

Calibração linear assimétrica

Cléber da Costa Figueiredo

TESE APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA

OBTENÇÃO DO GRAU

DE

DOUTOR EM CIÊNCIAS

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Heleno Bolfarine**

Co-Orientadora: **Prof^a . Dr^a . Mônica Carneiro Sandoval**

*Durante a elaboração deste trabalho o autor
recebeu apoio financeiro da CAPES e do CNPq*

São Paulo, março de 2009.

Calibração linear assimétrica

Este exemplar corresponde à redação
final da tese devidamente corrigida
e defendida por Cléber da Costa Figueiredo
e aprovada pela comissão julgadora.

São Paulo, março de 2009.

Banca examinadora:

- Titulares:

- Prof. Dr. Heleno Bolfarine (orientador) - IME-USP;
- Prof^ª . Dr^ª . Mônica Carneiro Sandoval (co-orientadora) - IME-USP;
- Prof. Dr. Héctor Waldimir Gómez Geraldo - Universidad Concepción - Chile;
- Prof. Dr. Filidor Edilfonso Vilca Labra - IMECC-UNICAMP;
- Prof^ª . Dr^ª . Claudia Regina Oliveira de Paiva Lima - DE-UFPE.

- Suplentes:

- Prof^ª . Dr^ª . Márcia D'Elia Branco - IME-USP;
- Prof. Dr. Gilberto Alvarenga Paula - IME-USP;
- Prof. Dr. Jorge Luis Bazán Guzmán - PUC-Peru;
- Prof. Dr. Víctor Hugo Lachos Dávila - IMECC-UNICAMP;
- Prof. Dr. Edwin Moises Marcos Ortega - ESALQ-USP.

*À minha mãe, à minha esposa,
aos meus familiares e
aos meus amigos.*

Agradecimentos

Quero neste agradecimento ser muito mais objetivo do que fui no passado. É claro que durante quatro anos e meio de doutoramento muita gente passa pela vida da gente, deixa um rastro e precisa ser lembrada. Outras passam pela vida da gente e se esquecem de deixar uma marca no caminho... ainda há aquelas que já eram muito importantes e continuam sendo; para não ser muito extenso, Figueiredo (2004) trás uma série de referências a essas pessoas. Puxa! Você deve estar a se perguntar: até nos agradecimentos há referência!? É claro! Isso aprendi nesse tempo todo!! Poupa-nos um grande esforço!

Alguns outros agradecimentos precisam ser feitos e serão, talvez redundantes, caso você tenha consultado a referência supracitada. Portanto, quero agradecer ou (re)agradecer...

Ao meu orientador, Prof. Dr. Heleno Bolfarine, que aceitou o convite de orientação, quando precisei de sua ajuda para poder continuar com a coorientação da Prof^ª . Dr^ª . Mônica Sandoval, que na época estava impossibilitada de me orientar por questões burocráticas e administrativas do IME-USP. À Prof^ª . Dr^ª . Mônica Carneiro Sandoval que me orienta desde a iniciação científica, quando procurei a Prof^ª . Dr^ª . Denise Botter (outra figuraça!) para ser orientado, ainda na graduação.

À minha família pela tolerância à minha personalidade bastante divergente às idéias deles. Em especial, à minha mãe, aos meus irmãos, à minha cunhada, à

minha sobrinha e à minha sogra. Ainda, à minha recente esposa Alleid, que foi minha namorada e noiva durante esse período todo de estudo e que me ajudou muito a ter vontade de acabar com o doutorado o mais rápido possível!

Aos colegas e amigos que conheci na USP e que ainda são um “calo no meu sapato”... Márcio, Salomé, Sandra, Ana Paula, Lázaro, Rogério, Amanda, Jefferson, Thiago, Frederico, Livia, Tatiana, Gisela, Betsabé, João, Miriam e Maria Kelly. Ainda aos amigos Marcos e Daniela, com os quais estudei muito para obter os créditos iniciais e à Mayra por ser minha dupla em todas as matérias que fizemos juntos e ter me ajudado muitíssimo para o ingresso na carreira profissional. Mayra, muito obrigado por tudo!

Termino, agradecendo à banca que gentilmente aceitou o convite de comparecer à minha defesa em plena sexta-feira da semana de carnaval e aos comentários valiosos que foram incorporados ao texto. Ao CNPq e à CAPES pelo suporte financeiro. A Deus por Sua ajuda incondicional. Ao CEPEUSP pela excelente piscina.

Resumo

A presente tese aborda aspectos teóricos e aplicados da estimação dos parâmetros do modelo de calibração linear com erros distribuídos conforme a distribuição normal-assimétrica (Azzalini, 1985) e t-normal-assimétrica (Gómez, Venegas e Bolfarine, 2007). Aplicando um modelo assimétrico, não é necessário transformar as variáveis a fim de obter erros simétricos. A estimação dos parâmetros e das variâncias dos estimadores do modelo de calibração foram estudadas através da visão freqüentista e bayesiana, desenvolvendo algoritmos tipo EM e amostradores de Gibbs, respectivamente. Um dos pontos relevantes do trabalho, na ótica freqüentista, é a apresentação de uma reparametrização para evitar a singularidade da matriz de informação de Fisher sob o modelo de calibração normal-assimétrico na vizinhança de $\lambda = 0$. Outro interessante aspecto é que a reparametrização não modifica o parâmetro de interesse. Já na ótica bayesiana, o ponto forte do trabalho está no desenvolvimento de medidas para verificar a qualidade do ajuste e que levam em consideração a assimetria do conjunto de dados. São propostas duas medidas para medir a qualidade do ajuste: o ADIC (*Asymmetric Deviance Information Criterion*) e o EDIC (*Evident Deviance Information Criterion*), que são extensões da idéia de Spiegelhalter *et al.* (2002) que propôs o DIC ordinário que só deve ser usado em modelos simétricos.

Palavras-chave: distribuições normal e t-normal assimétricas; singularidade da matriz de informação de Fisher; critérios de informação; algoritmo EM; amostrador de Gibbs.

Abstract

This thesis focuses on theoretical and applied estimation aspects of the linear calibration model with skew-normal (Azzalini, 1985) and skew-t-normal (Gómez, Venegas e Bolfarine, 2007) error distributions. Applying the asymmetrical distributed error methodology, it is not necessary to transform the variables in order to have symmetrical errors. The frequentist and the Bayesian solution are presented. The parameter estimation and its variance estimation were studied using the EM algorithm and the Gibbs sampler, respectively, in each approach. The main point, in the frequentist approach, is the presentation of a new parameterization to avoid singularity of the information matrix under the skew-normal calibration model in a neighborhood of $\lambda = 0$. Another interesting aspect is that the reparameterization developed to make the information matrix nonsingular, when the skewness parameter is near to zero, leaves the parameter of interest unchanged. The main point, in the Bayesian framework, is the presentation of two measures of goodness-of-fit: ADIC (*Asymmetric Deviance Information Criterion*) and EDIC (*Evident Deviance Information Criterion*). They are natural extensions of the ordinary DIC developed by Spiegelhalter *et al.* (2002).

Keywords: skew-normal and skew-t-normal distributions; singularity of the information matrix; information criteria; EM algorithm; Gibbs sampler.

Prefácio

Embora prefácios sejam dispensáveis em teses de doutoramento, acreditamos que algumas preliminares sejam naturalmente relevantes devido à forma como conduzimos o presente trabalho. O pontapé inicial foi estudar modelos de calibração linear para análise de dados com distribuições caracterizadas pela não simetria, onde propusemos considerar aspectos teóricos e aplicações a conjuntos de dados reais. Nesse âmbito é que desenrolamos nosso projeto de pesquisa no final de 2004 e início de 2005, aproveitando uma parceria existente entre o Instituto de Matemática e Estatística da USP e o Departamento de Estatística da UFPE, parceria essa suportada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico, através do Fundo Setorial de Infra-Estrutura (CT-INFRA) desse órgão, que visava apoiar a pós-graduação desse último departamento. O projeto havia sido batizado por PAPER (muito oportuno, digamos!), que são as iniciais de: Projeto de Apoio à Pós-graduação em Estatística no Recife e que teria a duração de dois anos, mais precisamente, de junho de 2004 a junho de 2006.

A efervescência do momento era estudar a distribuição normal-assimétrica que havia entrado no cenário internacional estatístico através de Azzalini (1985), mas havia experimentado um momento de latência de aproximadamente quinze anos até explodir com inúmeros trabalhos, servindo como base para diversas teses, dissertações e artigos tanto em reuniões científicas como em revistas de prestígio internacional e

tem sido o palco de diversas discussões, conferências, sessões temáticas, seminários entre outros, valendo ressaltar que o estudo da assimetria e, já podemos dizer que também o estudo da curtose, ou de distribuições que possuem caudas mais pesadas que a própria normal-assimétrica, culminou com o prêmio IASI em 2008 oferecido ao jovem doutor Lachos, V. H. pelos inúmeros estudos realizados nesse campo em tão pouco tempo.

Nossa idéia inicial era propor um modelo de calibração onde a distribuição dos erros do modelo seria assimétrica (normal-assimétrica estritamente), ou seja, propusemos:

1. Apresentar os principais resultados teóricos do modelos de calibração linear, considerando que o vetor de erros seria normal-assimétrico;
2. Desenvolver alguma rotina computacional para a estimação dos parâmetros do modelo. Uma idéia inicial poderia ser encontrar os estimadores de máxima verossimilhança da forma usual, desenvolvendo a matriz de segundas derivadas para estimar variâncias através de métodos iterativos, ou ainda, encontrar os estimadores de máxima verossimilhança através do método EM. Dessa necessidade de maximizar a verossimilhança para encontrar um estimador para os parâmetros do modelo, surgiria a necessidade da utilização de programas como o Ox ou a linguagem R para essa tarefa que parecia não ser simples;
3. Estender os resultados para a abordagem bayesiana; e
4. Aplicar os modelos estudados a conjunto de dados reais e verificar a adequabilidade do modelo.

O projeto foi escrito dada a experiência dos orientadores (Bolfarine, H. e Sandoval, M. C.) em calibração linear, além de problemas de calibração estatística serem freqüentes em centros de estatística aplicada e em diversas áreas do conhecimento,

sendo que as suposições de normalidade ou de outras distribuições simétricas, tais como distribuições elípticas simétricas podiam não ser adequadas a conjunto de dados de notável assimetria. O arquivo de dados que havia sido analisado até então era o clássico conjunto de dados de 102 homens e 100 mulheres atletas selecionados pelo *Australian Institute of Sport*¹, onde um simples histograma poderia mostrar claramente a assimetria natural do conjunto e diversos trabalhos versaram sobre esse conjunto de dados. Aqui é necessário fazer um parêntese para atribuir ao menos uma nota de rodapé aos autores do conjunto de dados, embora muitos acreditem ser de autoria de Azzalini, A.. O presente trabalho pretende preencher as lacunas que ainda restam nesse imenso todo que compreende a pesquisa. O que veremos já no índice é que a proposta inicial foi vastamente explorada e, é claro, quatro anos é bastante tempo para que a pesquisa tome outros rumos e outros interesses sejam incorporados ao estudo. Dessa forma, acreditamos que o prefácio já se tornou indispensável e era esse nosso objetivo; torná-lo útil. Durante a qualificação do autor foi sugerida a incorporação do estudo do modelo de calibração que utilizava a distribuição t-normal-assimétrica pela banca composta pelos professores doutores Bolfarine H., Sandoval, M. C. e Lima, C. R. O. P., além de uma série de outros palpites terem feito parte do corpo do texto, em especial, as sugestões de alguns *referees* do *Journal of Applied Statistics* e do *Journal of Chemometrics* que são hospedeiros do Capítulo 1 e do Capítulo 2, respectivamente, desta tese. Esperamos que o leitor possa aproveitar o texto e, se encontrar incongruências, atribua-as ao autor.

São Paulo, 27 de fevereiro de 2009.

Cléber da Costa Figueiredo

¹TELFORD, R. e CUNNINGHAM, R. *apud* AZZALINI, A. Disponível em: << <http://pbil.univ-lyon1.fr/library/sn/html/ais.html> >>. Acesso em: 15 setembro 2006.

Índice

Agradecimentos	v
Resumo	vii
Abstract	viii
Prefácio	ix
1 Enfoque freqüentista do modelo de calibração normal-assimétrico	1
1.1 Introdução	1
1.2 O modelo de calibração usual	5
1.3 A distribuição normal-assimétrica	6
1.4 Modelo de calibração linear normal-assimétrico	8
1.4.1 Algoritmo EM	12
1.4.2 Passo E	13
1.4.3 Passo M	15
1.5 Estudo de simulação	17
1.5.1 Discussão dos resultados	17
1.6 Variância do estimador de máxima verossimilhança	18
1.7 Prevenção de viés	26
1.8 Aplicação	27
2 Enfoque bayesiano do modelo de calibração normal-assimétrico	32
2.1 Introdução	32

<i>Índice</i>	xiii
2.2 O modelo hierárquico e o amostrador de Gibbs	34
2.2.1 O amostrador de Gibbs	35
2.3 Aplicação	39
3 Modelo de calibração t-normal-assimétrico	52
3.1 Introdução	52
3.2 A distribuição t-normal-assimétrica	53
3.3 Enfoque freqüentista	54
3.3.1 Obtenção dos estimadores através do algoritmo EM	54
3.3.2 Variância dos estimadores	62
3.3.3 Aplicação	63
3.4 Enfoque bayesiano	65
4 Conclusões Finais	70
A Informação esperada e observada do modelo normal-assimétrico	72
B Informação observada do modelo t-normal-assimétrico	78
C Comandos: modelo t-normal-assimétrico, abordagem clássica	83
D Comandos: modelo t-normal-assimétrico, abordagem bayesiana	91
E DIC, ADIC e EDIC	96
Referências bibliográficas	103

Enfoque freqüentista do modelo de calibração normal-assimétrico

1.1 Introdução

Os modelos de calibração são de suma importância em diversas áreas do conhecimento humano. Na Biologia ou Química (vide, por exemplo, Thomas e Ge, 2000) a técnica de calibração é utilizada para determinar concentrações de substâncias. Na Física e Engenharia (vide, por exemplo, Guimarães-Filho e Helene, 2003), tem sido utilizada para calibração de instrumentos de grandezas físicas. Em Medicina (vide, por exemplo, Costa e Caldas, 2003), também é utilizada para a calibração de aparelhos que medem a pressão sanguínea, o nível de colesterol ou temperatura, entre outros. Em Economia ou Administração, pode ser utilizada para tomada de decisões em cenários de instabilidade já vividos, através da previsão das possíveis causas e rápida intervenção (vide, por exemplo, McDonald, 2009). Em *Marketing* (vide, Pillai e Hofacker, 2007) podemos calibrar a relação entre veracidade e confiança do conhecimento sobre os conteúdos que são disponibilizados na *web*, de modo que um indivíduo estará bem calibrado se for realística a avaliação que o mesmo faz do conhecimento que possui. Toda vez que quisermos determinar concentrações de materiais em soluções,

composições de materiais, ou fazer avaliações de ferramentas físicas, bolar estratégias para redução de prejuízo ou mesmo maximizar o lucro, estaremos frente a um problema que envolve calibração. Branco (1997) ilustra a utilização de calibração aplicada à genética, conjecturando a situação de um possível acidente nuclear, onde um certo trabalhador foi exposto a uma dose de radiação desconhecida, x_0 . O número de células efetivamente afetadas pela radiação foi determinado por meio de amostras de sangue do trabalhador e expresso por y_0 . Por questões éticas, o estimador da dose, x_0 , de radiação recebida pelo trabalhador deve ser obtido em um experimento que envolva medições laboratoriais que espelhem as mesmas características do hipotético acidente nuclear com cobaias, por exemplo. Assim, é suficiente expor as cobaias a doses conhecidas de radiação x_1, \dots, x_n e anotar os efeitos citogenéticos produzidos y_1, \dots, y_n . Este procedimento é conhecido na literatura por *calibração controlada*.

A calibração usual ou controlada é caracterizada pela observação de duas variáveis X e Y , que se relacionam através de uma função conhecida f . Neste trabalho X representa o valor de uma característica de interesse que está definida em um intervalo I . Este intervalo é chamado intervalo de calibração e Y , por outro lado, representa a variável resposta que está relacionada a X .

No primeiro passo, ou no *processo de calibração*, selecionamos n observações de uma variável aleatória Y , digamos (y_1, \dots, y_n) , por meio de valores pré-fixados de X , digamos (x_1, \dots, x_n) , a fim de estimar a função que relaciona ambas variáveis. Em um segundo momento, agora *de fato no experimento de calibração*, observamos (y_{01}, \dots, y_{0k}) , k ($k \geq 1$) variáveis aleatórias, que são k realizações de uma variável aleatória Y , associadas a um valor desconhecido de X , o qual denotaremos por x_0 . O interesse reside na estimação desse valor desconhecido, x_0 , baseados no conhecimento da função f que foi estimada no primeiro passo da calibração. Naturalmente, o processo de estimação é baseado nos pares (x_i, y_i) , $i = 1, \dots, n$ e valores y_{0j} , $j = 1, \dots, k$. Em muitas partes do texto, não faremos distinção entre a variável aleatória,

escrita em letra maiúscula, e sua realização, em letra minúscula.

