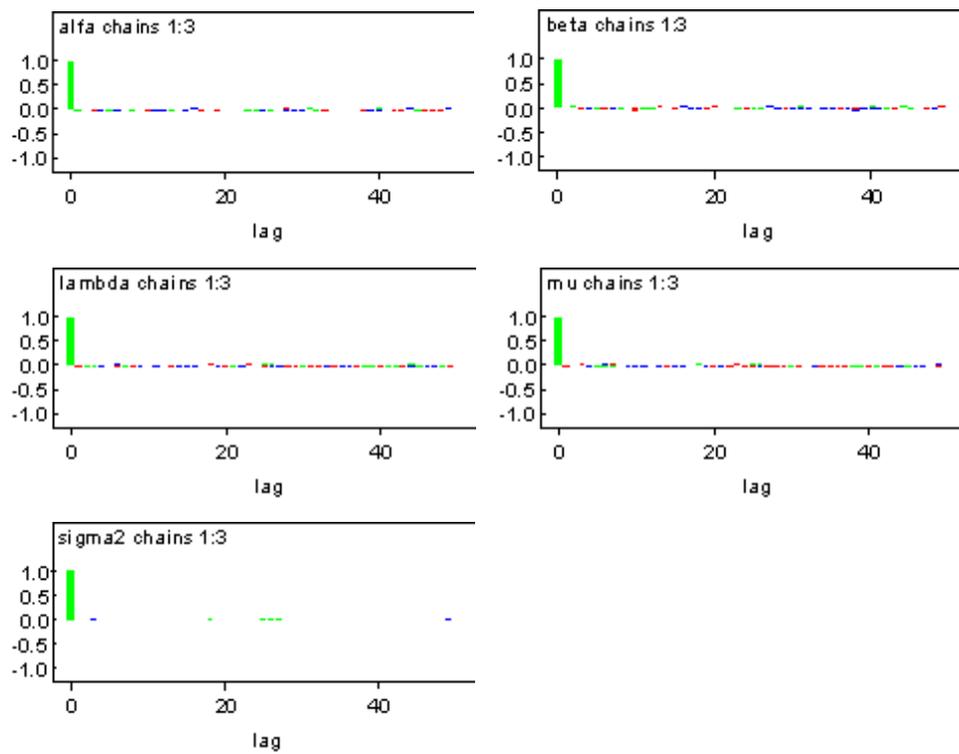


Figura 5.6 : Correlações das cadeias geradas.

Capítulo 6

Modelos alternativos para o erro

Neste capítulo são propostos três outros modelos como alternativa aos modelos (5.1) e (5.3). O primeiro é o modelo flexível de Ma & Genton [22], mais exatamente o modelo FGSN de ordem 3 que prevê bimodalidade (veja seção 3.7.1). No segundo modelo é proposto que a variável resposta é estritamente positiva, supondo que os erros são normais e é adicionado um efeito aleatório com distribuição normal para para cada observação. No último modelo os erros são normais assimétricos e têm uma componente aleatória que, para o conjunto de dados reais considerados nesta dissertação, é devida ao fator sexo.

6.1 Modelo flexível de Ma & Genton (2004)

Considere-se o modelo

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim FGSN(\mu, \sigma^2, \beta_1, \beta_3), \quad i = 1, \dots, n. \quad (6.1)$$

É importante observar que para os modelos flexíveis de Ma & Genton ainda não existem na literatura expressões fechadas para a esperança e para a variância. Aliás, esta é uma classe de densidades ainda em discussão e seu histórico de exploração ainda é recente.

6.1.1 Distribuição a posteriori de $\alpha, \beta, \sigma^2, \mu, \beta_1$ e β_3

Os parâmetros de interesse são $\alpha, \beta, \sigma^2, \mu, \beta_1$ e β_3 e a função de verossimilhança é dada por

$$L(\alpha, \beta, \sigma^2, \mu, \beta_1, \beta_3 / \mathcal{D}) \propto \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \beta_1 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right) + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right)^3 \right\}.$$

Supondo $\alpha, \beta, \sigma^2, \mu, \beta_1$ e β_3 independentes, são consideradas as seguintes priors:

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \beta \sim N(\mu_\beta, \sigma_\beta^2), \quad \sigma^2 \sim IG(u, v),$$

$$\mu \sim N(\mu_\mu, \sigma_\mu^2), \quad \beta_1 \sim N(\mu_1, \sigma_1^2) \quad \text{e} \quad \beta_3 \sim N(\mu_3, \sigma_3^2).$$

A distribuição a posteriori conjunta para $\alpha, \beta, \sigma^2, \mu, \beta_1$ e β_3 é dada por

$$\pi(\alpha, \beta, \sigma^2, \mu, \beta_1, \beta_3 / \mathcal{D}) \propto (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 - \frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2 - \frac{1}{2\sigma_\mu^2} (\mu - \mu_\mu)^2 \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma_1^2} (\beta_1 - \mu_1)^2 - \frac{1}{2\sigma_3^2} (\beta_3 - \mu_3)^2 - \frac{u}{\sigma^2} \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right)^3 \right\}.$$

As distribuições condicionais necessárias à implementação do algoritmo de Metropolis-Hastings são dadas por

$$\pi(\alpha | \beta, \sigma^2, \mu, \beta_1, \beta_3, \mathcal{D}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma} \right)^3 \right\},$$

$$\begin{aligned}
\pi(\beta|\alpha, \sigma^2, \mu, \beta_1, \beta_3, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{1}{2\sigma_\alpha^2} (\beta - \mu_\beta)^2\right\} \\
&\quad \times \prod_{i=1}^n \Phi\left\{\beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right)^3\right\}, \\
\pi(\sigma^2|\alpha, \beta, \mu, \beta_1, \beta_3, \mathcal{D}) &\propto (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{u}{\sigma^2}\right\} \\
&\quad \times \prod_{i=1}^n \Phi\left\{\beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right)^3\right\}, \\
\pi(\mu|\alpha, \beta, \sigma^2, \beta_1, \beta_3, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \mu)^2 - \frac{1}{2\sigma_\mu^2} (\mu - \mu_\mu)^2\right\} \\
&\quad \times \prod_{i=1}^n \Phi\left\{\beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right)^3\right\}, \\
\pi(\beta_1|\alpha, \beta, \sigma^2, \mu, \beta_3, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma_1^2} (\beta_1 - \mu_1)^2\right\} \\
&\quad \times \prod_{i=1}^n \Phi\left\{\beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right)^3\right\}
\end{aligned}$$

e

$$\begin{aligned}
\pi(\beta_3|\alpha, \beta, \sigma^2, \mu, \beta_1, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma_3^2} (\beta_3 - \mu_3)^2\right\} \\
&\quad \times \prod_{i=1}^n \Phi\left\{\beta_1 \frac{y_i - \alpha - \beta x_i - \mu}{\sigma} + \beta_3 \left(\frac{y_i - \alpha - \beta x_i - \mu}{\sigma}\right)^3\right\}.
\end{aligned}$$

6.2 Modelo com efeito aleatório da observação

Considere-se o modelo

$$y_i = \alpha + \beta x_i + \gamma_i + \varepsilon_i, \quad \gamma_i \stackrel{iid}{\sim} N(0, \sigma_\gamma^2), \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2), \quad \gamma_i \perp \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (6.2)$$

onde ε_i é o erro de cada observação e γ_i o seu efeito aleatório.

Neste modelo o efeito de cada observação é desvinculado do erro e na prática pode

haver restrição no modelo, como é o caso da variável *Bfat* que é a variável resposta do conjunto de dados reais aqui considerado (veja capítulo 6). A variável *Bfat* representa a porcentagem de gordura no organismo e deve ser positiva. Neste caso temos que

$$\begin{aligned} Y_i &\geq 0 \\ \implies E(\alpha + \beta x_i + \gamma_i + \varepsilon_i | \gamma_i) &\geq 0 \\ \implies \alpha + \beta x_i + \gamma_i &\geq 0 \\ \implies \gamma_i &\geq -(\alpha + \beta x_i), \end{aligned}$$

ou seja,

$$\begin{aligned} \gamma_i &> -\tau_i \\ \implies \gamma_i + \tau_i &> 0. \end{aligned} \tag{6.3}$$

onde $\tau_i = \alpha + \beta x_i$, $i = 1, 2, \dots, n$.

O objetivo é obter a distribuição condicional de Y dado que $\gamma_i \geq -(\alpha + \beta x_i)$, ou, de outra forma, obter a distribuição de $Y_i - \alpha - \beta x_i | \gamma_i + \tau_i$. Para isto defina-se o vetor aleatório normal bivariado

$$Z_i = \begin{pmatrix} \gamma_i \\ Y_i - \alpha - \beta x_i \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} \gamma_i \\ \varepsilon_i \end{pmatrix},$$

de modo que

$$Z_i \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_\gamma^2 & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\varepsilon^2 + \sigma_\gamma^2 \end{bmatrix} \right).$$

Considere-se as reparametrizações

$$U_0 = \frac{\gamma_i}{\sigma_\gamma} \quad \text{e} \quad U_1 = \frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}.$$

Logo,

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \frac{\sigma_\gamma}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \\ \frac{\sigma_\gamma}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} & 1 \end{bmatrix} \right).$$

A quantidade

$$\delta = \frac{\sigma_\gamma}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} = \left(\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2 + \sigma_\gamma^2} \right)^{1/2} \quad (6.4)$$

mede a correlação entre as variáveis aleatórias marginais U_0 e U_1 . Desta, obtemos outra quantidade que, similarmente ao λ da normal assimétrica também será chamada de λ e é dada pela relação

$$\lambda = \frac{\delta}{\sqrt{1 - \delta^2}} = \frac{\sigma_\gamma}{\sigma_\varepsilon}.$$