Quase toda a literatura que trata do problema de calibração assume que os erros associados a variável aleatória Y são normalmente distribuídos. Exemplos disso podem ser encontrados em Eisenhart (1939), Krutchkoff (1967) e Brown (1993) para citar alguns. O ponto de vista que queremos estabelecer neste trabalho é a inserção de suposições que fogem do padrão simétrico da normal ou de outras distribuições que buscam modelar o problema de calibração pressupondo que a distribuição dos erros aleatórios é simétrica em torno de um eixo central. Lima *et al.* (2008) realizaram um estudo do problema de calibração assumindo que os erros do modelo poderiam ser modelados através de uma distribuição t de *Student*. A classe das distribuições elípticas foi proposta por vários autores e aqui citamos Branco (1997) no contexto bayesiano. Branco (1997) estudou o problema de calibração em dois cenários. No primeiro, assumiu que o vetor de erros, ε , era proveniente de uma distribuição elíptica dependente e, no segundo, considerou que o vetor de erros era proveniente de um modelo elíptico independente. Esse estudo generalizou o uso de diversas distribuições que poderiam ser utilizadas para modelar o erro aleatório do ponto de vista bayesiano e podemos citar a distribuição t de *Student*, bem como a Cauchy como caso particular; no entanto, as distribuições até então estudadas eram todas simétricas. Uma recente publicação nacional relacionada às distribuições elípticas assimétricas pode ser encontrada em Arellano-Valle e Branco (2004).

Azzalini (1985) definiu a distribuição normal-assimétrica e suas propriedades básicas para o modelo de locação-escala. Arellano-Valle *et al.* (2005) apresentou os principais resultados teóricos para modelos de regressão com erros nas variáveis que incorporavam a assimetria por meio da suposição de que os erros eram normal-assimetricamente distribuídos.

Assim como em outras situações, a suposição de simetria dos erros do modelo de calibração pode ser questionada. Muitas são as situações em que as observações

não são oriundas de distribuições simétricas. Por exemplo, Azzalini e Dalla-Valle (1996) estudaram um conjunto de dados¹, onde um simples histograma mostra claramente a natureza assimétrica dos dados. Arellano-Valle *et al.* (2003, 2004) modelaram algumas outras aplicações através da incorporação da suposição de assimetria dos erros e utilizaram a distribuição normal-assimétrica para isso. Além disso, Lachos (2004) e Rodriguez (2005) estudaram modelos mistos normal-assimétricos (vide Arellano-Valle, Bolfarine e Lachos, 2005) e inferência bayesiana em modelos normal-assimétricos (vide Bayes e Branco, 2005), respectivamente.

Neste capítulo daremos algumas definições que serão úteis para o entendimento do texto. Iniciaremos com a definição usual do modelo de calibração e algumas propriedades da distribuição normal-assimétrica. O restante está organizado de modo a apresentar os principais resultados teóricos do modelo de calibração normal-assimétrico e implementar um algoritmo do tipo EM. Foram realizadas simulações que aparecem na Seção 1.5 deste capítulo, onde procuramos resumir o comportamento assintótico do estimador de máxima verossimilhança do parâmetro de calibração. As Seções 1.6 e 1.7 foram dedicadas às situações particulares de convergência do estimador de máxima verossimilhança. Em uma estudamos o comportamento do estimador numa vizinhança de zero, apresentando uma transformação para evitar a singularidade da matriz de informação de Fisher, que mostrou ser singular quando o estimador de máxima verossimilhança do parâmetro de assimetria tende a zero. O interessante da abordagem é não modificar o parâmetro de calibração, x_0 . O segundo cenário de convergência estudado na Seção 1.7 foi o comportamento do estimador no limite do infinito. Para encerrar o capítulo, aplicamos a metodologia proposta a um conjunto de dados reais, onde é ilustrada a melhor performance do modelo assimétrico de calibração. Este Capítulo se encontra em publicação no *Journal of*

¹TELFORD, R. e CUNNINGHAM, R. *apud* AZZALINI, A. Disponível em: << <http://pbil.univ-lyon1.fr/library/sn/html/ais.html> >>. Acesso em: 15 de setembro de 2006.

Applied Statistics (vide Figueiredo *et al.*, 2009).

1.2 O modelo de calibração usual

Algumas idéias básicas serão apresentadas na seqüência.

Definição 1: O modelo de calibração linear usual é caracterizado por

$$\begin{aligned} y_i &= \alpha + \beta x_i + \varepsilon_i, & i = 1, \dots, n; \\ y_{0j} &= \alpha + \beta x_0 + \varepsilon_{0j}, & j = 1, \dots, k; \end{aligned} \quad (1.1)$$

onde $\varepsilon_1, \dots, \varepsilon_n, \varepsilon_{01}, \dots, \varepsilon_{0k}$ são independentes e identicamente distribuídos (i.i.d) com média igual a zero e variância igual a σ^2 . Além disso, x_1, \dots, x_n são constantes conhecidas e $\alpha, \beta, x_0, \sigma^2$ são parâmetros desconhecidos.

Lema 1: Sob a Definição 1, o estimador de máxima verossimilhança (EMV) de x_0 , sob a suposição de normalidade, é chamado **estimador clássico** e é definido por

$$\hat{x}_{0C} = \frac{\bar{y}_0 - \hat{\alpha}}{\hat{\beta}}.$$

Definição 2: Sejam $\hat{\phi}$ e $\hat{\gamma}$ os EMV de ϕ e γ obtidos do modelo $x_i = \phi + \gamma y_i + \varepsilon_i$, $i = 1, \dots, n$. Então

$$\hat{x}_{0I} = \hat{\phi} + \hat{\gamma} \bar{y}_0$$

é um outro estimador de x_0 , chamado de **estimador inverso**.

O problema de calibração é muito antigo. Eisenhart (1939) estudou esse problema, e mais tarde, Krutchkoff (1967) trouxe à tona alguns argumentos para a utilização do estimador inverso quando o interesse reside em um problema de interpolação e, um pouco mais adiante, (Krutchkoff - 1969²), para o caso de extrapolação.

²KRUTCHKOFF, R. *apud* BRANCO, M. D. *Calibração: uma abordagem bayesiana*. São Paulo, 1997. Tese (Doutorado em Estatística) - Departamento de Estatística, IME-USP.

Como uma extensão natural do modelo de calibração com erros normalmente distribuídos vastamente estudado, consideramos que os erros do modelo de calibração linear seguem uma distribuição normal-assimétrica e o modelo de calibração normalmente distribuído será tratado como um caso particular.

Com esta nova suposição, o modelo de calibração linear, antes com quatro parâmetros, designados por α , β , x_0 e σ^2 passa a ter cinco parâmetros, designados por α , β , x_0 , σ^2 e λ . Este último é resultado da suposição de assimetria, que definiremos na próxima seção.

1.3 A distribuição normal-assimétrica

Nesta seção, vamos apresentar a distribuição normal-assimétrica, segundo a definição de Azzalini (1985). Assim, uma variável aleatória Z possui uma distribuição normal-assimétrica padrão se sua função densidade de probabilidade for dada por

$$\phi_{SN}(z|\lambda) = 2\phi(z)\Phi(\lambda z),$$

sendo que $\phi(\cdot)$ e $\Phi(\cdot)$ representam a função densidade de probabilidade (fdp) e a função de distribuição acumulada (fda) da distribuição normal padrão, respectivamente. Usaremos a notação $Z \sim SN(\lambda)$, onde o parâmetro λ controla a assimetria e a curtose. Quando positivo, $\lambda > 0$, temos um caso de assimetria à direita e no caso negativo, $\lambda < 0$, temos um caso de assimetria à esquerda. Se $\lambda = 0$ a distribuição é considerada simétrica e se reduz à distribuição normal padrão.

Algumas propriedades da distribuição normal-assimétrica são dadas em Azzalini (1985) e Henze (1986):

1. A média e a variância de Z são dadas, respectivamente, por

$$E[Z] = \sqrt{\frac{2}{\pi}}\delta \text{ e } Var[Z] = 1 - \frac{2}{\pi}\delta^2,$$

sendo que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$;

2. O coeficiente de assimetria e de curtose de Z são dados, respectivamente, por

$$\gamma = \frac{\sqrt{2}(4-\pi)}{\pi\sqrt{\pi}} \text{sinal}(\delta) \left(1 - 2\frac{\delta^2}{\pi}\right)^{-3/2} \quad \text{e} \quad \kappa = \frac{8(\pi-3)\delta^4}{(\pi-2\delta^2)^2}.$$

o que significa que $-0.9953 < \gamma < 0,9953$ e $0 < \kappa < 0.8692$, sendo que $\text{sinal}(\cdot)$ é uma função que atribui o valor 1, quando δ é positivo, e -1 , caso contrário;

3. A densidade de Z é log-côncava;
4. Uma importante representação estocástica apresentada por Andel *et al.* (1984) e estendida por Henze (1986) que é utilizada na grande maioria dos trabalhos que tratam da distribuição normal-assimétrica, tornando-se uma fundamental ferramenta para o desenvolvimentos dessa teoria, é apresentada como segue: sejam $X \sim HN(0;1)$ e $Y \sim N(0;1)$, respectivamente, variáveis aleatórias normal padrão truncada à esquerda de zero e normal padrão, respectivamente. Se X e Y são variáveis aleatórias independentes, então a distribuição de $Z = \delta X + (1 - \delta^2)^{1/2}Y$ será $SN(\lambda(\delta))$, com $\lambda = \frac{\delta}{(1-\delta^2)^{1/2}}$. Podemos representar os modelos usando λ (parametrização “lambda”) ou δ (parametrização “delta”);
5. A representação estocástica dada no item 4 também segue se considerarmos que a distribuição de $Z|X$ (Z dado X) é normal com média δx e variância $1 - \delta^2$, ou seja, se $Z|X = x \sim N(\delta x; 1 - \delta^2)$ e $X \sim HN(0;1)$, então Z tem a distribuição $SN(\lambda(\delta))$.

Observação 1. Se uma variável aleatória X segue uma distribuição normal assimétrica com parâmetro de locação μ e parâmetro de escala σ^2 , então a fdp de X é dada por

$$f_X(x) = \frac{2}{\sigma} \phi_1\left(\frac{x-\mu}{\sigma}\right) \Phi_1\left(\lambda \frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}, \quad (1.2)$$

com a notação $X \sim SN(\mu; \sigma^2; \lambda)$. A densidade acima é denotada por $\phi(x|\mu; \sigma^2; \lambda)$. Se $\lambda = 0$, a densidade de X é reduzida a uma fdp $N(\mu; \sigma^2)$.

Observação 2. Através de aplicações diretas das propriedades da normal-assimétrica e transformação de variáveis, segue que, se $Z \sim SN(\mu; \sigma^2; \lambda)$, então $Z^* = aZ + b \sim SN(a\mu + b; a^2\sigma^2; \text{sinal}(a)\lambda)$, onde $\text{sinal}(\cdot)$ foi definido acima.

1.4 Modelo de calibração linear normal-assimétrico

O modelo de calibração linear normal-assimétrico é caracterizado por duas etapas

$$\begin{aligned} y_i &= \alpha + \beta x_i + \varepsilon_i, & i &= 1, \dots, n; \\ y_{0j} &= \alpha + \beta x_0 + \varepsilon_{0j}, & j &= 1, \dots, k, \end{aligned} \tag{1.3}$$

sendo o interesse estimar o parâmetro x_0 , tal que, ε_i e ε_{0j} serão supostos independentes e identicamente distribuídos com distribuição $SN(0; \sigma^2; \lambda)$, onde 0 , σ^2 e λ são, respectivamente, os parâmetros de locação, de escala e de forma. Note que quando $\lambda = 0$, temos o modelo de calibração usual com erros normalmente distribuídos. Considera-se x_1, \dots, x_n constantes conhecidas e α , β , x_0 , σ^2 e λ parâmetros desconhecidos. As variáveis y_1, \dots, y_n e y_{01}, \dots, y_{0j} foram definidas no início do Capítulo. Com as suposições acima, segue que

$$\begin{aligned} y_i|x_i &\sim SN(\alpha + \beta x_i; \sigma^2; \lambda), & i &= 1, \dots, n; \\ y_{0j}|x_0 &\sim SN(\alpha + \beta x_0; \sigma^2; \lambda), & j &= 1, \dots, k. \end{aligned}$$

Conseqüentemente, a função de verossimilhança pode ser escrita como

$$L(\theta|y; y_0) = \prod_{i=1}^n \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right\} \times \prod_{i=1}^n \Phi\left(\frac{\lambda}{\sigma}(y_i - \alpha - \beta x_i)\right) \\ \times \prod_{j=1}^k \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_{0j} - \alpha - \beta x_0)^2\right\} \times \prod_{j=1}^k \Phi\left(\frac{\lambda}{\sigma}(y_{0j} - \alpha - \beta x_0)\right). \quad (1.4)$$

Os estimadores de máxima verossimilhança do vetor de parâmetros do modelo, $\theta = (\alpha, \beta, \sigma, \lambda, x_0)$, poderiam ser obtidos por maximização direta da função de verossimilhança acima. No entanto, Azzalini (1985) apresenta alguns problemas que podem ocorrer no processo de otimização, tal como, a singularidade da matriz de informação de Fisher, quando $\lambda = 0$, ou ainda, Sartori (2003), na abordagem freqüentista, ou Liseo e Loperfido (2006), na abordagem bayesiana, salientam que no processo de estimação de λ , o estimador pode tender a infinito com probabilidade positiva. Chiogna (2005) apresenta as verossimilhanças perfiladas em relação ao estimador de λ e pudemos perceber que nas proximidades de 0 temos um ponto de inflexão nessa verossimilhança, além dela ser razoavelmente *flat* a ponto de termos convergência para valores altos de $\hat{\lambda}$. Ou seja, a maximização direta parece não ser robusta com respeito aos valores iniciais, isto é, se não são utilizados valores iniciais adequados, o método de maximização pode não convergir. Sartori (2003) sugere um método de estimação direta em duas etapas para contornar o problema de não convergência. Um algoritmo do tipo EM está disponível na biblioteca `sn.zip`³ da linguagem R⁴ para a estimação dos parâmetros do modelo de locação-escala. Neste contexto, surge a idéia de estimar o vetor de parâmetros θ por meio do algoritmo EM, como uma extensão natural do que já existe. Vamos primeiro rememorar algumas propriedades básicas das distribuições truncadas.

³CRAN

⁴VENABLES, W. N. , SMITH, D. M. *An Introduction to R. R: A Programming Environment for Data Analysis and Graphics*. Versão 2.3.1. 2006. Disponível em: \ll <http://cran.r-project.org/doc/manuals/R-intro.pdf> \gg . Acesso em: 15 de setembro de 2006.

Lema 2: *Seja $T \sim TN(\mu; \sigma)I\{a < t < b\}$ uma variável normal duplamente truncada (vide Johnson et al., 1994). A função densidade de probabilidade de T é dada por:*

$$f_T(t) = \left\{ \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right\}^{-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}, \quad a < t < b,$$

onde a e b são os pontos inferior e superior de truncamento. Então, quando $a = 0$ e $b = \infty$ teremos uma variável com distribuição normal-truncada à esquerda; com função densidade de probabilidade

$$\begin{aligned} f_T(t) &= \left\{ 1 - \Phi\left(-\frac{\mu}{\sigma}\right) \right\}^{-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\} \\ &= \left\{ \Phi\left(\frac{\mu}{\sigma}\right) \right\}^{-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}, \quad 0 < t < \infty. \end{aligned}$$

Além disso, a esperança e o segundo momento de T são, respectivamente,

$$E(T) = \mu + \frac{\left\{ \phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right) \right\}}{\left\{ \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right\}} \sigma$$

e

$$E(T^2) = \mu^2 + \sigma^2 + \frac{\left\{ \left(\frac{a-\mu}{\sigma}\right) \phi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right) \phi\left(\frac{b-\mu}{\sigma}\right) \right\}}{\left\{ \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right\}} \sigma^2 + 2\mu\sigma \frac{\left\{ \phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right) \right\}}{\left\{ \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right\}}.$$

No caso particular em que $a = 0$ e $b = \infty$, temos que

$$E(T) = \mu + \frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} \sigma$$

e

$$E(T^2) = \mu^2 + \sigma^2 - \frac{\mu}{\sigma} \frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} \sigma^2 + 2\mu\sigma \frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}$$

$$= \mu^2 + \sigma^2 + \mu\sigma \frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}.$$

Do Lema 2, segue que quando $a = \mu$ e $b = \infty$ teremos uma distribuição truncada particular (notação devido ao inglês: $HN(\mu; \sigma)$) com densidade

$$f_T(t) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t - \mu)^2\right\}, \quad \mu < t < \infty,$$

que é, a distribuição de $\mu + \sigma|U|$ sendo U uma variável com distribuição normal padrão. No texto chamaremos essa distribuição de meio-normal. A palavra “meio”, mesmo que invariante, se refere à idéia de metade e não a idéia de “quase”, ou seja, uma normal pela metade e não uma distribuição que é quase uma normal; o que não faria sentido algum.

Usando a representação estocástica descrita na Propriedade 4 e, considerando U_i, U_{0j}, V_i e V_{0j} variáveis aleatórias independentes com distribuição normal padrão, temos que

$$\begin{aligned} \varepsilon_i &= \sigma\delta|U_i| + \sqrt{(1 - \delta^2)}V_i\sigma \sim SN(0; \sigma^2; \lambda), \quad i = 1, \dots, n; \\ \varepsilon_{0j} &= \sigma\delta|U_{0j}| + \sqrt{(1 - \delta^2)}V_{0j}\sigma \sim SN(0; \sigma^2; \lambda), \quad j = 1, \dots, k. \end{aligned} \quad (1.5)$$

Agora, utilizando (1.5) acima, podemos reescrever o modelo de calibração dado em (1.3), trocando ε_i e ε_{0j} em (1.3) tal que

$$y_i = \alpha + \beta x_i + \sigma\delta t_i + r_i \quad (1.6)$$

$$y_{0j} = \alpha + \beta x_0 + \sigma\delta t_{0j} + r_{0j} \quad (1.7)$$

sendo $t_i = |U_i|$, $t_{0j} = |U_{0j}|$, $r_i = \sigma(1 - \delta^2)^{1/2}V_i$ e $r_{0j} = \sigma(1 - \delta^2)^{1/2}V_{0j}$, com $i = 1, \dots, n$ e $j = 1, \dots, k$.

Temos que r_i e r_{0j} são normalmente distribuídos como $N(0; \sigma^2(1 - \delta^2))$ bem como t_i e t_{0j} são meio-normalmente distribuídos. Portanto, condicionando em t_i e t_{0j} e, além disso, substituindo $b = \sigma\delta$ e $\sigma_*^2 = \sigma^2(1 - \delta^2)$, teremos a seguinte forma hierárquica de representação do modelo de calibração:

$$y_i | t_i; x_i \sim N(\alpha + \beta x_i + bt_i; \sigma_*^2),$$

$$t_i \sim HN(0; 1) \quad i = 1, \dots, n;$$

$$y_{0j} | t_{0j} \sim N(\alpha + \beta x_0 + bt_{0j}; \sigma_*^2),$$

$$t_{0j} \sim HN(0; 1) \quad j = 1, \dots, k.$$

1.4.1 Algoritmo EM

A função de verossimilhança completa (aumentada), em relação a $\theta^* = (\alpha, \beta, \sigma_*, b, x_0)$, é essencial para a implementação do algoritmo EM e será dada por

$$L(\theta^* | z) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left\{-\frac{1}{2\sigma_*^2}(y_i - \alpha - \beta x_i - bt_i)^2\right\} \times \sqrt{2/\pi} \exp\left\{-\frac{1}{2}(t_i)^2\right\} \quad (1.8)$$

$$\times \prod_{j=1}^k \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left\{-\frac{1}{2\sigma_*^2}(y_{0j} - \alpha - \beta x_0 - bt_{0j})^2\right\} \times \sqrt{2/\pi} \exp\left\{-\frac{1}{2}(t_{0j})^2\right\}.$$

desde que $z = (y, y_0, t, t_0)$ seja o vetor com os dados aumentados, acoplando tanto as observações efetivamente observadas $y = y_1, \dots, y_n$ e $y_0 = y_{01}, \dots, y_{0j}$, quanto as variáveis latentes $t = t_1, \dots, t_n$ e $t_0 = t_{01}, \dots, t_{0j}$.

O algoritmo EM (Dempster *et al.*, 1977) é obtido pela implementação de dois passos chamados de Passo E e Passo M, respectivamente.

1.4.2 Passo E

Aplicando o logaritmo natural (notação: “log”) à verossimilhança aumentada, temos o logaritmo da função de verossimilhança (ℓ) dado por

$$\begin{aligned} \ell(\theta^*|z) = & -n \log(\sigma_*^2) - \sum_i \frac{1}{2\sigma_*^2} (y_i - \alpha - \beta x_i - bt_i)^2 - \frac{1}{2} \sum_i (t_i)^2 \\ & -k \log(\sigma_*^2) - \sum_j \frac{1}{2\sigma_*^2} (y_{0j} - \alpha - \beta x_0 - bt_{0j})^2 - \frac{1}{2} \sum_j (t_{0j})^2. \end{aligned}$$

No Passo E do algoritmo EM é obtida a esperança condicional, $E_{z|y}(\ell(\theta^*))$.

Portanto,

$$\begin{aligned} E_{z|y}(\ell(\theta^*)) = & \\ & -n \log(\sigma_*^2) - \frac{1}{2\sigma_*^2} \left\{ \sum_i (y_i - \alpha - \beta x_i)^2 - 2 \sum_i (y_i - \alpha - \beta x_i) b E(t_i|y_i) \right. \\ & \left. + b^2 \sum_i E(t_i^2|y_i) \right\} - \frac{1}{2} \sum_i E((t_i|y_i)^2) \\ & -k \log(\sigma_*^2) - \frac{1}{2\sigma_*^2} \left\{ \sum_j (y_{0j} - \alpha - \beta x_0)^2 - 2 \sum_j (y_{0j} - \alpha - \beta x_0) b E(t_{0j}|y_j) \right. \\ & \left. + b^2 \sum_j E(t_{0j}^2|y_j) \right\} - \frac{1}{2} \sum_j E((t_{0j}|y_j)^2) \end{aligned}$$

Como uma extensão do resultado descrito em Arellano *et al.* (2005) podemos mostrar que $T|Y$ tem distribuição normal-truncada.

Lema 3: *Com as suposições que foram incorporadas ao modelo (1.3)*

$$T|Y \sim TN\left((y - \alpha - \beta x) \frac{b}{b^2 + \sigma_*^2}; \frac{\sigma_*^2}{b^2 + \sigma_*^2}\right) I(0 < t < \infty).$$

Demonstração Como $f_{Y,T}(y, t) = f_{Y|T=t}(y) f_T(t) = \frac{1}{\sqrt{2\pi\sigma_*^2}} \sqrt{\frac{2}{\pi}} \exp\left[-\frac{(y-\alpha-\beta x-bt)^2}{2\sigma_*^2} - \frac{t^2}{2}\right]$, assim a idéia básica é fatorar esta expressão em dois fatores, a fim de obtermos $f_{T|Y=y}(t) f_Y(y)$. Então $f_{Y,T}(y, t)$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma_*^2}} \sqrt{\frac{2}{\pi}} \exp \left\{ -\frac{1}{2\sigma_*^2} [(y - \alpha - \beta x - bt)^2 + \sigma_*^2 t^2] \right\} \\
&= \frac{1}{\pi\sigma_*} \exp \left\{ -\frac{1}{2\sigma_*^2} [(y - \alpha - \beta x)^2 - 2(y - \alpha - \beta x)bt + b^2 t^2 + \sigma_*^2 t^2] \right\} \\
&= \frac{1}{\pi\sigma_*} \exp \left\{ -\frac{1}{2\sigma_*^2} (y - \alpha - \beta x)^2 \right\} \times \exp \left\{ -\frac{1}{2\sigma_*^2} (b^2 + \sigma_*^2) \left[t^2 - \frac{2(y - \alpha - \beta x)bt}{(b^2 + \sigma_*^2)} \right] \right\} \\
&= \frac{1}{\pi\sigma_*} \exp \left\{ -\frac{1}{2\sigma_*^2} (y - \alpha - \beta x)^2 \right\} \times \exp \left\{ +\frac{(b^2 + \sigma_*^2)}{2\sigma_*^2} \left[\frac{b^2 (y - \alpha - \beta x)^2}{(b^2 + \sigma_*^2)^2} \right] \right\} \\
&\times \exp \left\{ -\frac{(b^2 + \sigma_*^2)}{2\sigma_*^2} \left[t - \frac{(y - \alpha - \beta x)b}{(b^2 + \sigma_*^2)} \right]^2 \right\} \\
&= \frac{1}{\pi\sigma_*} \exp \left\{ -\frac{1}{2\sigma_*^2} \left[(y - \alpha - \beta x)^2 - \frac{b^2 (y - \alpha - \beta x)^2}{(b^2 + \sigma_*^2)} \right] \right\} \\
&\quad \times \exp \left\{ -\frac{(b^2 + \sigma_*^2)}{2\sigma_*^2} \left[t - \frac{(y - \alpha - \beta x)b}{(b^2 + \sigma_*^2)} \right]^2 \right\} \\
&= \frac{1}{\pi\sigma_*} \exp \left\{ -\frac{(y - \alpha - \beta x)^2}{2(b^2 + \sigma_*^2)} \right\} \times \exp \left\{ -\frac{(b^2 + \sigma_*^2)}{2\sigma_*^2} \left[t - \frac{(y - \alpha - \beta x)b}{(b^2 + \sigma_*^2)} \right]^2 \right\}
\end{aligned}$$

Note que o último fator é o núcleo de uma $TN\left((y - \alpha - \beta x)\frac{b}{b^2 + \sigma_*^2}; \frac{\sigma_*^2}{b^2 + \sigma_*^2}\right)$ e o restante não depende de T , só de Y . Portanto, depois de integrarmos em Y , teremos

$$f_{T|Y=y}(t) = \sqrt{\frac{2(b^2 + \sigma_*^2)}{\pi\sigma_*^2}} \exp \left\{ -\frac{(b^2 + \sigma_*^2)}{2\sigma_*^2} \left[t - \frac{(y - \alpha - \beta x)b}{(b^2 + \sigma_*^2)} \right]^2 \right\}.$$

□

Utilizando os Lemas 2 e 3, obtemos o primeiro e o segundo momentos da distribuição de $T|Y$ que resultam em

$$\hat{w}_{1i} = E(t|y = y_i; \hat{\theta}^*) = (y_i - \hat{\alpha} - \hat{\beta}x_i) \frac{\hat{b}}{\hat{b}^2 + \hat{\sigma}_*^2} + \frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2 + \hat{\sigma}_*^2}} \frac{\phi(\cdot)}{\Phi(\cdot)} \quad e$$

$$\begin{aligned}\hat{w}_{2i} = E(t^2|y = y_i; \hat{\theta}^*) &= (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \frac{\hat{b}^2}{(\hat{b}^2 + \hat{\sigma}_*^2)^2} + \frac{\hat{\sigma}_*^2}{\hat{b}^2 + \hat{\sigma}_*^2} \\ &+ (y_i - \hat{\alpha} - \hat{\beta}x_i) \frac{\hat{b}}{\hat{b}^2 + \hat{\sigma}_*^2} \frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2 + \hat{\sigma}_*^2}} \frac{\phi(\cdot)}{\Phi(\cdot)},\end{aligned}$$

onde $(\cdot) = \left(\frac{\hat{b}}{\hat{\sigma}_*} \frac{(y_i - \hat{\alpha} - \hat{\beta}x_i)}{\sqrt{\hat{b}^2 + \hat{\sigma}_*^2}}\right)$, quando estamos na etapa de regressão, isto é, quando $i = 1, \dots, n$, e

$$\begin{aligned}\hat{w}_{1j} = E(t|y = y_{0j}; \hat{\theta}^*) &= (y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0) \frac{\hat{b}}{\hat{b}^2 + \hat{\sigma}_*^2} + \frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2 + \hat{\sigma}_*^2}} \frac{\phi(\cdot)}{\Phi(\cdot)} \quad \text{e} \\ \hat{w}_{2j} = E(t^2|y = y_{0j}; \hat{\theta}^*) &= (y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2 \frac{\hat{b}^2}{(\hat{b}^2 + \hat{\sigma}_*^2)^2} + \frac{\hat{\sigma}_*^2}{\hat{b}^2 + \hat{\sigma}_*^2} \\ &+ (y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0) \frac{\hat{b}}{\hat{b}^2 + \hat{\sigma}_*^2} \frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2 + \hat{\sigma}_*^2}} \frac{\phi(\cdot)}{\Phi(\cdot)},\end{aligned}$$

onde $(\cdot) = \left(\frac{\hat{b}}{\hat{\sigma}_*} \frac{(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)}{\sqrt{\hat{b}^2 + \hat{\sigma}_*^2}}\right)$, quando estamos na etapa de calibração, isto é, quando $j = 1, \dots, k$.

1.4.3 Passo M

Para implementar o passo M é necessário maximizar $\ell(\theta^*|z)$ em relação aos parâmetros do vetor θ^* . Assim, depois de algumas manipulações algébricas e com $m = 0, 1, 2, \dots$ indicando a etapa do processo iterativo, temos que

$$\hat{b}^{(m+1)} = \frac{Syw_{(i)} + Sy_0w_{(j)} - \frac{SxySxw_{(i)}}{Sxx}}{\sum \hat{w}_{2i} + \sum \hat{w}_{2j} - \bar{w}_{1(i)} \sum \hat{w}_{1i} - \bar{w}_{1(j)} \sum \hat{w}_{1j} - \frac{(Sxw_{(i)})^2}{Sxx}};$$

desde que $Syw_{(i)} = \sum (y_i - \bar{y})(\hat{w}_{1i} - \bar{w}_{1(i)})$, $Sxw_{(i)} = \sum (x_i - \bar{x})(\hat{w}_{1i} - \bar{w}_{1(i)})$, $Sxy = \sum (x_i - \bar{x})(y_i - \bar{y})$, $Sxx = \sum (x_i - \bar{x})^2$, $Sy_0w_{(j)} = \sum (y_{0j} - \bar{y}_0)(\hat{w}_{1j} - \bar{w}_{1(j)})$, $\bar{w}_{1(i)} = \frac{\sum \hat{w}_{1i}}{n}$ e $\bar{w}_{1(j)} = \frac{\sum \hat{w}_{1j}}{k}$. Além disso,

$$\hat{\beta}^{(m+1)} = \frac{Sxy}{Sxx} - \frac{\hat{b}^{(m)} Sxw_{(i)}}{Sxx};$$

$$\hat{\alpha}^{(m+1)} = \bar{y} - \hat{\beta}^{(m)}\bar{x} - \hat{b}^{(m)}\bar{w}_{1(i)};$$

$$\hat{x}_0^{(m+1)} = \frac{\bar{y}_0 - \hat{\alpha}^{(m)}}{\hat{\beta}^{(m)}} - \frac{\hat{b}^{(m)}\bar{w}_{1(j)}}{\hat{\beta}^{(m)}}; \quad \text{e}$$

$$\hat{\sigma}_*^{(m+1)} = \left[\frac{\left(\sum_i (y_i - \hat{\alpha}^{(m)} - \hat{\beta}^{(m)}x_i)^2 - 2 \sum_i (y_i - \hat{\alpha}^{(m)} - \hat{\beta}^{(m)}x_i)\hat{b}^{(m)}\hat{w}_{1i} + \hat{b}^{2(m)} \sum_i \hat{w}_{2i} \right)}{(n+k)} \right. \\ \left. + \frac{\left(\sum_j (y_{0j} - \hat{\alpha}^{(m)} - \hat{\beta}^{(m)}\hat{x}_0)^2 - 2 \sum_j (y_{0j} - \hat{\alpha}^{(m)} - \hat{\beta}^{(m)}\hat{x}_0)\hat{b}^{(m)}\hat{w}_{1j} + \hat{b}^{2(m)} \sum_j \hat{w}_{2j} \right)}{(n+k)} \right]^{\frac{1}{2}}.$$

Com estas expressões, o algoritmo EM segue como o seguinte processo iterativo:

1. Atribua valores iniciais $\hat{b}^{(0)}$, $\hat{\alpha}^{(0)}$, $\hat{\beta}^{(0)}$, $\hat{\sigma}_*^{(0)}$ e $\hat{x}_0^{(0)}$ para o vetor de parâmetros θ^* .
2. Calcule \hat{w}_{1i} e \hat{w}_{2i} , $\forall i$, também \hat{w}_{1j} e \hat{w}_{2j} , $\forall j$ e todas as expressões que dependem destes valores;
3. Obtenha novos valores para os estimadores \hat{b} , $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_*$ e \hat{x}_0 ; e
4. Repita os passos 2 e 3 até a convergência.

Os valores obtidos após a convergência para os estimadores \hat{b} , $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_*^2$ e \hat{x}_0 são as estimativas de máxima verossimilhança do modelo de calibração. Note que, quando $b = 0$, os estimadores serão os mesmos dados pelo modelo de calibração com erros normalmente distribuídos (vide Lima, 1996, ou Achic, 2005), dados por

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}};$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x};$$

$$\hat{x}_0 = \frac{\bar{y}_0 - \hat{\alpha}}{\hat{\beta}}; \quad \text{e}$$

$$\hat{\sigma} = \left[\frac{\sum_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum_j (y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2}{(n+k)} \right]^{\frac{1}{2}}.$$

1.5 Estudo de simulação

Estudamos o comportamento do EMV de x_0 através do vício e do EQM. Utilizando a linguagem R, geramos 1000 amostras de acordo com o modelo de calibração normal-assimétrico considerando $\alpha = 3$, $\beta = 1,5$, $x_0 = 15$ e 40 , $\sigma^2 = 0,1; 1$ e 10 e $\lambda = 0,5; 2$ e 5 para $n = 30; 50; 100$ e 200 . Os valores x_1, \dots, x_n foram gerados de uma distribuição uniforme $U(10; 30)$.

Para cada amostra, os estimadores foram obtidos. A média das estimativas, o vício e o EQM foram avaliados por $\sum_{r=1}^{1000} \hat{x}_{0r}/1000$, $\sum_{r=1}^{1000} (x_0 - \hat{x}_{0r})/1000$ e $\sum_{r=1}^{1000} (x_0 - \hat{x}_{0r})^2/1000$, respectivamente. As Tabelas 1.1 e 1.2 trazem os resultados do estudo de simulação.