Temos também que

$$\delta^2 = \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2 + \sigma_\gamma^2}$$

e

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix} \right).$$

Da condição (6.3) temos que

$$U_0 + \frac{\tau_i}{\sigma_\gamma} > 0.$$

Portanto, devido ao resultado de Arellano-Valle & Azzalini [2], veja seção 2.5, a densidade de Y_i é dada por

$$f(y_i) = \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \phi \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \frac{\Phi \left(\frac{\tau_i}{\sigma_\gamma \sqrt{1 - \delta^2}} + \frac{\delta}{\sqrt{1 - \delta^2}} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right)}{\Phi \left(\frac{\tau_i}{\sigma_\gamma} \right)}. \quad (6.5)$$

Pela expressão de δ em (6.4) temos que

$$f(y_i) = \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \phi \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \frac{\Phi \left(\frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right)}{\Phi \left(\frac{\tau_i}{\sigma_\gamma} \right)}, \quad (6.6)$$

de forma que a função de verossimilhança é dada por

$$L(\alpha, \beta, \sigma_\varepsilon^2, \sigma_\gamma^2 \mid \mathcal{D}) \propto \frac{1}{(\sigma_\varepsilon^2 + \sigma_\gamma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right\}$$

$$\times \prod_{i=1}^n \frac{\Phi \left\{ \frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right\}}{\Phi \left(\frac{\tau_i}{\sigma_\gamma} \right)}.$$

6.2.1 Distribuição a posteriori de α , β , σ_ε^2 , σ_γ^2 , δ e λ

Supondo α , β , σ_ε^2 e σ_γ^2 independentes, são consideradas as seguintes distribuições a priori:

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \beta \sim N(\mu_\beta, \sigma_\beta^2), \quad \sigma_\varepsilon^2 \sim IG(u_\varepsilon, v_\varepsilon), \quad \text{e} \quad \sigma_\gamma^2 \sim IG(u_\gamma, v_\gamma)$$

A distribuição a posteriori conjunta para α , β , σ_ε^2 e σ_γ^2 é dada por

$$\begin{aligned} \pi(\alpha, \beta, \sigma_\varepsilon^2, \sigma_\gamma^2 | \mathcal{D}) &\propto \frac{1}{(\sigma_\varepsilon^2 + \sigma_\gamma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right\} \\ &\times \exp \left\{ -\frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 - \frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2 - \frac{u_\varepsilon}{\sigma_\varepsilon^2} - \frac{u_\gamma}{\sigma_\gamma^2} \right\} \\ &\times (\sigma_\varepsilon^2)^{-(v_\varepsilon+1)} (\sigma_\gamma^2)^{-(v_\gamma+1)} \prod_{i=1}^n \frac{\Phi \left\{ \frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right\}}{\Phi \left(\frac{\tau_i}{\sigma_\gamma} \right)}. \end{aligned}$$

As distribuições condicionais utilizadas no algoritmo de Metropolis-Hastings são as seguintes:

$$\begin{aligned} \pi(\alpha | \beta, \sigma_\varepsilon^2, \sigma_\gamma^2, \mathcal{D}) &\propto \exp \left\{ -\frac{1}{2\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 - \frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 \right\} \\ &\times \prod_{i=1}^n \Phi \left\{ \frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right\}, \\ \pi(\beta | \alpha, \sigma_\varepsilon^2, \sigma_\gamma^2, \mathcal{D}) &\propto \exp \left\{ -\frac{1}{2\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 - \frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2 \right\} \\ &\times \prod_{i=1}^n \Phi \left\{ \frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right\}, \end{aligned}$$

$$\pi(\sigma_\varepsilon^2 | \alpha, \beta, \sigma_\gamma^2, \mathcal{D}) \propto \frac{(\sigma_\varepsilon^2)^{-(v_\varepsilon+1)}}{(\sigma_\varepsilon^2 + \sigma_\gamma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 - \frac{u_\varepsilon}{\sigma_\varepsilon^2} \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right\},$$

e

$$\pi(\sigma_\gamma^2 | \alpha, \beta, \sigma_\varepsilon^2, \mathcal{D}) \propto \frac{(\sigma_\gamma^2)^{-(v_\gamma+1)}}{(\sigma_\varepsilon^2 + \sigma_\gamma^2)^{n/2}} \exp \left\{ -\frac{u_\gamma}{\sigma_\gamma^2} \right\} \\ \times \prod_{i=1}^n \frac{\Phi \left\{ \frac{\tau_i \sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}}{\sigma_\gamma \sigma_\varepsilon} + \frac{\sigma_\gamma}{\sigma_\varepsilon} \left(\frac{Y_i - \alpha - \beta x_i}{\sqrt{\sigma_\varepsilon^2 + \sigma_\gamma^2}} \right) \right\}}{\Phi \left(\frac{\tau_i}{\sigma_\gamma} \right)}.$$

6.3 Modelo com efeito aleatório de variável binária

Considere-se o modelo

$$Y_i = \alpha + \beta x_i + Z_i' \gamma + \sigma \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} SN(0, 1, \lambda), \quad i = 1, \dots, n, \quad (6.7)$$

onde

$$Z_i' = (I_i, 1 - I_i), \\ I_i = \begin{cases} 1, & \text{observação tipo 1} \\ 0, & \text{observação tipo 2.} \end{cases}$$

e

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix},$$

com

$$\gamma_1 \text{ II } \gamma_2.$$

O vetor γ é o efeito aleatório da variável binária no modelo. Para o conjunto de dados usado no capítulo 7 temos que

$$I_i = \begin{cases} 1, & \text{observação do sexo masculino} \\ 0, & \text{observação do sexo Feminino.} \end{cases}$$

6.3.1 Distribuição a posteriori de $\alpha, \beta, \sigma^2, \gamma_1, \gamma_2$ e λ

A função de verossimilhança é dada por

$$L(\alpha, \beta, \sigma^2, \gamma_1, \gamma_2, \lambda | \mathcal{D}) \propto \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma} \right\}.$$

Supondo $\alpha, \beta, \sigma^2, \gamma_1, \gamma_2$ e λ independentes, são consideradas as seguintes distribuições a priori:

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \beta \sim N(\mu_\beta, \sigma_\beta^2), \quad \sigma^2 \sim IG(u, v), \\ \gamma_1 \sim N(\mu_1, \sigma_1^2), \quad \gamma_2 \sim N(\mu_2, \sigma_2^2), \quad \text{e} \quad \lambda \sim N(\mu_\lambda, \sigma_\lambda^2).$$

A distribuição a posteriori conjunta para $\alpha, \beta, \sigma^2, \gamma_1, \gamma_2$ e λ é dada por

$$\pi(\alpha, \beta, \sigma^2, \gamma_1, \gamma_2, \lambda | \mathcal{D}) \propto (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 - \frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2 - \frac{1}{2\sigma_\lambda^2} (\lambda - \mu_\lambda)^2 \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma_1^2} (\gamma_1 - \mu_1)^2 - \frac{1}{2\sigma_2^2} (\gamma_2 - \mu_2)^2 - \frac{u}{\sigma^2} \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma} \right\},$$

de onde obtemos as distribuições condicionais

$$\pi(\alpha | \beta, \sigma^2, \gamma_1, \gamma_2, \lambda, \mathcal{D}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 - \frac{1}{2\sigma_\alpha^2} (\alpha - \mu_\alpha)^2 \right\} \\ \times \prod_{i=1}^n \Phi \left\{ \lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma} \right\},$$

$$\begin{aligned} \pi(\beta|\alpha, \sigma^2, \gamma_1, \gamma_2, \lambda, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 - \frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2\right\} \\ &\quad \times \prod_{i=1}^n \Phi\left\{\lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma}\right\}, \end{aligned}$$

$$\begin{aligned} \pi(\sigma^2|\alpha, \beta, \gamma_1, \gamma_2, \lambda, \mathcal{D}) &\propto (\sigma^2)^{-\left(\frac{n}{2} + v + 1\right)} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 - \frac{u}{\sigma^2}\right\} \\ &\quad \times \prod_{i=1}^n \Phi\left\{\lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma}\right\}, \end{aligned}$$

$$\begin{aligned} \pi(\gamma_1|\alpha, \beta, \sigma^2, \gamma_2, \lambda, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 - \frac{1}{\sigma_1^2} (\gamma_1 - \mu_1)^2\right\} \\ &\quad \times \prod_{i=1}^n \Phi\left\{\lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma}\right\}, \end{aligned}$$

$$\begin{aligned} \pi(\gamma_2|\alpha, \beta, \sigma^2, \gamma_1, \lambda, \mathcal{D}) &\propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - Z'_i \gamma)^2 - \frac{1}{\sigma_2^2} (\gamma_2 - \mu_2)^2\right\} \\ &\quad \times \prod_{i=1}^n \Phi\left\{\lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma}\right\} \end{aligned}$$

e

$$\pi(\lambda|\alpha, \beta, \sigma^2, \gamma_1, \gamma_2, \mathcal{D}) \propto \prod_{i=1}^n \Phi\left\{\lambda \frac{Y_i - \alpha - \beta x_i - Z'_i \gamma}{\sigma}\right\}.$$

Capítulo 7

Aplicação

O método dos momentos foi de fundamental importância para se obter a expressão matemática para esperança e para a variância da normal assimétrica deduzidos por Azzalini [3], bem como para a dedução dos estimadores dos parâmetros envolvidos. Estes estimadores fazem parte do Capítulo 3 deste trabalho e foram apresentados por Branco & Arellano-Valle [9]. Estes resultados já discutidos no Capítulo 4 apresentam estimadores que não se comportam muito bem no caso de uma distribuição normal. Branco & Arellano-Valle [9] propõem também o uso do algoritmo EM para a estimação dos parâmetros pelo método de máxima verossimilhança. Branco & Dey [11] usam o algoritmo EM para estimar os parâmetros de uma regressão linear simples onde os erros são independentes e têm distribuição t assimétrica. Neste trabalho daremos preferência à abordagem bayesiana devido aos melhores resultados obtidos. Desta forma, a menos de menção em contrário, por questão de comodidade quando se falar em métodos de inferência estaremos nos referindo à abordagem bayesiana.

Neste capítulo trabalhamos com um conjunto de dados extraído do arquivo "ais data", incluído na biblioteca "SN" do sistema R, disponível disponível para download em <http://azzalini.stat.unipd.it/SN/index.html>. Este conjunto de dados consiste de variáveis indicadoras do condicionamento físico de 202 atletas, sendo 100 do sexo feminino e 102 do sexo masculino. Trata-se de parte de um conjunto que contém 11 variáveis, das quais selecionamos as variáveis $Bfat$, indicadora da porcentagem de gordura corpórea, e Wt ,

indicadora da massa do atleta em quilogramas. A variável $Bfat$ é dada por

$$Bfat = \frac{Wt}{Ht^2},$$

onde Ht é dada em centímetros no conjunto de dados, mas na expressão acima precisa ser transformada em metros. Este trabalho consiste em ajustar ao conjunto de dados um modelo de regressão linear simples com $Bfat$ como variável resposta e Wt como variável independente, onde os erros são considerados independentes e para estes são propostos e discutidos os cinco modelos apresentados nos capítulos 5 e 6.

Na seção 7.1 são discutidos os resultados da estimação dos parâmetros de uma regressão linear simples para cada um dos cinco modelos apresentados, baseada no método dos uns, disponível na linguagem BUGS. O outro método também já mencionado usa a representação de Henze e tem se mostrado equivalente ao método dos uns. Na seção 7.2 a discussão será com relação ao tipo de modelo mais adequado ao conjunto de dados citado. Esta discussão será com base nas critérios de seleção de modelos **DIC** (*Deviance Information Criterion* [25]) e **BIC** (*Bayesian Information Criterion* [31]), veja Apêndice **D**. Na conclusão justificaremos o modelo com efeito aleatório do sexo como preferível aos outros modelos. Na Figura 7.1 são apresentados o histograma e a densidade da variável $Bfat$ e na Figura 7.2 é apresentado o gráfico de dispersão da relação entre as variáveis Wt e $Bfat$.

É conveniente aqui uma melhor nomenclatura dos modelos, que apresentamos abaixo.

Modelo 1: Modelo (5.1), com distribuição normal para os erros;

Modelo 2: Modelo (5.3), com distribuição normal assimétrica para os erros;

Modelo 3: Modelo (6.1), com distribuição flexível de Ma & Genton para os erros;

Modelo 4: Modelo (6.2), com distribuição normal assimétrica estendida (Arellano-Valle & Azzalini, [2]) para os erros e o efeito aleatório de cada indivíduo;

Modelo 5: Modelo (6.7), normal assimétrico com efeito aleatório de variável binária. Neste conjunto de dados o sexo compõe os dois tipos de variável.

7.1 Estimação dos parâmetros

Modelo 1

Usaremos agora o método bayesiano para estimar os parâmetros, através do programa do Apêndice **C.2**, supondo $\lambda = 0$, ou seja, supondo o Modelo (5.1) da seção 4.1.1. Foram consideradas as seguintes distribuições a priori:

$$\alpha \sim N(0.001, 1000),$$

$$\beta \sim N(0.001, 1000),$$

e

$$\sigma^2 \sim IG(0.01, 0.01).$$

Foram geradas três cadeias com os seguintes valores iniciais:

Cadeia 1: $(\tau = 0.1, \alpha = 2, \beta = 0)$,

Cadeia 2: $(\tau = 0.4, \alpha = 0.5, \beta = 3)$

e

Cadeia 3: $(\tau = 0.5, \alpha = 15, \beta = 2)$.

As estatísticas, DIC e BIC obtidos estão relacionados abaixo.

Tabela 7.1 : Sumário a posteriori para os parâmetros do **Modelo 1**.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\hat{R}
α	13,4	2,39	8,762	13,4	18,1	1,001
β	0,0014	0,031	-0,06	0,0014	0,064	1,001
σ^2	38,91	3,929	31,98	38,64	47,44	1,001

DIC = 1314,650; **BIC** = 1327,47.

Modelo 2

Supondo agora o modelo (5.3) da seção 4.1.2 apresentamos as estatísticas e os valores do DIC e BIC, programa do Apêndice C.3. Foram consideradas as seguintes distribuições a priori:

$$\alpha \sim N(0.001, 1000),$$

$$\beta \sim N(0.001, 1000),$$

$$\sigma^2 \sim IG(0.01, 0.01),$$

e

$$\lambda \sim N(0.001, 1000).$$

Foram geradas três cadeias com os seguintes valores iniciais:

Cadeia 1: $(\tau = 0.2, \lambda = 20, \alpha = 2, \beta = 0)$,

Cadeia 2: $(\tau = 0.9, \lambda = 15, \alpha = 0.5, \beta = 3)$

e

Cadeia 3: $(\tau = 0.5, \lambda = 60, \alpha = 15, \beta = 2)$.

Tabela 7.2 : Sumário a posteriori para os parâmetros do **Modelo 2**.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\hat{R}
α	12,18	1,131	10,07	12,12	14,55	1,000
β	0,0197	0,01434	-0,01077	0,02066	0,04586	1,001
δ	0,9997	2,908E-4	0,9989	0,9998	0,9999	1,001
λ	47,61	16,5	21,53	45,64	85,32	1,000
μ	-7,91	0,4005	-8,74	-7,891	-7,172	1,000
σ^2	98,59	10,03	80,87	97,87	120,0	1,001

DIC = 1230, 880; **BIC** = 1259, 33.

Modelo 3

Supondo agora o modelo (6.1) da seção 5.1 apresentamos as estatísticas e os valores do DIC e BIC, programa do Apêndice C.4. Foram consideradas as seguintes distribuições a priori:

$$\alpha \sim N(0.001, 1000),$$

$$\beta \sim N(0.001, 1000),$$

$$\sigma^2 \sim IG(0.01, 0.01),$$

$$\beta_1 \sim N(0.001, 1000)$$

e

$$\beta_3 \sim N(0.001, 1000).$$

Foram geradas três cadeias com os seguintes valores iniciais:

Cadeia 1: $(\tau = 0.2, \lambda = 20, \alpha = 2, \beta_1 = 0, \beta_3 = 1, \mu = 5)$,

Cadeia 2: $(\tau = 0.9, \lambda = 15, \alpha = 0.5, \beta_1 = 3, \beta_3 = 3, \mu = 7)$

e

Cadeia 3: $(\tau = 0.5, \lambda = 60, \alpha = 15, \beta_1 = 2, \beta_3 = 4, \mu = 6)$.

Tabela 7.3 : Sumário a posteriori para os parâmetros do **Modelo 3**.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\hat{R}
α	7,67	19,63	-39,03	8,594	41,58	1,006
β	0,01962	0,01411	-0,0109	0,02072	0,04512	1,000
β_1	48,4	16,79	21,53	46,36	86,09	0,999
β_3	22,02	20,06	-4,676	18,32	69,22	0,999
μ	-3,395	19,64	-37,26	-4,268	43,37	1,007
σ^2	98,32	9,973	80,58	97,67	119,7	1,001

DIC =1230, 590; **BIC** =1259, 32.

Modelo 4

Supondo o modelo (6.2) da seção 5.2 apresentamos as estatísticas e os valores do DIC e BIC, programa do Apêndice C.5. Foram consideradas as seguintes distribuições a priori:

$$\alpha \sim N(0.001, 1000),$$

$$\beta \sim N(0.001, 1000),$$

$$\sigma_\varepsilon^2 \sim IG(0.01, 0.01),$$

$$\sigma_\varepsilon^2 \sim IG(0.01, 0.01),$$

e

$$\sigma_\gamma^2 \sim IG(0.5, 0.1).$$

Foram geradas três cadeias com os seguintes valores iniciais:

Cadeia 1: $(\tau = 0.2, \alpha = 2, \beta = 0, \tau_2 = 0.1)$,

Cadeia 2: $(\tau = 0.9, \alpha = 0.5, \beta = 3, \tau_2 = 0.2)$

e

Cadeia 3: $(\tau = 0.5, \alpha = 15, \beta = 2, \tau_2 = 0.7)$.

Tabela 7.4 : Sumário a posteriori para os parâmetros do **Modelo 4**.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\hat{R}
α	13,13	2,704	7,824	13,13	18,46	0,998
β	2,618E-4	0,03536	-0,07038	5,781E-4	0,0697	1,003
δ	0,9089	0,2477	0,06701	0,9968	0,9999	0,821
λ	19,89	21,1	0,06716	12,39	78,24	0,995
σ_ε^2	4,454	10,9	0,007271	0,2816	40,1	0,904
σ_γ^2	39,19	13,5	0,1754	42,1	55,53	0,949

DIC = -1284, 500; **BIC** = 1335, 93.

Modelo 5

Supondo o modelo (6.7) da seção 5.3 temos as estatísticas e os valores do DIC e BIC, resultados do MCMC, programa do Apêndice C.6. Foram consideradas as seguintes distribuições a priori:

$$\alpha \sim N(0.001, 1000),$$

$$\beta \sim N(0.001, 1000),$$

$$\sigma^2 \sim IG(0.5, 0.01),$$

$$\lambda \sim N(0.001, 1000),$$

$$\gamma_1 \sim N(0.001, 1000)$$

e

$$\gamma_2 \sim N(0.001, 1000),$$

Foram geradas três cadeias com os seguintes valores iniciais:

Cadeia 1: ($\tau = 0.2, \lambda = 20, \alpha = 2, \beta = 0, \gamma_1 = 10, \gamma_2 = 2$),

Cadeia 2: ($\tau = 0.9, \lambda = 15, \alpha = 0.5, \beta = 3, \gamma_1 = 8, \gamma_2 = 9$)

e

Cadeia 3: ($\tau = 0.5, \lambda = 60, \alpha = 15, \beta = 2, \gamma_1 = 7, \gamma_2 = 6$).