Tabela 1.1: Resultados da simulação - $x_0 = 15$

Tabela 1.1: Resultados da simulação - $x_0 = 15$											
			$\sigma^2 = 0,1$			$\sigma^2 = 1$			$\sigma^2 = 10$		
n	λ	\hat{x}_0	Vício	EQM	\hat{x}_0	Vício	EQM	\hat{x}_0	Vício	EQM	
30	0,5	15,00	0,00	0,00	14,95	-0,05	0,26	14,30	-0,70	35,83	
30	2	15,00	0,00	0,00	14,98	-0,02	0,13	15,02	0,02	16,01	
30	5	15,00	0,00	0,00	15,00	0,00	0,10	16,27	1,27	13,61	
50	0,5	15,00	0,00	0,00	14,96	-0,04	0,22	14,58	-0,42	26,86	
50	2	15,00	0,00	0,00	14,97	-0,03	0,11	15,56	0,56	14,94	
50	5	15,00	0,00	0,00	15,01	0,01	0,10	16,56	1,56	10,77	
100	0,5	15,00	0,00	0,00	14,97	-0,03	0,21	14,97	-0,03	21,00	
100	2	15,00	0,00	0,00	15,01	0,01	0,11	15,47	0,47	12,08	
100	5	15,00	0,00	0,00	14,99	-0,01	0,09	16,63	1,63	10,49	
200	0,5	15,00	0,00	0,00	14,98	-0,02	0,20	14,93	-0,07	20,38	
200	2	15,00	0,00	0,00	15,01	0,01	0,11	15,72	0,72	11,16	
200	5	15,00	0,00	0,00	14,98	-0,02	0,08	16,43	1,43	9,92	

1.5.1 Discussão dos resultados

Podemos observar que, tanto para a interpolação de x_0 , ou seja, caso em que $x_0 = 15$ quanto para a extrapolação de x_0 , ou seja, caso em que $x_0 = 40$, independente da variabilidade, o EQM tende a diminuir à medida que aumenta a assimetria, ou

Tabela 1.2: Resultados da simulação - $x_0 = 40$

n	λ	$\sigma^2 = 0,1$			$\sigma^2 = 1$			$\sigma^2 = 10$		
		\hat{x}_0	Vício	EQM	\hat{x}_0	Vício	EQM	\hat{x}_0	Vício	EQM
30	0,5	40,01	0,01	0,00	40,06	0,06	0,40	41,85	1,85	69,76
30	2	40,01	0,01	0,00	40,14	0,14	0,24	41,61	1,61	29,82
30	5	40,02	0,02	0,00	40,16	0,16	0,17	42,20	2,20	22,03
50	0,5	40,00	0,00	0,00	40,03	0,03	0,32	40,79	0,79	38,29
50	2	40,01	0,01	0,00	40,07	0,07	0,19	41,14	1,14	20,65
50	5	40,02	0,02	0,00	40,17	0,17	0,14	41,92	1,92	17,10
100	0,5	40,00	0,00	0,00	40,04	0,04	0,24	40,61	0,61	26,69
100	2	40,00	0,00	0,00	40,07	0,07	0,12	40,63	0,63	13,99
100	5	40,01	0,01	0,00	40,16	0,16	0,13	41,66	1,66	12,81
200	0,5	40,01	0,01	0,00	40,00	0,00	0,21	40,17	0,17	21,55
200	2	40,00	0,00	0,00	40,05	0,05	0,12	40,62	0,62	12,39
200	5	40,01	0,01	0,00	40,14	0,14	0,10	41,39	1,39	10,42

seja, quanto mais assimétrico os dados, mais preciso é o estimador; além disso, o EQM de x_0 tende a diminuir à medida que o tamanho da amostra aumenta, ou seja, o estimador é consistente.

Tanto o vício quanto o EQM aumentam à medida que a variabilidade aumenta, independente da assimetria e do tamanho da amostra.

Embora a extrapolação de x_0 seja pouco utilizada na prática, foi verificado também que o EQM é maior na extrapolação de x_0 , quando $\sigma^2 \geq 1$, independente de n .

1.6 Variância do estimador de máxima verossimilhança

Aqui, a idéia é derivar uma aproximação assintótica para a variância do estimador de máxima verossimilhança, utilizando a função escore do vetor de parâmetros θ , digamos $S_\theta = \partial l(\theta)/\partial \theta$, um vetor de dimensão 5×1 . É vastamente sabido que no caso univariado o Limite Inferior de Cramer-Rao para $\hat{\theta}$, um estimador não viciado

de θ , baseado numa função de verossimilhança regular, especifica que

$$\text{var}(\hat{\theta}) \geq I_F^{-1}(\theta)$$

onde o limite inferior é alcançado se e só se $\hat{\theta}$ é uma combinação linear da função escore $S_\theta(\theta)$. $I_F(\theta)$ a matriz de informação de Fisher correspondente ao vetor de parâmetros θ . Também é largamente sabido que sob certas condições de regularidade o EMV atinge assintoticamente o Limite Inferior de Cramer-Rao (Sen e Singer, 1993).

Algumas idéias e discussões sobre a variabilidade do vetor de parâmetros do modelo de calibração linear serão dadas na seqüência. Como o logaritmo da função de verossimilhança do Modelo (1.3) pode ser escrito como

$$\ell(\theta) = \sum_i l_i(\theta) + \sum_j l_{0j}(\theta),$$

onde

$$\sum_i l_i(\theta) = n \log \left(\frac{2}{\sqrt{2\pi\sigma^2}} \right) - \frac{1}{2\sigma^2} \sum_i (y_i - \alpha - \beta x_i)^2 + \sum_i \log \left(\Phi \left(\frac{\lambda}{\sigma} (y_i - \alpha - \beta x_i) \right) \right)$$

e

$$\sum_j l_{0j}(\theta) = k \log \left(\frac{2}{\sqrt{2\pi\sigma^2}} \right) - \frac{1}{2\sigma^2} \sum_j (y_{0j} - \alpha - \beta x_0)^2 + \sum_j \log \left(\Phi \left(\frac{\lambda}{\sigma} (y_{0j} - \alpha - \beta x_0) \right) \right),$$

então as componentes do vetor escore $S_\theta(\theta|y_l; x_l) = (S_\alpha, S_\beta, S_\sigma, S_\lambda, S_{x_0})$, onde $S_\theta = \partial \ell(\theta) / \partial \theta$, são dadas por

$$\begin{aligned} S_\alpha &= \sum_{i=1}^n \left[\frac{1}{\sigma^2} (y_i - \alpha - \beta x_i) - \frac{\lambda \phi(\cdot)}{\sigma \Phi(\cdot)} \right] + \sum_{j=1}^k \left[\frac{1}{\sigma^2} (y_{0j} - \alpha - \beta x_0) - \frac{\lambda \phi(\cdot)}{\sigma \Phi(\cdot)} \right]; \\ S_\beta &= \sum_{i=1}^n \left[\frac{x_i}{\sigma^2} (y_i - \alpha - \beta x_i) - \frac{\lambda x_i \phi(\cdot)}{\sigma \Phi(\cdot)} \right] + \sum_{j=1}^k \left[\frac{x_0}{\sigma^2} (y_{0j} - \alpha - \beta x_0) - \frac{\lambda x_0 \phi(\cdot)}{\sigma \Phi(\cdot)} \right]; \\ S_\sigma &= \sum_{i=1}^n \left[\frac{1}{\sigma} + \frac{1}{\sigma^3} (y_i - \alpha - \beta x_i)^2 - \frac{\lambda x_i}{\sigma^2} (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} \right] \end{aligned}$$

$$+ \sum_{j=1}^k \left[\frac{1}{\sigma} + \frac{1}{\sigma^3} (y_{0j} - \alpha - \beta x_0)^2 - \frac{\lambda x_0}{\sigma^2} (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} \right];$$

$$S_\lambda = \sum_{i=1}^n \left[\frac{1}{\sigma} (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} \right] + \sum_{j=1}^k \left[\frac{1}{\sigma} (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} \right];$$

e

$$S_{x_0} = \sum_{j=1}^k \left[\frac{\beta}{\sigma^2} (y_{0j} - \alpha - \beta x_0) - \frac{\lambda \beta}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} \right];$$

com $(\cdot) = \lambda(y_i - \alpha - \beta x_i)/\sigma$ e $(\cdot) = \lambda(y_{0j} - \alpha - \beta x_0)/\sigma$.

As expressões das segundas derivadas bem como de suas esperanças são facilmente calculadas e a matriz de informação esperada é obtida, calculando-se $I_F(\theta) = E[S(\theta)S(\theta)^\top]$, cujos cálculos aparecem no apêndice. No entanto, um estimador consistente da matriz de covariâncias assintótica deveria ser obtido por meio do cálculo da matriz inversa de $[S(\hat{\theta})^\top S(\hat{\theta})]$, tipicamente conhecida como a matriz de informação observada, mas estas expressões não podem ser obtidas devido a uma série de dificuldades que a distribuição normal-assimétrica apresenta. Assim, as condições de regularidade tipicamente associadas com o método de máxima verossimilhança não são válidas. Como consequência, propriedades assintóticas dos EMV podem não ser válidas, em particular, consistência e normalidade assintótica. Algumas dificuldades que foram observadas serão destacadas abaixo.

Observação 1. Note que a expressão da matriz de segundas derivadas quando a avaliamos numa vizinhança de $\lambda = 0$, usando $\hat{\sigma}^2 = [\sum_{i=1}^n (y_i - \alpha - \beta x_i) + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)] / (n + k)$ para simplificar a expressão, é dada pela matriz

$$\begin{bmatrix} -\frac{(n+k)}{\hat{\sigma}^2} & -\frac{\sum_{i=1}^n x_i + kx_0}{\hat{\sigma}^2} & 0 \\ -\frac{\sum_{i=1}^n x_i + kx_0}{\hat{\sigma}^2} & -\frac{\sum_{i=1}^n x_i^2 + kx_0^2}{\hat{\sigma}^2} & 0 \\ 0 & 0 & -\frac{3[\sum_{i=1}^n (y_i - \alpha - \beta x_i) + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)] - (n+k)\hat{\sigma}^2}{\hat{\sigma}^4} \\ -\frac{(n+k)}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} & -\frac{\sum_{i=1}^n x_i + kx_0}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} & 0 \\ -\frac{k\beta}{\hat{\sigma}^2} & -\frac{\sum_{j=1}^k (2\beta x_0 - y_{0j} + \alpha)}{\hat{\sigma}^2} & 0 \end{bmatrix}$$

$$\begin{array}{cc}
\begin{array}{c}
-\frac{(n+k)}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} \\
-\frac{\sum_{i=1}^n x_i + kx_0}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} \\
0 \\
-\frac{2}{\pi} \left[\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i) + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)}{\hat{\sigma}^2} \right] \\
-\frac{k\beta}{\hat{\sigma}} \frac{2}{\pi}
\end{array}
&
\begin{array}{c}
-\frac{k\beta}{\hat{\sigma}^2} \\
-\frac{\sum_{j=1}^k (2\beta x_0 - y_{0j} + \alpha)}{\hat{\sigma}^2} \\
0 \\
-\frac{k\beta}{\hat{\sigma}} \frac{2}{\pi} \\
-\frac{k\beta^2}{\hat{\sigma}^2}
\end{array}
\end{array}
\left. \vphantom{\begin{array}{c} \\ \\ \\ \\ \\ \end{array}} \right]$$

que tem claramente determinante menor que zero. Logo, podemos encontrar estimativas quaisquer para os outros parâmetros α , β , σ e x_0 , nessa vizinhança de interesse, tal que teremos um ponto de sela e não um máximo. Esse resultado é garantido pelo conhecido critério básico para identificação de máximos e mínimos baseado no Hessiano de uma matriz (vide Morettin, Hazzan e Bussab (2003), p 284, por exemplo).

Observação 2. Além disso, como relatado por Azzalini (1985), o modelo com três parâmetros apresenta matriz de informação de Fisher singular na vizinhança de $\lambda = 0$. Recentemente, Rodriguez (2005) apresentou algumas discussões sobre a matriz de informação de Fisher para o modelo de regressão múltipla no cenário $\lambda = 0$. Aqui, generalizamos essa informação de Fisher, estendendo-a com a inserção do parâmetro de calibração e investigando o comportamento dela no mesmo cenário $\lambda = 0$ que engloba os estudos de Azzalini⁵ e Rodriguez⁶. Assim, quando λ aproxima-se de zero, a matriz de informação, calculada de forma usual, do modelo de calibração normal-assimétrico, supondo que um experimento com independência foi realizado com $n + k$

⁵*Op. cit.*

⁶*Op. cit.*

elementos, será dada por

$$\begin{bmatrix} \frac{(n+k)}{\sigma^2} & \frac{\sum_{i=1}^n x_i + kx_0}{\sigma^2} & 0 & \frac{(n+k)}{\sigma} \sqrt{\frac{2}{\pi}} & \frac{k\beta}{\sigma^2} \\ \frac{\sum_{i=1}^n x_i + kx_0}{\sigma^2} & \frac{\sum_{i=1}^n x_i^2 + kx_0^2}{\sigma^2} & 0 & \frac{\sum_{i=1}^n x_i + kx_0}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} & \frac{k\beta x_0}{\sigma^2} \\ 0 & 0 & \frac{2(n+k)}{\sigma^2} & 0 & 0 \\ \frac{(n+k)}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} & \frac{\sum_{i=1}^n x_i + kx_0}{\sigma} \sqrt{\frac{2}{\pi}} & 0 & \frac{2(n+k)}{\pi} & \frac{k\beta}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} \\ \frac{k\beta}{\sigma^2} & \frac{k\beta x_0}{\sigma^2} & 0 & \frac{k\beta}{\hat{\sigma}} \sqrt{\frac{2}{\pi}} & \frac{k\beta^2}{\hat{\sigma}^2} \end{bmatrix}$$

que tem a primeira e a terceira linhas, correspondentes aos parâmetros α e λ linearmente dependentes, conseqüentemente, levando a uma matriz de informação singular. Um outro aspecto importante é que quando $k = 0$, ou seja, não há a segunda etapa, essa matriz se reduz a matriz de informação do modelo de regressão; o que faz bastante sentido, (vide, por exemplo, Rodriguez, 2005) e quando β e x_0 são ambos nulos teremos a matriz de informação do modelo de locação-escala (vide, por exemplo, Azzalini, 1985).

Azzalini (1985) propôs uma reparametrização para evitar o problema da singularidade no modelo de locação-escala, $Z \sim SN(\mu; \sigma^2; \lambda)$, considerando

$$Y = \theta_1 + \theta_2(Z - E(Z))/\sqrt{V(Z)},$$

com

$$\theta_1 = \mu + \sigma c \delta;$$

$$\theta_2 = \sigma[1 + \lambda^2(1 - c^2)]^{1/2}(1 + \lambda^2)^{-1/2};$$

e

$$\gamma_1 = \frac{(4 - \pi)}{2}(c\lambda)^3[1 + \lambda^2(1 - c^2)]^{-3/2};$$

onde $c = \sqrt{\frac{2}{\pi}}$, e o parâmetro λ é substituído pelo coeficiente de assimetria γ_1 . Mais tarde, essa parametrização ficou conhecida na literatura como *centralização*, ou *parametrização centrada* como descrito em Azzalini e Capitanio (1999) (vide também Arellano-Valle e Azzalini, 2008). Rotnitzky *et al.* (2000) propuseram um processo

iterativo para encontrar uma reparametrização a fim de evitar a singularidade da matriz de informação de Fisher. Chiogna (2005) deu detalhes desse processo para o modelo normal-assimétrico de locação-escala, que leva a mesma reparametrização proposta por Azzalini (1985). Portanto, podemos estender o processo iterativo de Rotnitzky *et al.* (2000) para o nosso modelo e encontrar a melhor reparametrização para evitar singularidade na vizinhança de $\lambda = 0$.

Proposição 1: A reparametrização $\tilde{\theta} = (\tilde{\alpha} = \alpha + \sigma^* \lambda \sqrt{2/\pi}, \beta, \tilde{\sigma} = \sigma - \sigma^* \lambda^2/\pi, \lambda, x_0)$ tem a terceira derivada em relação a λ do vetor escore $l(\tilde{\theta})$, designada $\tilde{S}_{\lambda\lambda\lambda}$, linearmente independente de quaisquer outras funções escores do modelo de calibração avaliada no ponto $\lambda = 0$.