Tabela 7.5 : Sumário a posteriori para os parâmetros do **Modelo 5**.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\hat{R}
α	-5,076	18,52	-36,26	-6,377	34,71	1,001
β	0,2149	0,02163	0,1727	0,2149	0,2572	1,001
δ	0,9188	0,07748	0,802	0,9322	0,9766	1,001
λ	2,668	0,8365	1,343	2,575	4,544	1,000
γ_1	-6,781	18,52	-46,39	-5,357	24,46	1,081
γ_2	4,263	18,5	-35,38	5,65	35,4	1,008
σ^2	27,03	4,744	18,47	26,78	37,02	1,000

DIC = 1069, 270; **BIC** = 1100, 93.

7.2 Comparação de modelos

Na simulação apresentada no capítulo 4 foram considerados os mesmos modelos (5.1) e (5.3), para os dados simulados. Os dados foram simulados em R, portanto os parâmetros envolvidos eram conhecidos previamente. A simulação do modelo (5.1) com o mesmo programa **C.2**, não está apresentada aqui mas pode ser repetida sob as mesmas condições. Portanto pôde ser feita uma comparação empírica entre a estimativa dos parâmetros pelos dois modelos e os seus valores previamente conhecidos. Observou-se que o modelo normal compromete a estimativa de σ . Quanto ao modelo assimétrico, as estimativas da média ocorreram muito próximo dos valores esperados para todos os parâmetros envolvidos, e as estimativas das medianas de cada parâmetro estão próximas dos valores das respectivas médias estimadas. Houve convergência das estatísticas de Gelman Rubin. Portanto, naquele caso é verificada por simulação a validade do modelo (5.3).

Com relação ao conjunto de dados reais que está sendo trabalhado neste capítulo, apresentamos primeiramente a comparação entre o **Modelo 1** e o **Modelo 2**. Para cada modelo foram geradas 130000 interações e foram descartadas as primeiras 10000 interações para fazer inferências com as distribuições a posteriori. Para evitar problemas de correlação nas cadeias geradas foi dado um salto de tamanho 10 em cada simulação. Nos dois casos os históricos das cadeias de Markov para os parâmetros indicam que não existem problemas de convergência nas cadeias geradas, e visualmente houve convergência das estatísticas de Gelman-Rubin. Os valores dos critérios de informação BIC e DIC favorecem o **Modelo 2** (veja tabela 7.6) e justifica a suposição de que os dados têm um comportamento assimétrico.

Tabela 7.6 : Comparação entre os modelos **1** e **2**.

Modelos	DIC	BIC
Modelo 1	1314,650	1327,47
Modelo 2	1230,880	1259,33

É importante explicar as diferenças no valor de σ^2 nestes dois modelos. Pelo **Modelo**

1 temos que $\sigma^2 = 38,91$ e é o valor da variância de $Bfat$ (variável resposta) por este modelo. Pelo **Modelo 2** temos que $\sigma^2 = 98,59$. A variância de $Bfat$ por este **Modelo 2** é dada por

$$Var(Bfat) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2\right).$$

(veja seção 4.1). Então, substituindo os valores das médias a posteriori dos parâmetros envolvidos temos que

$$Var(Bfat) \approx 35,86.$$

O **Modelo 3** detecta a assimetria e a bimodalidade dos dados. De fato o histograma e a densidade de $Bfat$ apresentam bimodalidade como pode ser visto na Figura 6.1. Por este modelo foram geradas 130000 iterações e foram descartadas as primeiras 10000 iterações para fazer inferências a posteriori. Para evitar problemas de correlação nas cadeias geradas foi dado um salto de tamanho 10 na simulação. Os históricos das cadeias de Markov para os parâmetros indicam que não existem problemas de convergência nas cadeias geradas, e visualmente houve convergência das estatísticas de Gelman-Rubin. Os valores dos critérios de informação DIC e BIC para este modelo praticamente se igualam aos respectivos valores para o **Modelo 2**, o que sugere que estes dois modelos explicam de forma semelhante a relação entre as variáveis. Este **Modelo 3**, que pertence a uma família que detecta assimetria e multimodalidade dos dados pode ser uma alternativa aos modelos de mistura, embora ainda pouco estudado. Nesta simulação foi considerada a independência entre os parâmetros, fato que é passível de discussão. Também há que se considerar que ainda não se tem expressão fechada para função geradora de momentos e portanto também para a média e a variância e este e outros problemas deixam a discussão deste modelo em aberto.

Quanto ao **Modelo 4** o respectivo valor do critério de seleção BIC é positivo mas valor do critério de seleção DIC é negativo. O critério DIC é bastante criticado como método de seleção de modelos complexos e parece estar sujeito à interferência do comportamento dos dados, como pode ser este o caso. Mas por outro lado este modelo apresenta também problemas de convergência nas cadeias geradas. Decidimos então descartá-lo da comparação, mas por ser este um modelo do que se esperaria ser mais preciso que os outros passa

a ser tratado como proposta futura no próximo capítulo.

Por último, dentre estes cinco modelos os critérios de seleção DIC e BIC favorecem o **Modelo 5** como o melhor modelo, justificando a suposição de que os dados além de terem um comportamento assimétrico, também estão sujeitos ao efeito do sexo de cada indivíduo (veja tabela 7.7). Por este modelo foram geradas 130000 interações e foram descartadas as primeiras 10000 interações para fazer inferências a posteriori. Para evitar problemas de correlação nas cadeias geradas foi dado um salto de tamanho 10 na simulação. Os históricos das cadeias de Markov para os parâmetros indicam que não existe problemas de convergência nas cadeias geradas (veja Figuras 7.3 e 7.4) e visualmente houve convergência das estatísticas de Gelman-Rubin, o que pode ser verificado na Figura 7.5. Comparado com o **Modelo 3**, este modelo detecta a presença de duas populações distintas no conjunto de dados, enquanto aquele modelo detecta apenas a bimodalidade, sem explicá-la. Neste caso tem-se uma expressão fechada para a variância e para as médias da variável resposta, ao contrário dos modelos 3 e 4. O parâmetro de escala da variável resposta por este **Modelo 5**, que é igual a 27,03 é menor se comparada com a escala da mesma variável resposta pelo **Modelo 2**. Isso se explica pelo fato de no **Modelo 2** ter sido considerada apenas uma população, enquanto que neste são consideradas duas populações distintas, com a mesma variância. A tabela a seguir apresenta os quatro modelos considerados na classificação e estes estão ordenados segundo os valores dos critérios DIC e BIC.

Tabela 7.7 : Comparação entre os modelos considerados.

Modelos	DIC	BIC
Modelo 5	1069,27	1100,93
Modelo 3	1230,59	1259,32
Modelo 2	1230,88	1259,33
Modelo 1	1314,65	1327,47

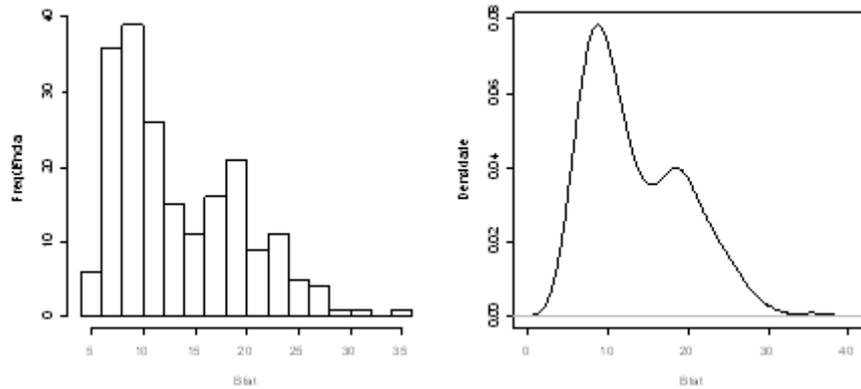


Figura 7.1 : Histograma e densidade estimada para a variável *Bfat*.

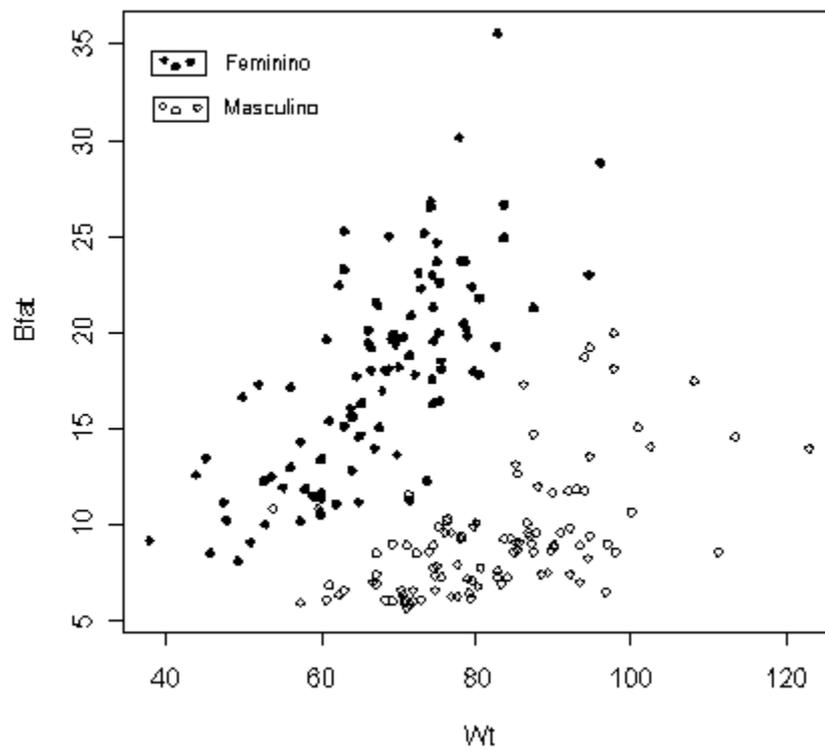


Figura 7.2 : Gráfico de dispersão para *Wt* vs *Bfat*.

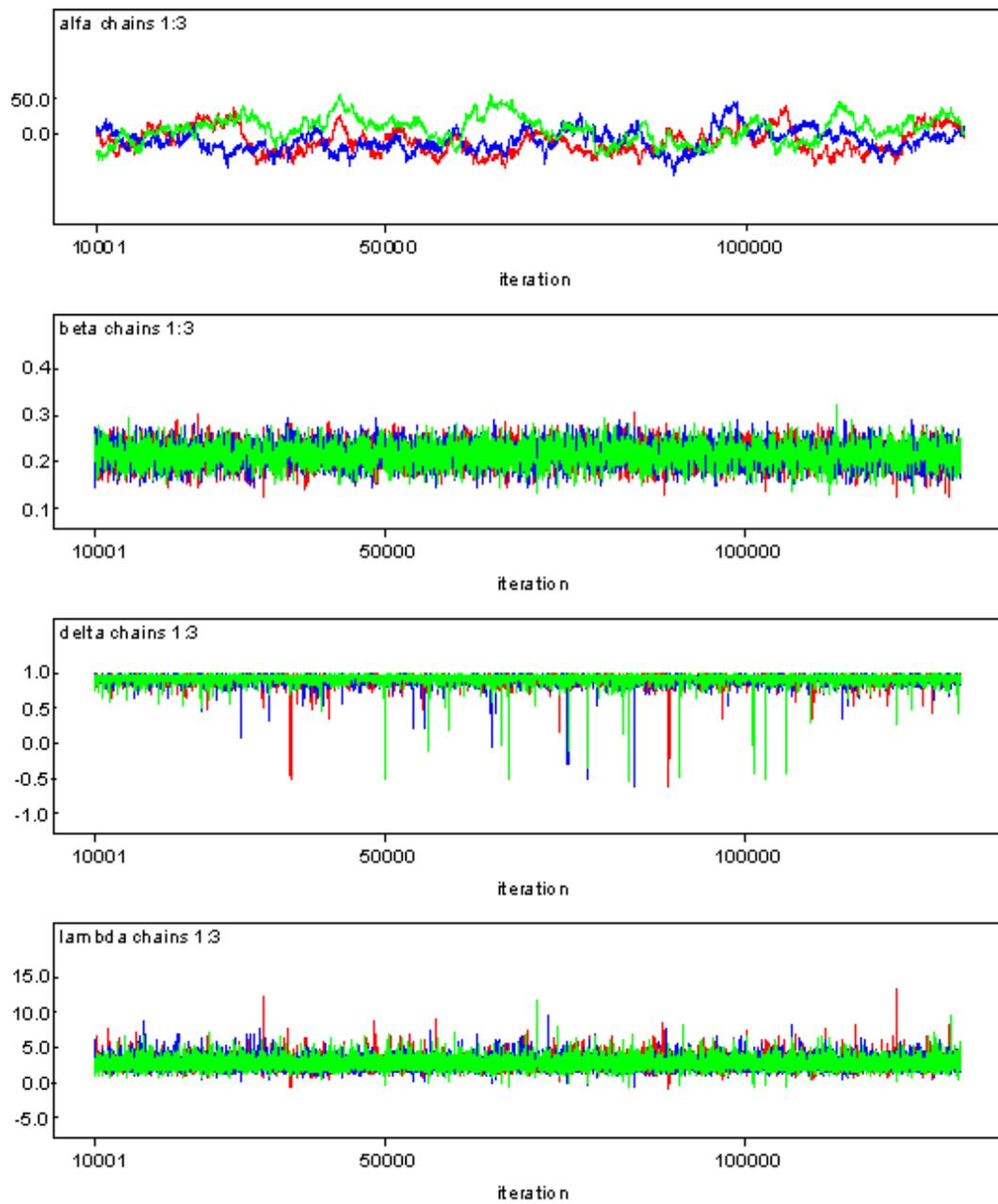


Figura 7.3 : Históricos das cadeias de Markov para α , β , δ e λ sob o modelo 5.

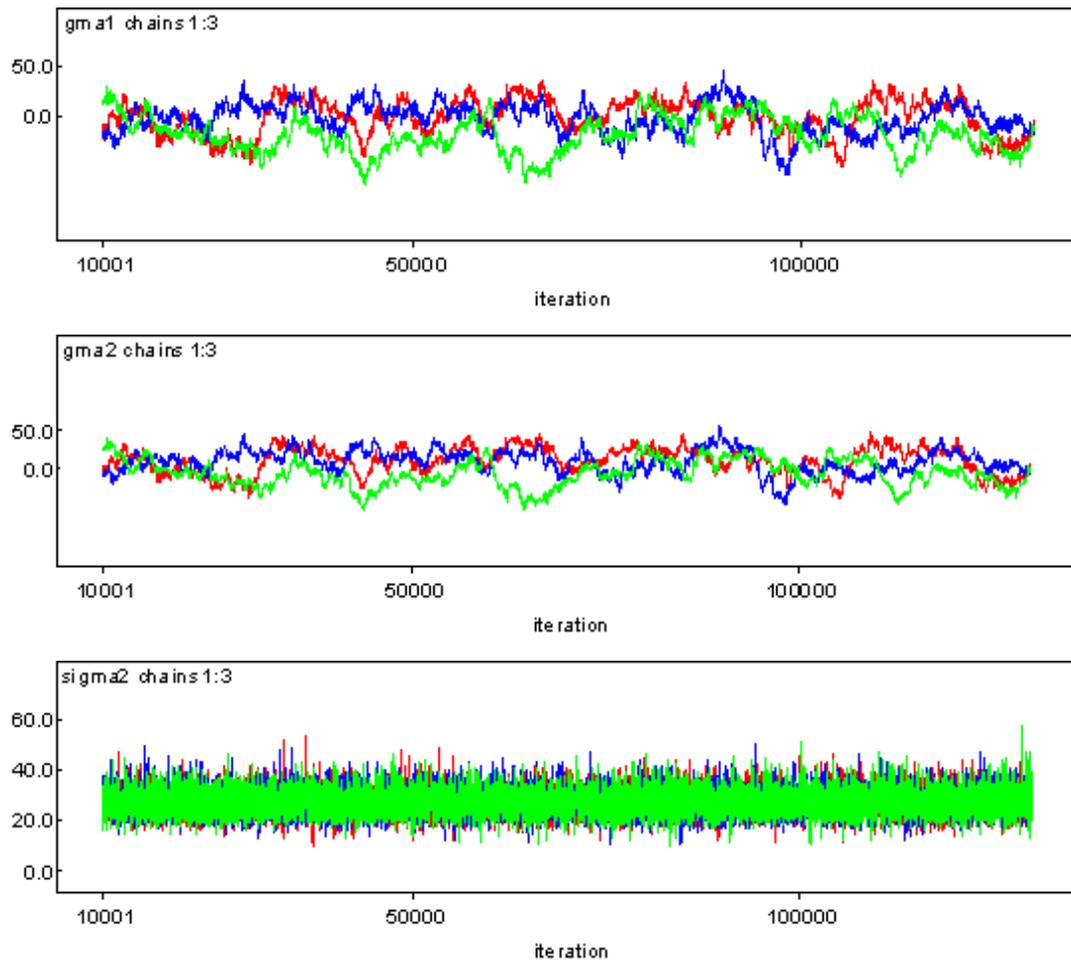


Figura 7.4 : Históricos das cadeias de Markov para γ_1 , γ_2 e σ^2 sob o modelo 5.

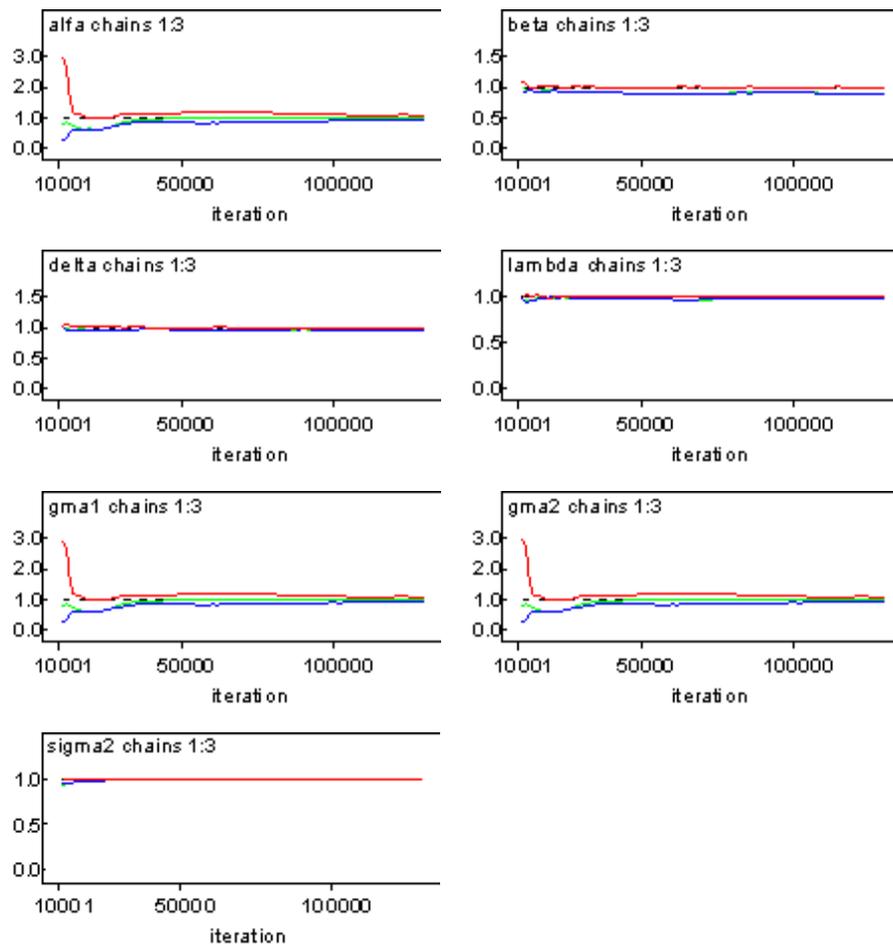


Figura 7.5 : Estatísticas de Gelman-Rubin para os parâmetros do modelo 5.