Demonstração: Suponha que estamos avaliando o vetor escore no ponto $\theta^* = (\alpha^*, \beta^*, \sigma^*, \lambda^* = 0, x_0^*)$. Segue diretamente que a terceira derivada em relação a λ da função escore $l(\tilde{\theta})$ avaliada no ponto $\tilde{\theta} = \theta^*$ é

$$\tilde{S}_{\lambda\lambda\lambda}^* = \sqrt{\frac{2}{\pi}} \left[\left(\frac{4}{\pi} - 1 \right) \left(\sum_i \left(\frac{y_i - \alpha^* - \beta^* x_i}{\sigma^*} \right)^3 + \sum_j \left(\frac{y_{0j} - \alpha^* - \beta^* x_0^*}{\sigma^*} \right)^3 \right) - \frac{6}{\pi} \left(\sum_i \left(\frac{y_i - \alpha^* - \beta^* x_i}{\sigma^*} \right) + \sum_j \left(\frac{y_{0j} - \alpha^* - \beta^* x_0^*}{\sigma^*} \right) \right) \right]$$

que não é combinação linear de

$$\begin{aligned} \tilde{S}_{\tilde{\alpha}}^* &= \sum_i \left(\frac{y_i - \alpha^* - \beta^* x_i}{\sigma^{*2}} \right) + \sum_j \left(\frac{y_{0j} - \alpha^* - \beta^* x_0^*}{\sigma^{*2}} \right); \\ \tilde{S}_{\beta}^* &= \sum_i x_i \left(\frac{y_i - \alpha^* - \beta^* x_i}{\sigma^{*2}} \right) + \sum_j x_0^* \left(\frac{y_{0j} - \alpha^* - \beta^* x_0^*}{\sigma^{*2}} \right); \\ \tilde{S}_{\tilde{\sigma}}^* &= \left(\sum_i \left(\frac{y_i - \alpha^* - \beta^* x_i}{\sigma^*} \right)^2 + \sum_j \left(\frac{y_{0j} - \alpha^* - \beta^* x_0^*}{\sigma^*} \right)^2 - (n+k) \right) / \sigma^*; \text{ e} \\ \tilde{S}_{x_0}^* &= \sum_j \beta^* \left(\frac{y_{0j} - \alpha^* - \beta^* x_0^*}{\sigma^{*2}} \right). \end{aligned}$$

□

Proposição 2: O vetor $\{(n+k)^{1/2}[\hat{\alpha} - (\alpha^* + \sigma^* \hat{\lambda} \sqrt{2/\pi})], (n+k)^{1/2}[\hat{\beta} - \beta^*], (n+k)^{1/2}[\hat{\sigma} - (\sigma^* - \sigma^* \hat{\lambda}^2/\pi)], (n+k)^{1/6} \hat{\lambda}, (n+k)^{1/2}[\hat{x}_0 - x_0^*]\}$ converge em distribuição para $(w_1, w_2, w_3, w_4, w_5)$, quando $n, k \rightarrow \infty$, cuja distribuição é normal multivariada com vetor de médias nulo e matriz de covariâncias

$$\begin{bmatrix} \frac{(n+k)}{\sigma^2} & \frac{\sum_{i=1}^n x_i + kx_0}{\sigma^2} & 0 & \frac{(n+k)}{\sqrt{2\pi\sigma}} \left[\frac{2}{\pi} - 1 \right] & \frac{k\beta}{\sigma^2} \\ \frac{\sum_{i=1}^n x_i + kx_0}{\sigma^2} & \frac{\sum_{i=1}^n x_i^2 + kx_0^2}{\sigma^2} & 0 & \frac{\sum_{i=1}^n x_i + kx_0}{\sqrt{2\pi\sigma}} \left[\frac{2}{\pi} - 1 \right] & \frac{k\beta x_0}{\sigma^2} \\ 0 & 0 & \frac{2(n+k)}{\sigma^2} & 0 & 0 \\ \frac{(n+k)}{\sqrt{2\pi\sigma}} \left[\frac{2}{\pi} - 1 \right] & \frac{\sum_{i=1}^n x_i + kx_0}{\sqrt{2\pi\sigma}} \left[\frac{2}{\pi} - 1 \right] & 0 & \frac{(n+k)}{6\pi} \left[\frac{44}{\pi^2} - \frac{28}{\pi} + 5 \right] & \frac{k\beta}{\sqrt{2\pi\sigma}} \left[\frac{2}{\pi} - 1 \right] \\ \frac{k\beta}{\sigma^2} & \frac{k\beta x_0}{\sigma^2} & 0 & \frac{k\beta}{\sqrt{2\pi\sigma}} \left[\frac{2}{\pi} - 1 \right] & \frac{k\beta^2}{\hat{\sigma}^2} \end{bmatrix}^{-1}.$$

Demonstração: Suponha novamente que estamos avaliando o vetor escore no ponto $\theta^* = (\alpha^*, \beta^*, \sigma^*, \lambda^* = 0, x_0^*)$. Aqui, precisamos fazer uso do Teorema 3 de Rotnitzky *et al.* (2000) que afirma que sob certas condições de regularidade, quando a ordem de derivadas, designada s , do escore de algum parâmetro, θ , que torna a matriz de informação singular, quando ele tende a θ^* , é ímpar e essa derivada do escore avaliada numa vizinhança de θ^* é diferente de zero, então

- (a) O estimador de máxima verossimilhança $\hat{\theta}$ de θ é único e é consistente, quando θ tende a θ^* .
- (b) A distribuição do vetor $\{(n)^{1/2s}[\hat{\theta}_1 - \theta_1^*], \dots, (n)^{1/2}[\hat{\theta}_p - \theta_p^*]\}^\top$ converge em distribuição para $(w_1^{1/s}, \dots, w_p)^\top$, cuja distribuição é normal multivariada com vetor de médias nulo e matriz de covariâncias igual a I^{-1} , onde I é a matriz de covariâncias de $(S_1^{(s)}/s!, S_2, \dots, S_p)^\top$.
- (c) A estatística do teste da razão de verossimilhanças, $2\{\ell(\hat{\theta}) - \ell(\theta^*)\}$, converge em distribuição para χ_p^2 .

Note que este resultado ocorrerá sob certas condições de regularidade e, para nosso modelo, precisamos checá-las. Assim, note que a reparametrização $\tilde{\theta}$, avaliada no ponto θ^* , torna-se θ^* . Além disso, em $\theta = \theta^*$, $\tilde{S}_\alpha^* = S_\alpha^*$, $\tilde{S}_\beta^* = S_\beta^*$, $\tilde{S}_\sigma^* = S_\sigma^*$, e $\tilde{S}_{x_0}^* = S_{x_0}^*$. Note também que $\tilde{S}_\lambda^* = 0$, $\tilde{S}_{\lambda\lambda}^* = 0$, e $\tilde{S}_{\lambda\lambda\lambda}^*$ não se anula nem é combinação linear de outros escores como suposto pela Proposição 1. Como estas duas condições estão satisfeitas, o resultado do Teorema 3 de Rotnitzky *et al.* (2000) está também satisfeito e nossa proposição segue diretamente após vários cálculos algébricos para encontrar a matriz de covariâncias de $I = (S_1^{(s)}/s!, S_2, \dots, S_p)^\top$. \square

Com a Proposição 2 derivamos uma matriz não singular quando λ se aproxima de zero. É evidente que as estimativas de variâncias obtidas desta matriz de informação transformada não podem ser utilizadas para fazermos inferências sobre os parâmetros α e σ , mas essa matriz pode ser assintoticamente utilizada em inferências sobre x_0 que não necessitou ser transformado, além de termos condições de realizar um teste para checarmos se λ é ou não nulo, já que dispomos da estatística de teste. Sempre que possível é interessante calcular as estimativas das variâncias dos parâmetros do modelo usual com erros normalmente distribuídos para podermos comparar as amplitudes dos intervalos nas duas situações. A matriz de informação de Fisher para o vetor de parâmetros, $\theta = (\alpha, \beta, \sigma, x_0)$, considerando o modelo normal usual com erros normalmente distribuídos, é dada por

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} n+k & kx_0 + \sum_{i=1}^n x_i & 0 & k\beta \\ kx_0 + \sum_{i=1}^n x_i & kx_0^2 + \sum_{i=1}^n x_i^2 & 0 & kx_0\beta \\ 0 & 0 & 2(n+k) & 0 \\ k\beta & kx_0\beta & 0 & k\beta^2 \end{bmatrix}.$$

Como o estimador de máxima verossimilhança $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{x}_0)$ tem distribuição aproximadamente normal com média θ e matriz de covariância $I(\theta)^{-1}$, conseguimos encontrar estimativas para as variâncias dos parâmetros do modelo usual com er-

ros normalmente distribuídos e comparar com as estimativas do modelo normal-assimétrico.

1.7 Prevenção de viés

O estimador de máxima verossimilhança de $\hat{\lambda}$ pode tender a infinito com probabilidade positiva como apresentou Sartori (2003) que implementou o método de Firth (Firth, 1993) para a prevenção de viés no caso do modelo de locação-escala normal assimétrico e ressaltou que a existência do estimador de máxima verossimilhança para λ não está garantida por questões teóricas. Sartori (2003) propôs uma abordagem baseada em quase-verossimilhança associada a uma modificação da função escore na construção de intervalos de confiança.

No modelo de calibração, o vício, $b(\lambda)$, pode ser obtido da mesma maneira que no modelo de locação-escala normal assimétrico, tal que sua ordem seja $O(n^{-1})$. Temos que

$$b(\lambda) = -\frac{1}{2}I_{\lambda}^{-2}(\lambda) \{E_Z [S^3(\lambda)] - E_Z [S(\lambda)S'(\lambda)]\} \quad (1.9)$$

já que $S(\lambda)$ é a função escore associada a λ e o índice Z especifica a natureza normal-assimétrica da distribuição, como na Seção anterior.

Sartori (2003) mostrou como evitar a super-estimação de λ no modelo de locação-escala, adicionando uma modificação, $M(\lambda)$, à função escore levando a $S^*(\lambda) = S(\lambda) + M(\lambda)$, com

$$M(\lambda) = -I_{\lambda}(\lambda)b(\lambda).$$

Para o cálculo de $M(\lambda)$, precisamos das seguintes expressões,

$$\begin{aligned} I_{\lambda}(\lambda) &= (n+k)a_2; \\ E_Z [S^3(\lambda)] &= (n+k)E_Z \left[Z^3 \frac{\phi^3(\lambda Z)}{\Phi^3(\lambda Z)} \right]; \end{aligned}$$

e

$$E_Z [S(\lambda)S'(\lambda)] = (n+k) \left\{ \lambda E_Z \left[Z^4 \frac{\phi^2(\lambda Z)}{\Phi^2(\lambda Z)} \right] + E_Z \left[Z^3 \frac{\phi^3(\lambda Z)}{\Phi^3(\lambda Z)} \right] \right\},$$

onde a_2 está definido no Apêndice, que levará a uma expressão que não depende do primeiro e segundo passos da calibração. Essa expressão é dada por

$$M(\lambda) = \frac{\lambda}{2} \frac{E_Z \left[Z^4 \frac{\phi^2(\lambda Z)}{\Phi^2(\lambda Z)} \right]}{a_2} \quad (1.10)$$

que pode ser substituída pela versão observada, $M^{obs}(\lambda)$, obtida através da substituição de $I_\lambda(\lambda)$ e $b(\lambda)$ em (1.9) pelas versões observadas $S^2(\hat{\lambda})$ e $-1/2[S^3(\hat{\lambda}) - S^2(\hat{\lambda})]$, respectivamente. A versão observada da modificação apresentou vários problemas de instabilidade numérica, que nem sempre garante a existência de estimativas, como discutiu Sartori (2003).

1.8 Aplicação

O conjunto de dados que apresentaremos a seguir foi extraído de Chipkevitch *et al.* (1996) que obteve medidas do volume testicular em 42 adolescentes, em mililitros cúbicos. Medir o volume testicular é um caminho bastante oportuno para estimarmos a espermatogênese na puberdade e o menos invasivo. Os autores supra-citados usaram várias ferramentas de medida e as compararam. Para nossos propósitos, apenas duas dessas medidas serão utilizadas: o dimensionamento do volume testicular obtido com um calibrador, que é o modo mais simples de se fazer um dimensionamento clínico e, por outro lado, o mais preciso dentre os utilizados pelos autores: a ultra-sonografia. Uma vez calibrado, o método de dimensionamento com o calibrador pode ser um melhor caminho de obtenção do volume testicular. A Figura 1.1(a) mostra o histograma de medições obtidas via ultra-sonografia do testículo direito de 42 adolescentes apenas para ilustrar a notável assimetria presente nesse tipo

de medição. Em Chipkevitch *et al.*⁷ os dados foram transformados tomando-se a raiz cúbica do volume para que a normalidade fosse assegurada. Claramente, nós evitaremos a transformação se utilizarmos a metodologia proposta. A Figura 1.1(b) tenta ilustrar a possível relação linear existente entre essas duas ferramentas de medida. O histograma dos resíduos de um modelo usual normal, assumindo a relação descrita na Figura 1.1(b), aparece na Figura 1.1(c) e claramente mostra a existência de assimetria à direita. Os dados estão apresentados na Tabela 1.3.

O coeficiente de assimetria dos resíduos foi também calculado, utilizando $\tilde{\lambda} = (n/(n-1)(n-2)) \sum((y - \bar{y})/DP(y))^3$ que resultou em $\tilde{\lambda} = 0,32$. Como apresentado na Figura 1.1(c), bem como utilizando o coeficiente de assimetria estimado, somos levados a considerar que a distribuição dos erros não é simétrica. Para termos alguma evidência estatística, utilizamos o teste não-paramétrico de Lilliefors para normalidade; vide, por exemplo, Conover (1999). O resultado do teste trouxe evidências para a rejeição da suposição de normalidade, já que a estatística de Lilliefors (0,8391) foi maior do que a esperada sob normalidade (0,1356).

Primeiramente, nós olhamos para um dado cujo volume tivesse sido observado mais de uma vez por meio da inspeção via ultra-sonografia. Encontramos o valor 16,4. Esses valores aparecem destacados na Tabela 1.3, em negrito. Na seqüência, consideramos o valor 16,4 como um valor desconhecido para essas duas observações. Nosso interesse está em estimar o valor $x_0 = 16,4$, que foi considerado desconhecido com os correspondentes valores de y_0 : $y_{01} = 10,3$ e $y_{02} = 17,3$. Os outros valores foram utilizados na primeira etapa do experimento. A estimação foi realizada através do algoritmo EM. As estimativas dos parâmetros do modelo aparecem na Tabela 1.4. Nós calculamos as estimativas dos parâmetros com duas especificações: uma normal-assimétrico e outra normal.

O algoritmo EM foi implementado na linguagem R. A fim de avaliar estes

⁷*Op. cit.*

Figura 1.1: Análise descritiva

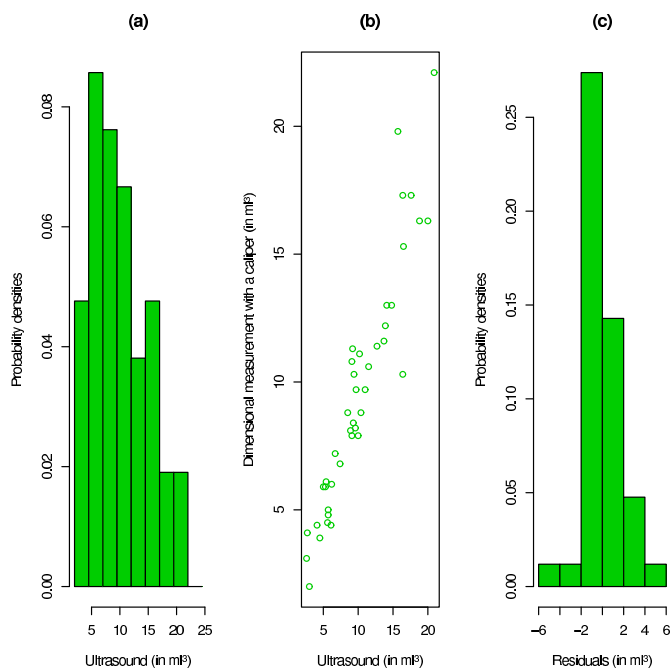


Tabela 1.3: Medições obtidas por dimensionamento (DM) por meio do calibrador e ultra-sonografia (US) do testículo direito de 42 adolescentes, ambas em mililitros cúbicos.

DM	US	DM	US	DM	US
5,9	5	17,3	16,4	7,2	6,7
4,8	5,7	17,3	17,6	5,9	5,3
6,8	7,4	7,9	10	16,3	20
3,1	2,6	4,4	4,1	16,3	18,8
5	5,7	11,4	12,7	12,2	13,9
4,4	6,1	4,1	2,7	10,3	9,4
6	6,2	11,1	10,2	10,8	9,1
8,8	10,4	15,3	16,5	13	14,1
7,9	9,1	3,9	4,5	8,4	9,3
13	14,8	4,5	5,6	22,1	20,9
10,3	16,4	9,7	11	10,6	11,5
8,2	9,6	11,3	9,2	9,7	9,7
19,8	15,7	8,8	8,5	11,6	13,7
2	3	6,1	5,4	8,1	8,9

Tabela 1.4: Estimativas e os desvios padrões delas.

Parâmetro	Modelo normal-assimétrico		Modelo normal	
	Estimativa	Desvio padrão	Estimativa	Desvio padrão
α	-0,69	-	0,32	0,56
β	0,86	0,07	0,92	0,05
σ	2,13	-	1,55	0,17
x_0	12,66	1,81	14,58	1,24
λ	2,16	1,73	-	-

dois modelos, utilizamos três critérios. Calculamos o critério de informação de Akaike ($AIC = -2\ell(\hat{\theta}) + 2p$), conhecido na literatura estatística por AIC (*Akaike Information Criterion*); suas iniciais em inglês. Calculamos também o critério de informação de Schwarz, ou também conhecido como critério de informação bayesiana ($BIC = -2\ell(\hat{\theta}) + p\log(N)$), portanto, conhecido como BIC, das iniciais em inglês (*Bayesian Information Criterion*). Por último, o critério de Hannan-Quinn ($HQ = -2\ell(\hat{\theta}) + 2\log(p\log(N))$), menos utilizado que os anteriores. Em todos eles, p e $\hat{\theta}$ são o número de parâmetros e o vetor das estimativas dos parâmetros sob o modelo em questão, respectivamente, e N é o número de observações nos dois passos do processo de calibração, ou melhor, $N = n + k$. A Tabela 1.5 mostra os valores destes critérios.

Tabela 1.5: Critérios de Informação.

Critério	Modelo normal-assimétrico	Modelo normal usual
	Valor	Valor
AIC	160,69	163,74
BIC	169,38	170,69
HQ	156,55	161,15

Assim, analisando estes três critérios de informação, encontramos que o modelo de calibração normal assimétrico superou o modelo usual normal, já que AIC , BIC e HQ foram sempre menores para o modelo normal-assimétrico quando comparado com o modelo de calibração normal usual. Este fato mostra que nossas suspeitas de possível assimetria presente nos dados é verdadeira e a distribuição normal-

assimétrica apresentou um melhor ajuste.

Enfoque bayesiano do modelo de calibração normal-assimétrico

2.1 Introdução

Neste capítulo desenvolvemos o modelo de calibração linear normal-assimétrico sob a ótica bayesiana e reescrevemos o modelo segundo uma representação hierárquica necessária para a implementação de um algoritmo do tipo MCMC (Markov Chain Monte Carlo), mais precisamente o amostrador de Gibbs. Esta representação hierárquica que já foi apresentada no capítulo anterior, será outra vez apresentada aqui para fins de organização. Além disso, aplicamos o modelo estudado a um conjunto de dados reais. Este Capítulo se encontra publicado no *Journal of Chemometrics*, (vide Figueiredo *et al.*, 2008).

O estudo de calibração segundo a abordagem bayesiana data de algum tempo. Hoadley (1970) foi o precursor no estudo do problema da calibração controlada sob a abordagem bayesiana. Um resumo dos principais resultados para o modelo de calibração pode ser encontrado em Brown (1993). Branco (1997) realizou um estudo sobre calibração linear sob a ótica bayesiana, levando em consideração dois cenários (vide também Branco *et al.*, 2000). Numa primeira especificação, Branco (1997) considerou

que o vetor de erros, ε , era proveniente de uma distribuição elíptica dependente e, numa segunda especificação, Branco *et al.* (2000) consideraram que o vetor de erros, ε , era proveniente de um modelo elíptico independente. Recentemente, Arellano-Valle e Branco (2004) publicaram um estudo sobre distribuições elípticas assimétricas. Lachos (2004) e Rodriguez (2005) estudaram modelos lineares mistos com distribuição dos erros assimétrica e inferência bayesiana em modelos normais-assimétricos, respectivamente. Além disso, Rodriguez (2005) propôs uma distribuição *a priori* própria t de *Student* para λ , fundamental na construção de um modelo hierárquico para estimação dos parâmetros do modelo de regressão linear. Uma outra distribuição *a priori* para λ que utilizaremos no trabalho é a distribuição *a priori* de referência proposta por Jeffreys e desenvolvida por Liseo e Loperfido (2006) para o modelo normal-assimétrico.

Do mesmo modo que no Capítulo anterior, consideraremos que o vetor de erros ε segue uma distribuição normal-assimétrica como foi considerada em (1.3). Conseqüentemente, podemos reescrever o modelo de calibração normal-assimétrico de acordo com a representação estocástica de Henze (1986) que facilita a implementação de um algoritmo do tipo MCMC, onde o nosso objetivo é implementar o amostrador de Gibbs. Esta representação foi apresentada inicialmente por Andel *et al.* (1984)¹ e mais adiante estendida por Henze (1986). Bazan *et al.* (2006) estudaram o modelo probito-assimétrico e sugeriram que a utilização da parametrização *delta*, ao invés da parametrização *lambda*, seria mais eficiente que a anterior na estimação dos parâmetros. Para chegar a essa conclusão, eles utilizaram a noção da razão das médias do tamanho amostral efetivamente gerado por segundo, designado por $(ESS/s1/ESS/s2)$, para comparar duas especificações de um mesmo modelo. Quanto maior esta medida, mais eficiente seria a primeira especificação quando com-

¹ANDEL, D. J., NETUKA, I. e ZVARA, K *apud* AZZALLINI, A. A class of distributions which includes the normal one. *Scandinavian Journal of Statistics*, 12, p. 171-178. 1985.

parada com a segunda. Por esta razão, não consideraremos a parametrização λ no nosso estudo, mas discutiremos alguns aspectos dela na aplicação que será apresentada na Seção 3.

Desse modo, neste capítulo, apresentamos a abordagem bayesiana para a estimação dos parâmetros do modelo de calibração normal-assimétrica, bem como as distribuições *a posteriori* condicionais que serão úteis para a implementação do amostrador de Gibbs. Desta forma, teremos mais uma metodologia para evitarmos a transformação de dados através do uso da metodologia proposta. Nossa proposta é utilizar uma nova medida para a aferição da qualidade do ajuste que é apenas uma correção do tão utilizado critério de informação da desviância, conhecido como DIC, das iniciais em inglês: *Deviance Information Criterion*. A proposta é utilizar o critério de informação da desviância assimétrica, que chamaremos de ADIC, devido às iniciais do nome em inglês (*Asymmetric Deviance Information Criterion*) apenas uma outra forma de calcular o DIC no caso assimétrico. No final do capítulo, apresentamos uma aplicação da metodologia proposta utilizando um problema de engenharia que apresenta a necessidade real de calibração, onde se estuda a relação entre a resistência e a elasticidade de vigas de concreto.

2.2 O modelo hierárquico e o amostrador de Gibbs

Como relatado por diversos autores na literatura (vide Azzalini e Capitanio, 1999), existem situações em que a modelagem normal parece inadequada devido a quantidade de assimetria presente nos dados. Esta assimetria pode afetar modelos de regressão, ou generalizando, os modelos de calibração. Neste contexto, o modelo de calibração, escrito na parametrização λ , definido pelas equações,

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, n;$$

$$y_{0j} = \alpha + \beta x_0 + \varepsilon_{0j}, \quad j = 1, \dots, k,$$

pode ser mais apropriado em algumas situações, onde ε_i e ε_{0j} são independentes e identicamente distribuídos segundo uma $SN(0; \sigma^2; \lambda)$, onde 0 , σ^2 e λ são, respectivamente, os parâmetros de locação, escala e forma.

O modelo de calibração acima pode ser reescrito utilizando a representação estocástica considerada por Henze (1986) e utilizando a parametrização *delta*, como

$$y_i = \alpha + \beta x_i + \sigma \delta t_i + r_i$$

$$y_{0j} = \alpha + \beta x_0 + \sigma \delta t_{0j} + r_{0j}$$

onde $t_i = |U_i|$, $t_{0j} = |U_{0j}|$, $r_i = \sigma(1 - \delta^2)^{1/2}V_i$ e $r_{0j} = \sigma(1 - \delta^2)^{1/2}V_{0j}$ com $i = 1, \dots, n$ e $j = 1, \dots, k$ com U e V independentes e distribuídos segundo normais-padrões.

Como conseqüências da representação acima, temos $r_i \sim N(0; \sigma^2(1 - \delta^2))$ e t_i , bem como t_{0j} , seguindo uma distribuição meio-normal. Nosso interesse está na estimação do vetor $\theta = (\alpha, \beta, \sigma, \lambda, x_0)$. Então, condicionando em t_i e t_{0j} e utilizando a reparametrização $b = \sigma\delta$ e $\sigma_*^2 = \sigma^2(1 - \delta^2)$, teremos a seguinte representação hierárquica do modelo de calibração:

$$y_i | t_i; x_i \sim N(\alpha + \beta x_i + b t_i; \sigma_*^2), \quad (2.1)$$

$$t_i \sim HN(0; 1), \quad i = 1, \dots, n;$$

$$y_{0j} | t_{0j} \sim N(\alpha + \beta x_0 + b t_{0j}; \sigma_*^2), \quad (2.2)$$

$$t_{0j} \sim HN(0; 1), \quad j = 1, \dots, k.$$

Portanto, a função de verossimilhança completa, com $\theta^* = (\alpha, \beta, \sigma_*^2, b, x_0)$, é dada por (1.8).

2.2.1 O amostrador de Gibbs

Para a implementação do algoritmo de Gibbs é necessário que seja realizado um estudo de sensibilidade sobre as performances das distribuições *a priori* que serão

utilizadas. Vamos nos referir à distribuição *a priori* apenas por *priori*, a partir de agora. O mesmo será válido para o termo latino *a posteriori*, que chamaremos apenas *posteriori*, quando o entendimento do texto não for prejudicado. Mais que isto, os termos também serão aportuguesados, para evitar a necessidade de enfatizar as expressões quando forem citadas, ou seja, tratá-las-emos também por distribuições à *priori* e à *posteriori*, a partir de agora.

Rodriguez (2005) conduziu um estudo semelhante a este para modelos de regressão. Podemos dizer que este trabalho está apoiado grande parte nos estudos realizados por Rodriguez (2005) e é uma extensão para modelos de calibração. Aliás, uma melhor referência para o trabalho de Rodriguez (2005) pode ser obtida em Bayes e Branco (2007), por estar mais atualizada e revista (vide também o relatório que foi gerado desse trabalho, Bayes e Branco, 2005). Lá, Rodriguez (2005), defendeu o uso da distribuição à *priori* para λ definida por

$$\pi(\lambda) \propto \left(1 + \frac{\lambda^2}{k\sigma_t^2}\right)^{-\frac{k+1}{2}}, \quad (2.3)$$

onde σ_t^2 e k são os parâmetros de escala e os graus de liberdade devido ao uso da distribuição *t* de *Student*, respectivamente.

Nesse estudo² a *priori* (2.3) foi reescrita em forma hierárquica, resultando em

$$\begin{aligned} \lambda|w &\sim N\left(0; \frac{\sigma_t^2}{w}\right) \\ w &\sim \text{Gama}\left(\frac{k}{2}; \frac{k}{2}\right). \end{aligned} \quad (2.4)$$

Este resultado segue se acreditarmos que δ segue uma distribuição uniforme ($U[-1; 1]$) e através de transformação, nós poderemos ver, diretamente, que λ terá distribuição *t* de *Student*, especificada por, $\lambda \sim t(0; 1/2; 2)$. Mais que isto, podemos fazer uso de uma *priori* de Jeffreys para λ , que foi concebida em Liseo e Loperfido (2006) para o caso do modelo de locação-escala normal assimétrico. Neste caso, a

²*Op. cit.*

priori de Jeffreys, $\pi^J(\lambda)$, é dada por:

$$\pi^J(\lambda) \propto \sqrt{\int_{-\infty}^{\infty} 2z^2 \phi(z) \frac{\phi(\lambda z)}{\Phi(\lambda z)} dz}$$

Chaibub Neto e Branco (2003) desenvolveram após longas manipulações algébricas uma aproximação para a quantidade acima que aparece designada por $\pi^J(\lambda)$, sendo esta quantidade dada por

$$\frac{1}{\pi} \phi(z) / \sqrt{\Phi(z)[1 - \Phi(z)]} \approx \frac{2}{\pi} \phi\left(\frac{2}{\pi}z\right).$$

Pode ser mostrado que esta aproximação leva à distribuição à priori especificada por $\pi^J(\lambda) \sim t(0; \pi^2/4; 1/2)$, como foi mostrado em Rodriguez (2005).

Como já foi dito na introdução do Capítulo, Hoagly (1970), foi o precursor do estudo do problema de calibração sob a abordagem bayesiana e propôs uma distribuição à priori t de *Student* para o parâmetro x_0 que conduzia a uma distribuição à posteriori também t de *Student*, onde o estimador bayesiano sob a função de perda quadrática era dado pelo estimador inverso, apresentado no Lema 2 do Capítulo anterior (vide Branco *et al.*, 2000). Essa priori, designada $t(\bar{x}; (S_x^2) * \frac{n+1}{n-3}; n-3)$ será utilizada aqui como uma priori para x_0 , desde que $n > 3$. Na parametrização *delta*, o parâmetro b funcionará como o indicador de assimetria. Então, utilizaremos as mesmas idéias de especificações à priori para o parâmetro λ , como em Rodriguez (2005), e, conseqüentemente, para a especificação à priori relacionada ao parâmetro b .

Como um resultado geral, temos o

Teorema 1: Assumindo distribuições à priori independentes entre os parâmetros do modelo de calibração linear, ou seja, fazendo, então, suposições de que $\alpha \sim N(\mu_\alpha; \sigma_\alpha^2)$, $\beta \sim N(\mu_\beta; \sigma_\beta^2)$, $x_0 \sim N(\mu_{x_0}; \sigma_{x_0}^2)$, $1/\sigma_*^2 \sim Gama(A; B)$ e $b \sim t(0; \sigma_t^2; g)$, onde a priori para b pode ser obtida através da mistura $b|w \sim N(0; \frac{\sigma_t^2}{w})$ e $w \sim Gama(\frac{g}{2}; \frac{g}{2})$, teremos a distribuição à posteriori dada por

$$\pi(\theta^* = (\alpha, \beta, x_0, b, \sigma_*^2) | y_i; y_{0j}; x_i) \propto \pi(\theta^*) \times L(\theta^*; z)$$

que conduz as seguintes distribuições à posteriori condicionais próprias:

- (a) $\pi(\alpha|\beta; b; \sigma_*^2; w; x_0; t_{0j}; t_i; y_i; y_{0j}; x_i)$
 $\sim N\left(\frac{\sigma_\alpha^2[\sum(y_i - \beta x_i - bt_i) + \sum(y_{0j} - \beta x_0 - bt_{0j})] + \sigma_*^2 \mu_\alpha}{\sigma_\alpha^2(n+k) + \sigma_*^2}; \frac{\sigma_\alpha^2 \sigma_*^2}{\sigma_\alpha^2(n+k) + \sigma_*^2}\right);$
- (b) $\pi(\beta|\alpha; b; \sigma_*^2; w; x_0; t_{0j}; t_i; y_i; y_{0j}; x_i)$
 $\sim N\left(\frac{\sigma_\beta^2[\sum x_i(y_i - \alpha - bt_i) + \sum x_0(y_{0j} - \alpha - bt_{0j})] + \sigma_*^2 \mu_\beta}{\sigma_\beta^2(\sum x_i^2 + kx_0) + \sigma_*^2}; \frac{\sigma_\beta^2 \sigma_*^2}{\sigma_\beta^2(\sum x_i^2 + kx_0) + \sigma_*^2}\right);$
- (c) $\pi(\frac{1}{\sigma_*^2}|\alpha; \beta; b; w; x_0; t_{0j}; t_i; y_i; y_{0j}; x_i)$
 $\sim \text{Gama}\left(\frac{n+k}{2} + A; \frac{1}{2}(\sum(y_i - \alpha - \beta x_i - bt_i) + \sum(y_{0j} - \alpha - \beta x_0 - bt_{0j}) + \frac{b^2 w}{\sigma_t^2}) + B\right);$
- (d) $\pi(x_0|\alpha; \beta; b; \sigma_*^2; w; t_i; t_{0j}; y_i; y_{0j}; x_i)$
 $\sim N\left(\frac{\sigma_{x_0}^2 \beta \sum(y_{0j} - \alpha - bt_{0j}) + \sigma_*^2 \mu_{x_0}}{\sigma_{x_0}^2(k\beta^2) + \sigma_*^2}; \frac{\sigma_{x_0}^2 \sigma_*^2}{\sigma_{x_0}^2(k\beta^2) + \sigma_*^2}\right);$
- (e) $\pi(b|\alpha; \beta; \sigma_*^2; w; x_0; t_{0j}; t_i; y_i; y_{0j}; x_i)$
 $\sim N\left(\frac{\sum(y_i - \alpha - \beta x_i)t_i + \sum(y_{0j} - \alpha - \beta x_0)t_{0j}}{\sum t_i^2 + \sum t_{0j}^2 + \frac{w}{\sigma_t^2}}; \frac{\sigma_*^2}{\sum t_i^2 + \sum t_{0j}^2 + \frac{w}{\sigma_t^2}}\right); e$
- (f) $\pi(w|\alpha; \beta; b; \sigma_*^2; x_0; t_i; t_{0j}; y_i; y_{0j}; x_i)$
 $\sim \text{Gama}\left(\frac{g+1}{2}; \frac{1}{2}\left(\frac{b^2}{\sigma_*^2 \sigma_t^2} - g\right)\right).$

Demonstração O resultado segue após direta e exaustiva manipulação algébrica, trabalhando diretamente com a função de verossimilhança completa. Sabemos que $f(\theta^*, z) = f(\theta^*|z) \times f(z)$ pode ser escrito como $f(\theta^*, z) = f(z|\theta^*) \times f(\theta^*)$, onde $f(z|\theta^*)$ é a função de verossimilhança e $f(\theta^*)$ a distribuição à priori. No entanto, $f(\theta^*|z) = \frac{f(z|\theta^*) \times f(\theta^*)}{f(z)} \propto L(\theta^*|z) \times f(\theta^*)$. Desse modo obtemos a distribuição condicional de um particular parâmetro, θ_i^* , condicionando em relação às outras variáveis e deixando θ_i^* livre. \square

Corolário: Sob as mesmas condições anteriores, as distribuições à posteriori condicionais relacionadas a t_i e t_{0j} e que foram definidas em (2.1) e (2.2) são dadas por

$$(i) \pi(t_i | \alpha; \beta; b; \sigma_*^2; w; x_0; t_{0j}; y_i; y_{0j}; x_i) \\ \sim HN\left(\frac{(y_i - \alpha - \beta x_i)b}{\sigma_*^2 + b^2}; \frac{\sigma_*^2}{\sigma_*^2 + b^2}\right), \forall i, \quad i = 1, \dots, n; \text{ e}$$

$$(ii) \pi(t_{0j} | \alpha; \beta; b; \sigma_*^2; w; x_0; t_i; y_i; y_{0j}; x_i) \\ \sim HN\left(\frac{(y_{0j} - \alpha - \beta x_0)b}{\sigma_*^2 + b^2}; \frac{\sigma_*^2}{\sigma_*^2 + b^2}\right), \forall j, \quad j = 1, \dots, k.$$

Demonstração: Segue após direta manipulação algébrica, como anteriormente. \square

Com o que foi exposto acima, temos as condições necessárias para garantir a convergência do amostrador de Gibbs e este pode ser implementado do seguinte modo:

1. Atribua valores iniciais para o vetor $\theta^{*(0)} = (\alpha^{(0)}, \beta^{(0)}, b^{(0)}, 1/\sigma_*^{2(0)}, x_0^{(0)},)$;
2. Gere $w^{(j)}, t_i^{(j)}, t_{0j}^{(j)}$, utilizando os valores iniciais fixados;
3. Atualize $\theta^{*(j)}$ através de suas respectivas distribuições;
4. Repita os passos 2 e 3 para $j = 1, \dots, N$; e
5. Descarte os B primeiros valores ($B \leq N$).