Capítulo 8

Considerações finais

Para o conjunto de dados que usamos no capítulo 6, na comparação dos modelos, os critérios de seleção DIC e BIC selecionam o **Modelo 5** (6.7) como o modelo que melhor descreve a relação entre as variáveis. Neste **Modelo 5** o valor do coeficiente de assimetria λ é pequeno se comparado ao do modelo **2**. Mas no **Modelo 3** os valores dos coeficientes β_1 e β_3 são também altos se comparados com o coeficiente de assimetria do **modelo 5**. Quanto ao valor da variância de $Bfat$ sob o **Modelo 5**, são consideradas neste caso duas populações com variâncias iguais. Para estimar esta variância usamos a expressão

$$Var(Bfat^*) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2 \right),$$

onde $Bfat^*$ é a notação usada para a população referente a cada um dos sexos. Substituindo os valores de σ^2 e δ^2 obtidos pelo MCMC temos que

$$Var(Bfat^*) \approx 12,50.$$

A diferença se explica pelo fato de que sob o **Modelo 2** e também sob os outros modelos é ignorada a presença de duas populações. Se é ignorada a presença de duas populações então o valor da variância é afetado pela locação de cada uma das populações. É o que acontece sob os modelos **1**, **2** e **4**. O **Modelo 3** embora seja um modelo de bastante interesse prático é pouco conhecido e ainda não se têm expressões para a sua função geradora de momentos, para sua média e sua variância. Enfim, os três modelos com

alguma assimetria considerados na seleção de modelos do capítulo 7 são propostas que devem ter as suas investigações continuadas. Quanto ao **modelo 4**, fazemos sobre este algumas considerações.

Devemos considerar que estes resultados não são conclusivos para o **modelo 4**. Este detecta o efeito aleatório de cada indivíduo, mas não detecta a presença de duas populações e apresenta problemas de convergência na geração das cadeias de Markov para o parâmetro σ_γ^2 . Se isso explica o valor negativo para o critério de seleção DIC é uma suposição que deve ser investigada. Foram atribuídos para σ_γ^2 os seguintes valores: $\sigma_\gamma^2 = 1$, $\sigma_\gamma^2 = 10$, $\sigma_\gamma^2 = 30$. Foram geradas cadeias de Markov para os demais parâmetros, com 20000 interações, com o descarte das 10000 primeiras interações. Os resultados para $\sigma_\gamma^2 = 1$ e $\sigma_\gamma^2 = 10$, são apresentados a seguir.

Tabela 8.1 : Sumários a posteriori para os parâmetros sob o **Modelo 4** supondo $\sigma_\gamma^2 = 1$.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\widehat{R}
α	13,3	2,54	8,039	13,29	17,94	1,011
β	0,002838	0,03329	-0,05872	0,002769	0,07208	1,005
δ	0,1609	0,008093	0,1452	0,1608	0,1772	1,001
λ	0,163	0,00842	0,1467	0,1629	0,18	1,001
σ_ϵ^2	37,93	3,953	30,86	37,67	46,45	1,001

DIC = 1314, 880; **BIC** = 1354, 62.

Note-se que para $\sigma_\gamma^2 = 1$ os resultados de α , β , σ e σ^2 estão conferido com os resultados dos mesmos parâmetros sob o **Modelo 1**, e isso é esperado para o **Modelo 4** quando $\sigma_\gamma^2 \rightarrow 0$ (veja densidade em (6.5) e (6.6)).

Tabela 8.2 : Sumários a posteriori para os parâmetros sob o

Modelo 4 supondo $\sigma_\gamma^2 = 10$.

Parâmetros	Média	D.P.	2.5%	mediana	97.5%	\hat{R}
α	13.45	2.394	8.927	13.31	18.36	1,020
β	8.637E-4	0.03144	-0.06426	0.002446	0.05942	1,015
δ	0.5095	0.02555	0.46	0.5095	0.56	0,999
λ	0.5933	0.04034	0.5181	0.5922	0.6758	0,999
σ_ϵ^2	28.81	3.931	21.9	28.52	37.26	1,001

DIC = 1314,610; **BIC** = 1353,62.

Os gráficos dos históricos das cadeias de Markov para os parâmetros em ambos os casos indicam que não há problemas de convergência e estão de acordo com os gráficos das estatísticas de Gelman-Rubin, que também mostram convergência. Os valores para o critério de seleção DIC são positivos e, na comparação com outros quatro modelos, levam à mesma classificação pelo critério de seleção BIC. Observe-se que nestes dois casos o valor da soma de $\sigma_\gamma^2 + \sigma_\epsilon^2$ coincide com o valor do parâmetro de escala sob o **Modelo 2**. Para $\sigma_\gamma^2 = 30$ há problemas de convergência.

Como proposta para resolver este problema de convergência, pretendemos utilizar em (6.5) a reparametrização: $\sigma_\gamma^2 = (\lambda\sigma_\epsilon)^2$ e utilizar as distribuições a priori não informativas propostas por Branco & Rodriguez (2005), descritas na seção 4.3.

Outras propostas de futuros trabalhos são listadas abaixo:

- O ajuste do **Modelo 4** para cada sexo separadamente.
- Considerar a variável *Bfat* truncada também à direita, dado que, por ser uma porcentagem deve ser menor ou igual a 100.
- Estimação dos parâmetros dos modelos através da representação de Henze (propriedade 15).

- Considerar, do mesmo conjunto de dados, outras variáveis explicativas para a variável $Bfat$, como é o caso das variáveis bmi (índice de massa corpórea), Ht (altura) e ssf (quantidade de dobras na barriga).

- Ajustar o conjunto de dados do apêndice **A1** ao **Modelo 2** e comparar com os resultados sob o **Modelo 1** apresentados na seção 5.2.1.

Referências Bibliográficas

- [1] Arellano-Valle, R. B., Gomez, H. and Iglesias, P. (2004). Shape mixtures of skewed distributions: a Bayesian analysis. Cimat, Guanajuato, Gto., México.
- [2] Arellano-Valle, R. B. and Azzalini, A. (2004). On the unification of families of skew-normal distributions. <http://azzalini.stat.unipd.it/SN/index.html>.
- [3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand J. Statist.* 12: 171-178.
- [4] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *J. R. Statist. Soc. B* 61, 579-602.
- [5] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on multivariate skew t distribution. *J. R. Statist. Soc. B* 65, 367-389.
- [6] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* 83, 715-726.
- [7] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. Università di Padova. NordStat 2004: Jyväskylä, 6-10 de junho de 2004.
- [8] Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). The new S language. Wadsworth & Brooks/Cole.
- [9] Branco, M. D. and Arellano-Valle, R. (2004). Distribuições elípticas assimétricas. 16º Simpósio Nacional de Probabilidade e Estatística. Associação Brasileira de Estatística. Brasil.

- [10] Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multiv. An.* 79, 99-113.
- [11] Branco, M. D. and Dey, D. K. (2002). Regression model under skew elliptical error distribution.
- [12] Branco, M. D. and Rodriguez, C.B. (2005). Bayesian inference for the skewness parameter on skew normal distribution. Relatório técnico, Universidade de São Paulo, São Paulo, Brasil.
- [13] Cartinhour, J. (1990). One dimensional marginal density function of a truncated multivariate normal density function. *Comm. Statist. - Theory & Methods* 19, 197-203.
- [14] Dávila, V. H. L. (2004). Modelos lineares mistos assimétricos. Tese de Doutorado, Universidade de São Paulo, Brasil.
- [15] Gelman, A., and Rubin, D. B. (1995). Avoiding model selection in Bayesian social research. *Sociological Methodology*, ed. P. V. Marsden, 165-188.
- [16] Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, 8, 519-540.
- [17] Gupta, A. K. (2003). A multivariate skew t -distributions. *Statistics*, 37(4), 359-363.
- [18] Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scand. J. Statist.* 13, 271-275.
- [19] Jeffreys, H. (1957). *Scientific Inference*, 2nd ed., Cambridge, England: Cambridge University Press.
- [20] Liseo, B. (1990). The skew-normal class of densities: inferencial aspects from a Bayesian viewpoint. *Statistica*, 50, 59-70.
- [21] Liseo, B. and Loperfido, N. (2004). Default Bayesian analysis of the skew-normal distribution. *J. Statist. Planning Infer.*, (pp. to appear).

- [22] Ma, Y. and Genton, M. G. (2004). Flexible class of skew-symmetric distributions. *Scad. J. Statist.*, 31, 459-468.
- [23] O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameters constraints. *Biometrika* 63, 201-202.
- [24] Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *Ann. Math. Statist.*, 27, 1075-1090.
- [25] Paulino, C. D., Turkman, M. A. A. e Murteira, B. (2003). Estatística Bayesiana. Fundação Calouste Gulbenkian. Lisboa, Portugal.
- [26] Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics* 27, 859-870.
- [27] Roberts, C. (1966). A correlation model useful in the study of twins. *J. Am. Statist. Assoc.*, 61, 1184-1190.
- [28] Rodriguez, C. L. B. (2005). Inferência bayesiana no modelo normal assimétrico. Dissertação de Mestrado, Universidade de São Paulo. São Paulo, Brasil.
- [29] Sahu, K., Dey, D. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canad. J. Statist.*, 31(2), 129-150.
- [30] Sartori, N. (2005). Bias prevention of maximum likelihood estimates: skew-normal and skew t -distributions. *J. Statist. Planning Infer.*, (pp. under revision).
- [31] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-466.
- [32] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Vanderlinde, A. (2002). Bayesian measures and fit. *Journal Royal Statistic Societ. B.* 64, 583-640.
- [33] Wang, J., Boyer, J. and Genton, M. G. (2004). A skew-symmetric representation of multivariate distribution. *Statistica Sinica*, 14, 1259-1270.
- [34] Zacks, S. (1981). Parametric statistical inference. Oxford: Pergamon Press.