Os $N - B$ valores obtidos serão utilizados para a estimação dos parâmetros de interesse.

2.3 Aplicação

O conjunto de dados que analisaremos a seguir foi retirado de Devore (2006)³. O valor absoluto da elasticidade ($ELA = y$), medida em GPa , e a resistência à flexão

³Effects of aggregates and microfillers on the flexural properties of concrete. *Magazine of Concrete Research*, p. 81-98, 1997, *apud* DEVORE, J. L. *Probabilidade e estatística para engenharia e ciências*. São Paulo: Thomson Learning, 2006.

($RES = x$), cuja medida é obtida da capacidade de resistência a falhas decorrentes de desdobramentos, em MPa , foi observada para uma amostra de 27 vigas de concreto (vide Figura 2.1(a)). Os dados estão apresentados na Tabela 2.1.

Figura 2.1: Análise Descritiva: Resíduos do modelo linear normal

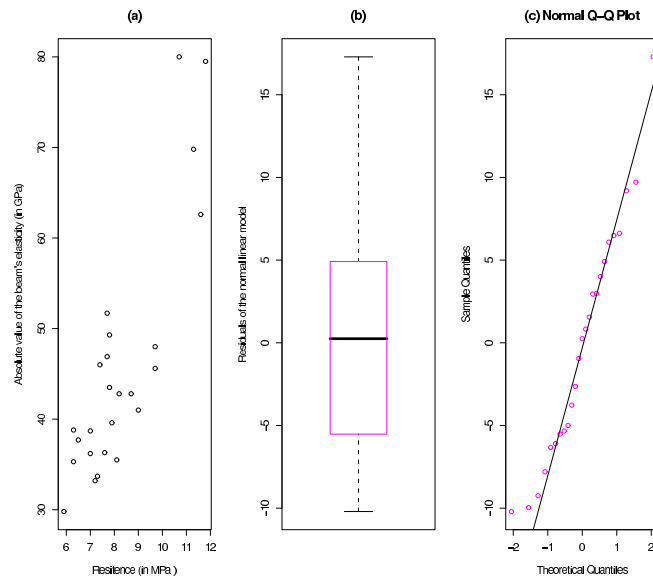


Tabela 2.1: Elasticidade (ELA) e resistência (RES) das vigas de concreto

ELA (em GPa)	RES (em MPa)	ELA	RES	ELA	RES
29,8	5,9	37,7	6,5	46	7,4
33,2	7,2	38,7	7	46,9	7,7
33,7	7,3	38,8	6,3	48	9,7
35,3	6,3	39,6	7,9	49,3	7,8
35,5	8,1	41	9	51,7	7,7
36,1	6,8	42,8	8,2	62,6	11,6
36,2	7	42,8	8,7	69,8	11,3
36,3	7,6	43,5	7,8	79,5	11,8
37,5	6,8	45,6	9,7	80	10,7

Primeiramente, procuramos por uma observação cujo o valor de RES tivesse sido observado mais de uma vez. Encontramos a observação 6,8. Os pares em que a observação aparece estão destacados, em negrito, na Tabela 2.1. A seguir,

consideramos o valor 6,8 como um valor desconhecido para esses dois pares. Nosso interesse será estimar o valor $x_0 = 6,8$, hipoteticamente considerado desconhecido e que está associado aos valores, que denotaremos por y_0 , $y_{01} = 36,1$ e $y_{02} = 37,5$. Na nossa análise descritiva, utilizamos a linguagem R⁴. Uma análise preliminar deste conjunto de dados (vide Figura 2.1(b) e 2.1(c)) mostrou que os resíduos do modelo linear de regressão normal eram, descritivamente, um pouco assimétricos à direita, i. e., há a indicação de uma sucinta cauda prolongada à direita, como mostra o *box-plot*, bem como podemos notar uma pequena tendência de existir mais observações sobre a reta bissetriz do gráfico de quantis. Também estimamos a assimetria, calculando $\tilde{\lambda} = (n/(n-1)(n-2)) \sum ((e_i - \bar{e})/DP(e))^3$, com e_i denotando o *i-ésimo* resíduo, \bar{e} sua média e $DP(e)$ seu desvio padrão, respectivamente. Esta medida resultou em $\tilde{\lambda} = 0,46$. Como ilustra a Figura 2.1, bem como, utilizando o coeficiente de assimetria estimado, somos conduzidos a considerar que a distribuição dos erros do modelo de calibração não pode ser simétrica. Naturalmente, procuraremos evitar o uso de uma transformação utilizando a metodologia proposta.

Como vem a seguir, utilizamos o amostrador de Gibbs para calcular as estimativas dos parâmetros do modelo. Nós utilizamos e implementamos o modelo com o *software WinBUGS*⁵, já que o *WinBUGS* pode ser utilizado para simular distribuições à posteriori e, hoje em dia, é vastamente utilizado por pesquisadores que implementam algoritmos MCMC.

Nós ajustamos sete modelos através do WinBUGS. Em todos eles, as distribuições à priori para α , β e σ^2 , consideradas no Teorema 1, foram especificadas por: $\alpha \sim N(\mu_\alpha; \sigma_\alpha^2)$, $\beta \sim N(\mu_\beta; \sigma_\beta^2)$, $1/\sigma_*^2 \sim Gama(A; B)$, com os hiperparâmetros

⁴VENABLES, W. N. , SMITH, D. M. *An Introduction to R. R: A Programming Environment for Data Analysis and Graphics*. Versão 2.3.1. 2006. Disponível em: << <http://cran.r-project.org/doc/manuals/R-intro.pdf> >>. Acesso em: 15 de setembro de 2006.

⁵WinBUGS. Version 1.4.1. Disponível em: << <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml> >>. Acesso em: 15 de março de 2007.

todos conhecidos. A escolha destas distribuições foram baseadas nas especificações usuais da teoria de regressão linear⁶. As diferenças entre os sete modelos aparecem nas especificações das prioris de x_0 e b , com exceção dos dois primeiros modelos que apenas foram implementados para visualizarmos as estimativas com a parametrização λ .

Utilizamos os seguintes valores para os hiperparâmetros: $\mu_\alpha = \bar{y} - \bar{x} * \mu_\beta$; $\mu_\beta = \sum(x_i - \bar{x}) * (y_i - \bar{y}) / \sum(x_i - \bar{x})^2$; $\mu_{x_0} = (\bar{y}_0 - \mu_\alpha) / \mu_\beta$; $\sigma_\alpha^2 = QMRes * (n^{-1} + (\bar{x})^2 / \sum(x_i - \bar{x})^2)$; $\sigma_\beta^2 = QMRes / \sum(x_i - \bar{x})^2$; $\sigma_{x_0}^2 = \sum(x_i - \bar{x})^2 / (n - 1)$, onde $QMRes$ simboliza o quadrado médio do resíduo⁷; $A = 0.05$; e $B = 3$. Estes dois últimos valores obtidos via método dos momentos.

As especificações à priori para os parâmetros x_0 e b são tais que:

MODELO I: $x_0 \sim N(\mu_{x_0}; \sigma_{x_0}^2)$ e $b \sim t(0; 1/2; 2)$ com parametrização λ ;

MODELO II: $x_0 \sim N(\mu_{x_0}; \sigma_{x_0}^2)$ e $b \sim t(0; \pi/4; 1/2)$ com parametrização λ ;

MODELO III: $x_0 \sim N(\mu_{x_0}; \sigma_{x_0}^2)$ e $b \sim t(0; 1/2; 2)$ com parametrização δ ;

MODELO IV: $x_0 \sim N(\mu_{x_0}; \sigma_{x_0}^2)$ e $b \sim t(0; \pi/4; 1/2)$ com parametrização δ ;

MODELO V: $x_0 \sim t(\bar{x}; (S_x^2) * \frac{n+1}{n-3}; n-3)$ e $b \sim t(0; 1/2; 2)$ com parametrização δ ;

MODELO VI: $x_0 \sim t(\bar{x}; (S_x^2) * \frac{n+1}{n-3}; n-3)$ e $b \sim t(0; \pi/4; 1/2)$ com parametrização δ ;

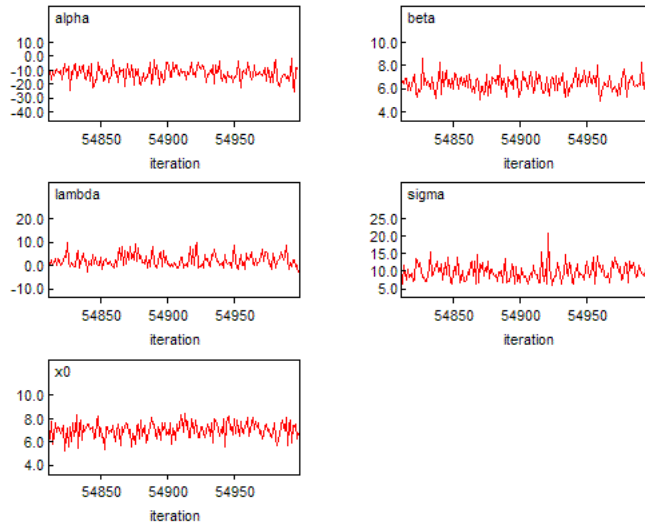
MODELO VII: Calibração usual, como definido em (1.1), com $x_0 \sim t(\bar{x}; (S_x^2) * \frac{n+1}{n-3}; n-3)$.

Foram geradas 5005000 iterações dentro do amostrador de Gibbs sendo descartadas as 5000 primeiras no processo de aquecimento das cadeias (*burn-in*). No entanto, dentro de cada cadeia, nós observamos a existência de baixo decaimento da função de auto-correlação, mostrando a grande correlação que existia entre as amostras para defasagens baixas. Em particular, a função de auto-correlação refe-

⁶NETER, J., KUTNER, M. H., NACHTSHEIN, C. J., LI, W. *Applied linear statistical models*. 4. ed. Boston: McGraw-Hill Irwin, p. 3-94, 1996.

⁷*Ibidem*, p 28.

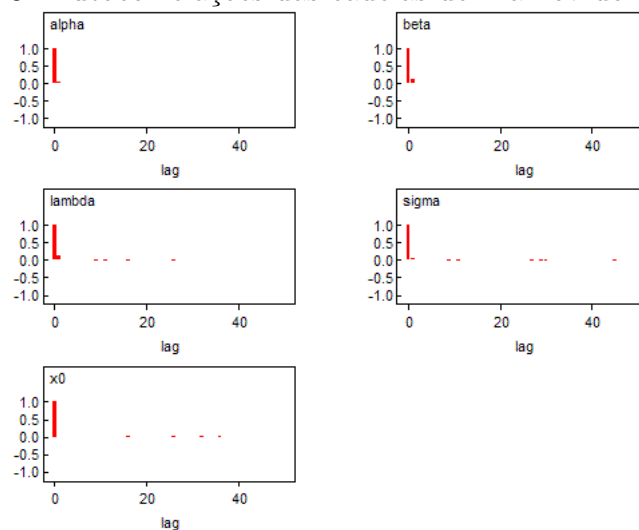
Figura 2.2: Convergência das cadeias de Markov para o modelo IV.



rente ao parâmetro de assimetria era a que apresentava o decaimento mais lento. Para contornar o problema, selecionamos as observações que efetivamente pertenceriam a amostra através de saltos com defasagem 100. Ou seja, a cada 100 valores gerados, retínhamos a centésima observação para fazer parte da amostra. Assim, a amostra final foi formada por cinquenta mil observações aproximadamente independentes.

Em geral, o comportamento da convergência das cadeias de Markov ocorreu,

Figura 2.3: Autocorrelações das cadeias de Markov do modelo IV.



como esperávamos para os modelos ajustados e a Figura 2.2 ressalta este fato para o modelo IV. A Figura 2.3 mostra o gráfico da função de auto-correlação para o modelo IV e indica que o procedimento foi suficiente para garantir amostras de Monte Carlo aproximadamente independentes. Para este modelo em questão. A Tabela 2.2 também mostra os desvios padrões, o erro de Monte Carlo, o percentil de ordem 2,5%, a mediana e o percentil de ordem 97,5% para a construção de intervalos HPD (do inglês, *highest probability density*).

A fim de comparar estes sete modelos (note que os dois primeiros não são novos modelos, apenas parametrizações distintas), utilizamos o DIC ($DIC = Dbar + \rho D = Dhat + 2 * \rho D$). Uma observação importante é que o programa WinBugs parece não calcular o DIC corretamente quando temos modelos hierárquicos e é irrealístico quando temos posterioris assimétricas para os parâmetros. Chegamos a essa conclusão através do cálculo do critério em diferentes plataformas. Logo, é necessário calculá-lo, utilizando nosso próprio código (escrevendo-o, manualmente) em algum outro programa que não seja o WinBugs. Nós nos convencemos disto, pelo fato do DIC parecer ser um dos mais realísticos critérios de avaliação de modelos bayesianos, já que combina a desviância esperada à posteriori, $Dbar$, que é a média das avaliações de desviância, calculadas a cada amostra de Monte Carlo que é gerada; a desviância das médias à posteriori, $Dhat$, que é a desviância avaliada com o estimador de Bayes sob perda quadrática, as médias à posteriori; e o efetivo número de parâmetros, ρD , como foi apresentado em Spiegelhalter *et al.* (2002). Em outras palavras, $Dbar$ é a média à posteriori de todas as desviâncias, onde a desviância é definida como $-2 * \log(vero.)$, $Dhat$ é um estimador pontual da desviância obtido pelas substituições das estimativas das médias à posteriori para os parâmetros do modelo e ρD é dado por $\rho D = Dbar - Dhat$. Quanto menor o valor do critério, melhor será o ajuste. No nosso modelo, não temos distribuições à posteriori para os parâmetros que sejam simétricas, como podemos ver na Tabela 2.2, onde médias e medianas não estão tão

Tabela 2.2: Estimador de Bayes sob perda quadrática, seu desvio padrão, o erro da cadeia de Markov bem como o estimador de Bayes sob perda absoluta e os percentis associados aos limites do intervalo de credibilidade HPD.