-
- [35] Zellner, A. (1971). An introduction to Bayesian inference in econometrics. John Wiley & Sons, INC. New York.
- [36] 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.

Apêndice A

Conjuntos de dados

São apresentados neste apêndice os dois conjuntos de dados usados nesta dissertação.

A.1 Renda Pessoal e investimento pessoal (Zelner[35])

Nesta tabela estão listados os valores das variáveis do conjunto de dados considerado na subseção 5.2.1. Os valores da renda pessoal (Y) e do investimento pessoal (X) são dados em dólar.

Indivíduo	X	Y	Indivíduo	X	Y
1	39	433	11	22	372
2	60	483	12	17	381
3	42	479	13	27	419
4	52	486	14	33	449
5	47	494	15	48	511
6	51	498	16	51	520
7	45	511	17	33	477
8	60	534	18	46	517
9	39	478	19	54	548
10	41	440	20	100	629

A.2 Dados usados na aplicação do capítulo 7

Este conjunto de dados é dado na forma de vetor que é o formato usados nos aplicativos **R** e **BUGS**. Os valores de Wt são dados em kilogramas. A ordem dos dados está dada segundo a ordem de cada indivíduo. Os primeiros 100 indivíduos são do sexo feminino e os 202 restantes do sexo masculino.

Bfat = c(19.75,
21.3, 19.88, 23.66, 17.64, 15.58, 19.99, 22.43, 17.95, 15.07,
28.83, 18.08, 23.3, 17.71, 18.77, 19.83, 25.16, 18.04, 21.79,
22.25, 16.25, 16.38, 19.35, 19.2, 17.89, 12.2, 23.7, 24.69, 16.58,
21.47, 20.12, 17.51, 23.7, 22.39, 20.43, 11.29, 25.26, 19.39,
19.63, 23.11, 16.86, 21.32, 26.57, 17.93, 24.97, 22.62, 15.01,
18.14, 26.78, 17.22, 26.5, 23.01, 30.1, 13.93, 26.65, 35.52,
15.59, 19.61, 14.52, 11.47, 17.71, 18.48, 11.22, 13.61, 12.78,
11.85, 13.35, 11.77, 11.07, 21.3, 20.1, 24.88, 19.26, 19.51,
23.01, 8.07, 11.05, 12.39, 15.95, 9.91, 16.2, 9.02, 14.26, 10.48,
11.64, 12.16, 10.53, 10.15, 10.74, 20.86, 19.64, 17.07, 15.31,
11.07, 12.92, 8.45, 10.16, 12.55, 9.1, 13.46, 8.47, 7.68, 6.16,
8.56, 6.86, 9.4, 9.17, 8.54, 9.2, 11.72, 8.44, 7.19, 6.46, 9,
12.61, 9.03, 6.96, 10.05, 9.56, 9.36, 10.81, 8.61, 9.53, 7.42,
9.79, 8.97, 7.49, 11.95, 7.35, 7.16, 8.77, 9.56, 14.53, 8.51,
10.64, 7.06, 8.87, 7.88, 9.2, 7.19, 6.06, 5.63, 6.59, 9.5, 13.97,
11.66, 6.43, 6.99, 6, 6.56, 6.03, 6.33, 6.82, 6.2, 5.93, 5.8,
6.56, 6.76, 7.22, 8.51, 7.72, 19.94, 13.91, 6.1, 7.52, 9.56,
6.06, 7.35, 6, 6.92, 6.33, 5.9, 8.84, 8.94, 6.53, 9.4, 8.18,
17.41, 18.08, 9.86, 7.29, 18.72, 10.12, 19.17, 17.24, 9.89, 13.06,
8.84, 8.87, 14.69, 8.64, 14.98, 7.82, 8.97, 11.63, 13.49, 10.25,
11.79, 10.05, 8.51, 11.5, 6.26),

Wt = c(78.9,
74.4, 69.1, 74.9, 64.6, 63.7, 75.2, 62.3, 66.5, 62.9, 96.3, 75.5,
63, 80.5, 71.3, 70.5, 73.2, 68.7, 80.5, 72.9, 74.5, 75.4, 69.5,
66.4, 79.7, 73.6, 78.7, 75, 49.8, 67.2, 66, 74.3, 78.1, 79.5,
78.5, 59.9, 63, 66.3, 60.7, 72.9, 67.9, 67.5, 74.1, 68.2, 68.8,
75.3, 67.4, 70, 74, 51.9, 74.1, 74.3, 77.8, 66.9, 83.8, 82.9,
64.1, 68.85, 64.8, 59, 72.1, 75.6, 71.4, 69.7, 63.9, 55.1, 60,
58, 64.7, 87.5, 78.9, 83.9, 82.8, 74.4, 94.8, 49.2, 61.9, 53.6,
63.7, 52.8, 65.2, 50.9, 57.3, 60, 60.1, 52.5, 59.7, 57.3, 59.6,
71.5, 69.7, 56.1, 61.1, 47.4, 56, 45.8, 47.8, 43.8, 37.8, 45.1,
67, 74.4, 79.3, 87.5, 83.5, 78, 78, 85, 84.7, 92, 72.3, 83, 96.9,
85.7, 85.4, 85.3, 93.5, 86.8, 87.9, 87.2, 53.8, 89.8, 91.1, 88.6,
92.3, 97, 89.5, 88.2, 92.2, 78.9, 90.3, 87, 113.7, 98, 100.2,
79.4, 90.3, 77.7, 83.9, 75.5, 60.6, 71, 71.8, 76.8, 102.7, 94.25,
79, 66.6, 71.8, 74.8, 68.2, 62.3, 61, 77.5, 57.4, 71.4, 70.3,
80.2, 84.2, 111.3, 80.7, 97.9, 123.2, 72.9, 83, 75.9, 70.7, 67.1,
69.2, 67.05, 70.5, 70.8, 71, 69.1, 62.9, 94.8, 94.6, 108.2, 97.9,
75.2, 74.8, 94.2, 76.1, 94.7, 86.2, 79.6, 85.3, 74.4, 93.5, 87.6,
85.4, 101, 74.9, 87.3, 90, 94.7, 76.3, 93.2, 80, 73.8, 71.1,
76.7))

Apêndice B

Programas em R

B.1 Programa para geração do gráfico de uma distribuição $SN(\lambda)$ dada por (2.3)

```
z<-seq(-4,4,length=80)
lam<-c(-5,0,5)
y<-matrix(NA,80,3)
for (i in 1:3)
y[,i]<-2*dnorm(z,0,1)*pnorm((lam[i]*z),0,1)
matplot(z,y,main="f(z)",type="l",xlab="z",ylab="f(z)",col=1,ylim=c(0,0.8))
legend(-3.8,0.6,paste("lambda=",lam),lty=1:3,cex=0.75)
```

B.2 Geração de amostra de regressão com erro normal assimétrico

```
set.seed(1050)
n <- 1000
m <- 10 # número de réplicas de Y
alpha <- 5
b <- 0.6
```

```
sigma <- 3
lambda <- -2
delta <- lambda/sqrt(1+lambda**2)
mu <- delta*sigma*sqrt(2/pi)
a <- alpha-mu
X <- matrix(sort(rep(1:m,trunc(n/m))),ncol=1)
U <- abs(rnorm(n,0,1))
V <- rnorm(n,0,1)
Z <- numeric()
E <- numeric()
Y <- numeric()

for(i in 1:n) {
  Z[i] <- delta*(U[i])+sqrt(1-delta**2)*V[i]
  E[i] <- Z[i]-delta*sqrt(2/pi)
  Y[i] <- alpha + b*X[i] + sigma*E[i]}

par(mfrow=c(2,2))
plot(density(Y),main="Densidade de Y",xlab="Y",ylab="Densidade")
hist(Y,main="Histograma de Y",xlab="Y",ylab="Frequência",col="green3")
plot(density(E),main="Densidade do Erro",xlab="Erro",ylab="Densidade")
hist(E,main="Histograma do Erro",xlab="Erro",ylab="Frequência",col="Blue")

  Y8 <- alpha + b*X
  Y7 <- a + b*X

points(X,Y,pch=20)
x11()
boxplot(Y~X,col="lavender",notch=TRUE,xlab="X",ylab="Y",main="BoxPLOT")
lines(X,Y7,type='l',lwd=2,col="red")
lines(X,Y8,type='l',lwd=2,col="blue").
```

Apêndice C

Programas em WinBUGS

C.1 Programa 1 para o Modelo 1

```
model
{
for (i in 1:N) {
Y[i] ~ dnorm(mu[i],tau)
mu[i] <- alfa+beta*x[i]
}
alfa ~ dnorm(350,0.001)
beta ~ dnorm(0.001,0.001)
tau ~ dgamma(0.01,0.01)
sigma2 <- 1/tau
sigma <- sqrt(sigma2)
}
```

Dados

```
list(N=20,Y=c( 433,483,479,486,494,498,511,534,478,440,
372,381,419,449,511,520,477,517,548,629),
x=c(39,60,42,52,47,51,45,60,39,41,22,
17,27,33,48,51,33,46,54,100))
```

Valores Iniciais

```
list(tau=1,alfa=100,beta=0)
```

```
list(tau=2,alfa=450,beta=3)
```

```
list(tau=4,alfa=40,beta=2)
```

C.2 Programa 2 para o Modelo 1**Model**

```
{
for (i in 1:N) {
uns[i] ~ dbern(p[i])
p[i] <- (2/sigma)*(1/(sqrt(6.28)))*exp(-0.5*pow(feio[i],2))*phi(lambda*feio[i])
feio[i] <- (e[i]-mu)/sigma
e[i] <- y[i]-(alfa+beta*x[i])
}
alfa ~ dnorm(0.001,0.001)
beta ~ dnorm(0.001,0.001)
tau ~ dgamma(0.01,0.01)
sigma2 <- 1/tau
sigma <- sqrt(sigma2)
delta <- lambda/sqrt(1+pow(lambda,2))
mu <- sigma*delta*0.7978846
lambda <- 0
}
```

Dados

```
list(N=1000,y=c( dados da variável Y),x=c( dados da variável X),uns=c(1,1,...,1,1))
```

Valores iniciais

```
list(tau=1,lambda=2,alfa=0,beta=0)
```

```
list(tau=2,lambda=-1,alfa=0.5,beta=3)
```

```
list(tau=4,lambda=1,alfa=1,beta=2)
```

C.3 Programa para o Modelo 2

Model

```
{  
for (i in 1:N) {  
uns[i] ~ dbern(p[i])  
p[i] <- (2/sigma)*(1/(sqrt(6.28)))*exp(-0.5*pow(feio[i],2))*phi(lambda*feio[i])  
feio[i] <- (e[i]-mu)/sigma  
e[i] <- y[i]-(alfa+beta*x[i])  
}  
alfa ~ dnorm(0.001,0.001)  
beta ~ dnorm(0.001,0.001)  
tau ~ dgamma(0.01,0.01)  
sigma2 <- 1/tau  
sigma <- sqrt(sigma2)  
delta <- lambda/sqrt(1+pow(lambda,2))  
mu <- sigma*delta*0.7978846  
lambda ~ dnorm(0.001,0.001)  
}
```

Dados

```
list(N=1000,y=c( dados da variável Y),  
x=c( dados da variável X),uns=c(1,1,...,1,1,1))
```

Valores iniciais

```
list(tau=1,alfa=0,beta=0)  
list(tau=2,alfa=0.5,beta=3)  
list(tau=4,alfa=1,beta=2).
```

C.4 Programa para o Modelo 3

```

model
{
for (i in 1:N) {
uns[i] ~ dbern(p[i])
p[i] <- (2/sigma)*(1/(sqrt(6.28)))*exp(-0.5*pow(feio[i],2))*
phi(lambda*feio[i]+eta1*pow(feio[i],3))
feio[i] <- (e[i]-mu)/sigma
e[i] <- y[i]-(alfa+beta*x[i])
}
alfa ~ dnorm(0.001,0.001)
beta ~ dnorm(0.001,0.001)
tau ~ dgamma(0.5,0.1)
sigma2 <- 1/tau
sigma <- sqrt(sigma2)
mu ~ dnorm(0.001,0.001)
lambda ~ dnorm(0.001,0.001)
eta1 ~ dnorm(0.001,0.001)
}

```

Dados

```

list(N=1000,y=c( dados da variável Y),
x=c( dados da variável X),uns=c(1,1,...,1,1,1))

```

Valores Iniciais

```

list(tau=0.2,lambda=20,alfa=2,beta=0,eta1=1,mu=5)
list(tau=0.9,lambda=15,alfa=0.5,beta=3,eta1=3,mu=7)
list(tau=0.5,lambda=60,alfa=15,beta=2,eta1=4,mu=6)

```

C.5 Programa Para o Modelo 4

```

model
{
for (i in 1:N) {
uns[i] ~ dbern(p[i])
p[i] <- (1/sqrt(sigma2+sigma2g))*(1/(sqrt(6.28)))*exp(-0.5*pow(feio[i],2))*
phi(tau1[i]*sqrt(sigma2+sigma2g)/sigma*sigmag+
(sigma2g/sigma)*feio[i])/phi(tau1[i]/sigmag)
feio[i] <- (e[i])/sqrt(sigma2+sigma2g)
e[i] <- y[i]-(alfa+beta*x[i])
tau1[i] <- alfa+beta*x[i]
}
alfa ~ dnorm(0,0.001)
beta ~ dnorm(0.001,0.001)
tau ~ dgamma(0.01,0.01)
sigma2 <- 1/tau
sigma <- sqrt(sigma2)
tau2 ~ dgamma(0.5,0.1)
sigma2g <- 1/tau2
sigmag <- sqrt(sigma2g)
delta <- sigmag/sqrt(sigma2g+sigma2)
lambda <- sigmag/sigma
}

```

Dados

```

list(N=1000,y=c( dados da variável Y),
x=c( dados da variável X),uns=c(1,1,...,1,1,1))

```

Valores Iniciais

```

list(tau=0.2,alfa=10,beta=0,tau2=0.1)
list(tau=0.9,alfa=12,beta=3,tau2=0.2)
list(tau=0.5,alfa=15,beta=0.1,tau2=0.7)

```

C.6 Programa para o Modelo 5

```

model
{
for (i in 1:N) {
uns[i] ~ dbern(p[i])
p[i] <- (2/sigma)*(1/(sqrt(6.28)))*exp(-0.5*pow(feio[i],2))*phi(lambda*feio[i])
feio[i] <- (e[i])/sigma
e[i] <- y[i]-(alfa+beta*x[i]+gma1*Z1[i]+gma2*Z2[i])
}
alfa ~ dnorm(0.001,0.001)
beta ~ dnorm(0.001,0.001)
tau ~ dgamma(0.5,0.01)
sigma2 <- 1/tau
sigma <- sqrt(sigma2)
delta <- lambda/sqrt(1+pow(lambda,2))
lambda ~ dnorm(0.001,0.001)
gma1 ~ dnorm(0.001,0.001)
gma2 ~ dnorm(0.001,0.001)
}

```

Dados

```

list(N=1000,y=c( dados da variável Y),
x=c( dados da variável X),Z1=c(dados da variável Z1),
Z2=c(dados da variável Z2),uns=c(1,1,...,1,1,1))

```

Valores Iniciais

```

list(tau=0.2,lambda=20,alfa=2,beta=0,gma1=10,gma2=2)
list(tau=0.9,lambda=15,alfa=0.5,beta=3,gma1=8,gma2=9)
list(tau=0.5,lambda=60,alfa=15,beta=2,gma1=7,gma2=6)

```

Apêndice D

Critérios de seleção de modelos

Após a obtenção das estimativas dos parâmetros sob os modelos, através das técnicas de MCMC, é importante que se consiga uma medida que possibilite fazer a escolha do melhor modelo. A escolha pode ser feita através de um critério de seleção de modelos. Neste trabalho são usados dois critérios de seleção: **DIC** (*deviance criterion information*) e **BIC** (*Bayesian information criterion*).

Spieghalter et al. [32] desenvolveram o critério de seleção **DIC**, definido por

$$\mathbf{DIC} = -2 \log \mathbf{p}(\mathbf{y}|\boldsymbol{\theta}) + \mathbf{p}_D,$$

onde $\mathbf{p}(\mathbf{y}|\boldsymbol{\theta})$ é a função de verossimilhança do modelo considerado. A parcela \mathbf{p}_D é definida por $\mathbf{p}_D = \bar{D} - D(\bar{\theta})$, onde $D(\theta) = -2 \log \mathbf{p}(\mathbf{y}|\boldsymbol{\theta})$ é a função *deviance*, $\bar{\theta} = E(\theta|y)$ e $\bar{D} = E[D(\theta)|y]$.

O critério de seleção **BIC** [31] é definido por

$$\mathbf{BIC} = -2 \log \mathbf{p}(\mathbf{y}|\boldsymbol{\theta}) + \mathbf{d} \log \mathbf{N},$$

onde $\mathbf{p}(\mathbf{y}|\boldsymbol{\theta})$ é a função de verossimilhança do modelo considerado, \mathbf{d} é a dimensão do vetor de parâmetros e \mathbf{N} o tamanho da amostra.

Como critério de seleção escolhe-se o modelo que apresentar o menor valor para o **DIC** e o menor valor para o **BIC**.

Apêndice E

Lista de distribuições

Tabela E.1: Expressões das funções de densidade de probabilidade a priori que foram utilizadas nesta dissertação

notação p/ priori	função densidade de probabilidades	$E(\theta)$	$V(\theta)$
$\pi(\theta) \sim \text{Bernoulli}(p)$	$\pi(\theta) = p^\theta (1-p)^{1-\theta}$ $\theta \in \{0, 1\}$	p	$p(1-p)$
$\pi(\theta) \sim \text{Uniforme}(a, b)$	$\pi(\theta) = \frac{1}{b-a}$ $a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\pi(\theta) \sim \text{Gama}(a, b)$	$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta)$ $a > 0, b > 0, \theta > 0$	$\frac{a}{b}$	$\frac{a}{b^2}$
$\pi(\theta) \sim \text{Inv-Gama}(a, b)$	$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} \exp(-b/\theta)$ $a > 0, b > 0, \theta > 0$	$\frac{b}{a-1}$	$\frac{b^2}{(a-1)^2(a-2)},$ $a > 2$
$\pi(\theta) \sim N(\mu, \sigma^2)$	$\pi(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\theta - \mu)^2\right]$ $\sigma > 0$	μ	σ^2
$\pi(\theta) \sim t(\mu, \sigma, v)$	$\pi(\theta) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sigma\sqrt{v\pi}} \left(1 + \frac{1}{v} \left(\frac{\theta-\mu}{\sigma}\right)^2\right)^{-\frac{v+1}{2}}$ $\sigma > 0$	μ	$\frac{v}{v-2}\sigma^2,$ $v > 2$