Parâmetro	média	desvio padrão	erro da cadeia	2,5%	mediana	97,5%
MODELO I						
α	-11,690	4,747	0,025	-21,030	-11,670	-2,410
β	6,747	0,586	0,004	5,574	6,758	7,870
λ	0,881	3,298	0,058	-1,344	0,279	6,388
σ	8,379	1,839	0,013	5,781	8,016	12,860
x_0	6,958	0,707	0,003	5,528	6,971	8,315
MODELO II						
α	-13,340	4,727	0,116	-22,540	-13,420	-3,948
β	6,352	0,618	0,012	5,213	6,315	7,645
λ	39,130	98,030	5,965	-1,317	5,309	330,200
σ	10,720	2,417	0,044	6,319	10,780	15,450
x_0	7,112	0,750	0,015	5,512	7,171	8,330
MODELO III						
α	-11,730	4,683	0,022	-20,910	-11,750	-2,514
β	6,748	0,570	0,003	5,610	6,754	7,847
λ	0,604	1,374	0,007	-1,281	0,317	4,313
σ	8,310	1,774	0,007	5,790	7,971	12,650
x_0	6,952	0,707	0,003	5,525	6,965	8,312
MODELO IV						
α	-12,500	4,803	0,023	-21,860	-12,550	-3,022
β	6,548	0,631	0,003	5,341	6,537	7,802
λ	2,038	2,681	0,012	-1,678	1,461	8,545
σ	9,629	2,313	0,010	6,059	9,382	14,720
x_0	6,987	0,728	0,003	5,462	7,020	8,298
MODELO V						
α	-11,730	4,725	0,020	-21,030	-11,770	-2,399
β	6,743	0,573	0,002	5,601	6,748	7,854
λ	0,621	1,400	0,006	-1,272	0,323	4,309
σ	8,351	1,795	0,008	5,801	8,009	12,690
x_0	6,956	0,787	0,004	5,352	6,978	8,449
MODELO VI						
α	-12,500	4,780	0,019	-21,760	-12,550	-3,000
β	6,547	0,629	0,003	5,351	6,537	7,799
λ	2,032	2,688	0,014	-1,688	1,445	8,603
σ	9,650	2,315	0,011	6,086	9,399	14,750
x_0	6,991	0,813	0,004	5,268	7,039	8,434
MODELO VII						
α	-10,920	4,276	0,018	-19,310	-10,930	-2,487
β	6,870	0,509	0,002	5,871	6,872	7,860
σ	7,190	1,065	0,005	5,472	7,064	9,635
x_0	6,948	0,783	0,003	5,396	6,954	8,474

próximas para alguns parâmetros e algumas especificações. Isto é evidenciado para o parâmetro λ , por exemplo, nos Modelos, II, IV e VI. As Figuras 2.4, 2.5, 2.6, 2.7, 2.8 e 2.9 mostram as densidades à posteriori para cada um dos modelos estudados e, claramente, podemos perceber a assimetria existente na distribuição à posteriori para o parâmetro λ em todos eles. O parâmetro de assimetria é aparentemente bastante sensível à elicitação das distribuições à priori, como foi observado por um *referee* do *Journal of Chemometrics*. Baseado nesta observação, propusemos uma modificação do DIC ordinário, quando tivermos modelos assimétricos. Nossa proposta é calcular o DIC, utilizando o estimador de Bayes sob perda absoluta: a mediana. Mas não apenas no cálculo do DIC, mas considerá-la em todas as análises, para evitar conclusões distorcidas pela utilização da média, influenciada por valores extremos. O novo DIC, que chamamos ADIC, precisa ser calculado avaliando $Dbar$ como a mediana à posteriori de todas as desviâncias calculadas a cada geração de uma amostra de Monte Carlo e $Dhat$ como a estimativa pontual da desviância, obtida pelo cálculo da desviância avaliada pelo estimador de Bayes sob perda absoluta: a mediana.

Tabela 2.3: Avaliação dos modelos sob o critério DIC

Modelo	$Dbar$	$Dhat$	ρD	DIC
Modelo I	181,674	182,514	-0,840	180,833
Modelo II	177,001	475,651	-298,651	-121,650
Modelo III	182,428	180,868	1,560	183,989
Modelo IV	179,276	182,390	-3,114	176,162
Modelo V	180,787	180,920	-0,133	180,654
Modelo VI	186,228	182,356	3,872	190,100
Modelo VII	182,469	179,277	3,191	185,660

Note que o critério ADIC parece ser mais realístico do que o DIC definido por Spiegelhalter *et al.* (2002) e que está implementado no programa WinBugs, quando temos especificações assimétricas. De fato, para o Modelo II, a estimativa de λ e o valor do DIC são 39,130 e $-121,650$, respectivamente, retratando um comportamento muito diferente dos outros modelos, como pode ser visto na Tabela 2.3. Claro, é um

Tabela 2.4: Avaliação dos modelos sob o critério ADIC

Modelo	$Dbar$	$Dhat$	ρD	$ADIC$
Modelo I	181,591	180,268	1,323	182,914
Modelo II	176,135	178,713	-2,578	173,557
Modelo III	182,155	180,094	2,061	184,217
Modelo IV	178,862	179,916	-1,054	177,809
Modelo V	180,527	180,148	0,379	180,907
Modelo VI	185,451	180,107	5,344	190,795
Modelo VII	181,789	179,155	2,634	184,423

resultado fictício e irreal, já que as especificações II e IV são as mesmas, exceto por uma diferente reparametrização do parâmetro de assimetria. Do mesmo modo, as especificações I e III são equivalentes. O único ganho nas especificações III e IV está na velocidade de geração das amostras das cadeias de Markov, como observou Bazan *et al.* (2006). Mesmo nos modelos simétricos, o ADIC mostrou ser mais realístico que o DIC, já que a densidade à posteriori de σ é também um tanto assimétrica, em geral, gama invertida, como podemos ver na Tabela 2.2 e Figura 2.10. A média e

Figura 2.4: Densidades à posteriori dos parâmetros do Modelo I

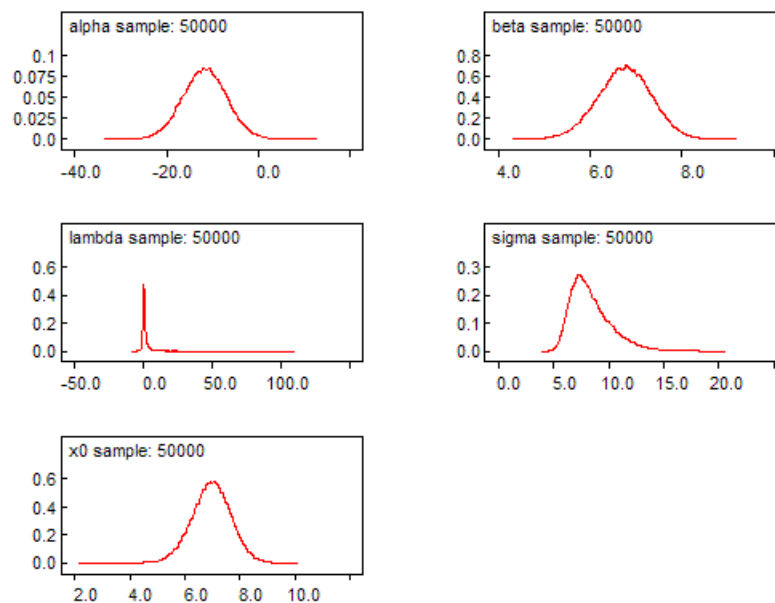


Figura 2.5: Densidades à posteriori dos parâmetros do Modelo II

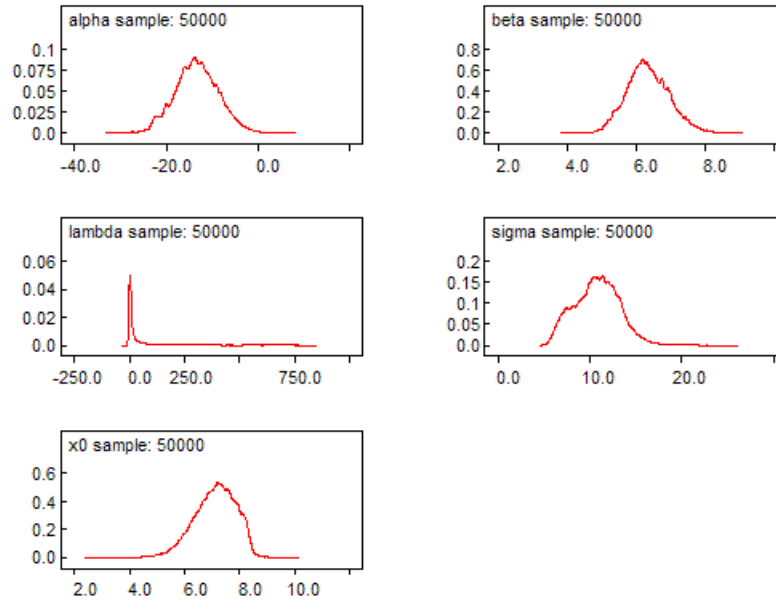


Figura 2.6: Densidades à posteriori dos parâmetros do Modelo III

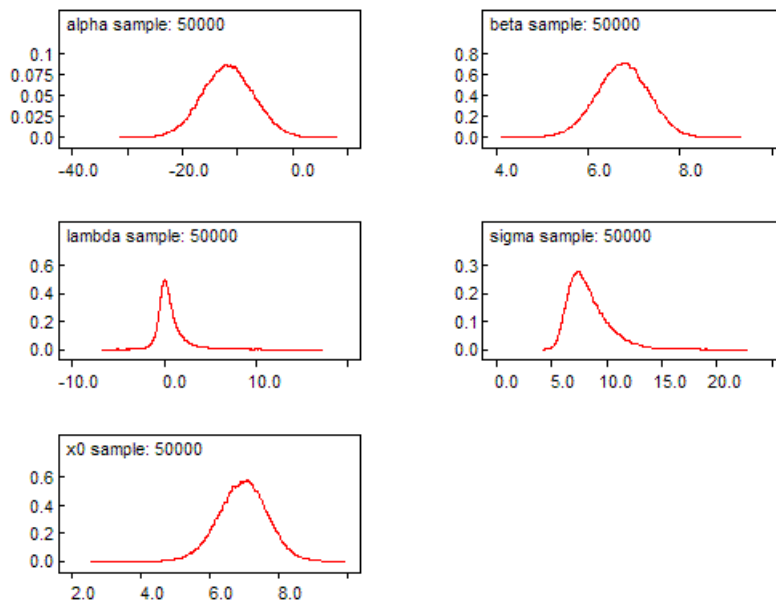


Figura 2.7: Densidades à posteriori dos parâmetros do Modelo IV

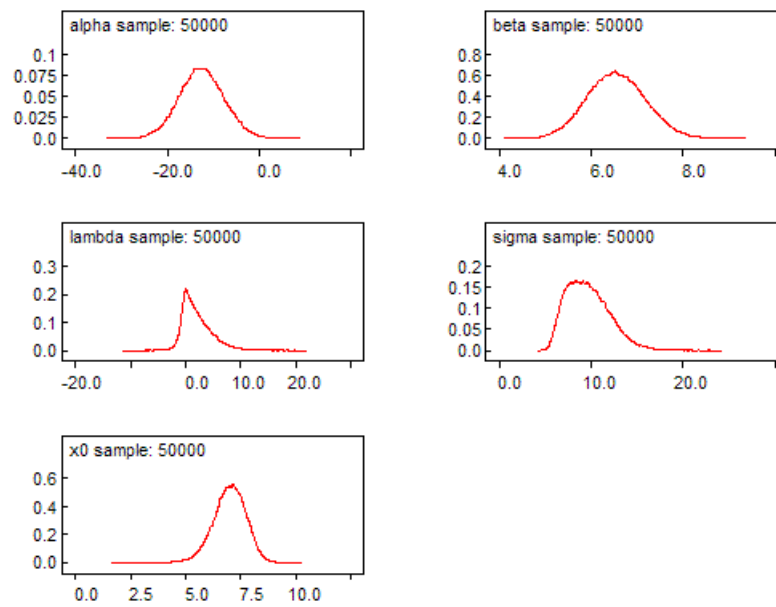


Figura 2.8: Densidades à posteriori dos parâmetros do Modelo V

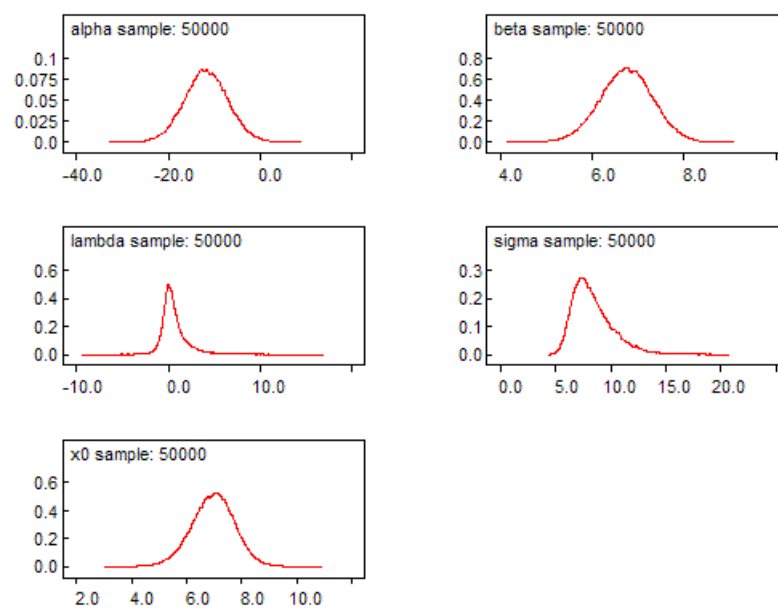


Figura 2.9: Densidades à posteriori dos parâmetros do Modelo VI

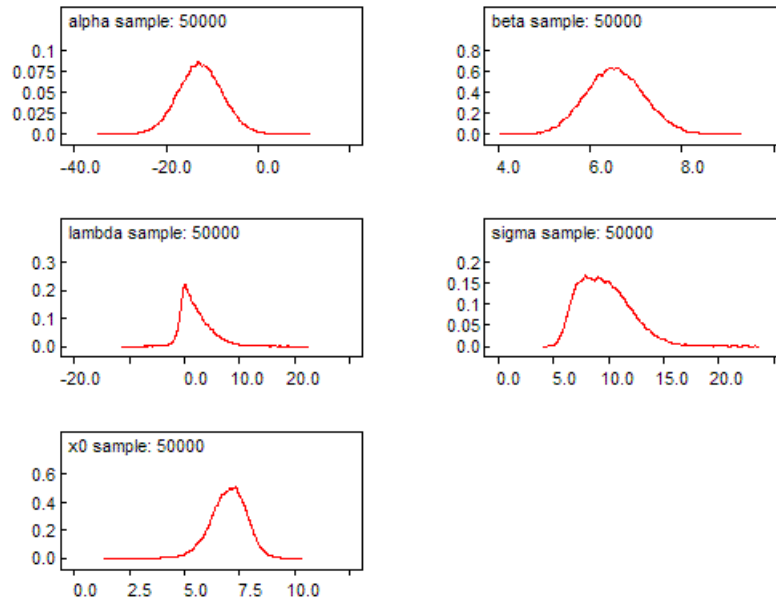
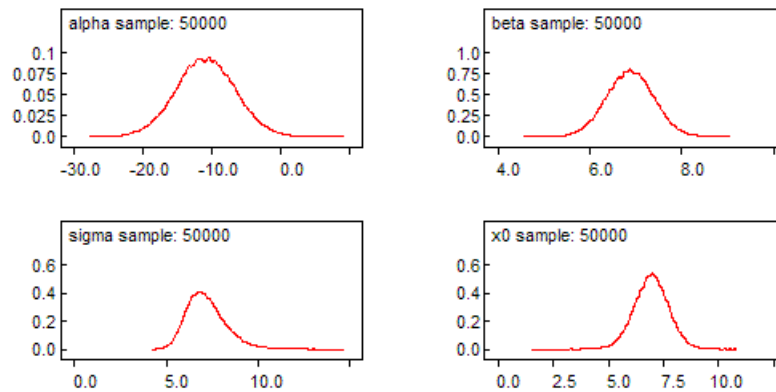


Figura 2.10: Densidades à posteriori dos parâmetros do Modelo VII



mediana para σ no modelo simétrico, designado por Modelo VII, foram dadas por 7,190 e 7,064, levando a um menor valor de ADIC (184,423), quando comparado com o DIC (185,660). As Figuras 2.4, 2.5, 2.6, 2.7, 2.8 e 2.9 mostram que as densidades à posteriori para σ também são ligeiramente assimétricas em cada um dos modelos assimétricos estudados. Assim, como pode ser mostrado na Tabela 2.4, obtivemos

que o modelo normal-assimétrico superou em performance o modelo de calibração usual, definido pelo modelo VII , quando temos a especificação à priori de Jeffreys para b (Modelos II e IV), já que os valores do critério ADIC são menores para estas duas especificações do que para o modelo usual de calibração com a priori t de *Student* proposta por Hoadley (1970), embora combinado com a priori t de *Student* para x_0 o modelo com a priori de Jeffreys não tenha apresentado um bom ajuste (Modelo VI). Por outro lado, os modelos I, III e V com equivalentes especificações à priori para o parâmetro b apresentaram melhores ajustes aos dados quando comparados com o Modelo VII, por possuírem menores valores dos critérios de informação. Além disso, esta especificação à priori para o parâmetro b combinada com a priori t de *Student* para x_0 apresentou a melhor performance neste grupo (Modelo V).

Por outro lado, em todas as especificações, a estimativa de x_0 esteve um pouco acima do valor real, como mostra a Tabela 2.2. Assim, o modelo que melhor se ajustou aos dados foi o Modelo IV, já que é o mais rápido para ser gerado e apresenta o menor ADIC.

Também é importante notar que Spiegelhalter *et al.* (2002) mencionou que quando ρD é menor que zero, isto pode indicar que há um conflito entre as escolhas feitas à priori e os dados. No nosso caso, isso foi gerado por uma má utilização do estimador de Bayes e, nós corrigimos este problema, substituindo a média pela mediana nas análises. Claramente, é a explicação para pequenos valores de DIC na Tabela 2.3.

Modelo de calibração

t-normal-assimétrico

3.1 Introdução

Neste capítulo, apresentamos uma modelagem que leva em consideração não só a fuga da suposição de simetria dos erros, implícita na suposição de normalidade usual, mas que também acomode eventuais distribuições com caudas mais pesadas que a normal.

O objetivo deste capítulo é propormos um modelo de calibração linear, onde faremos a suposição de que os erros são provenientes de uma distribuição t-normal-assimétrica (como referência temos Gomez, Venegas e Bolfarine, 2007). Uma outra abordagem poderia ser o uso da distribuição t-assimétrica (como referência, Branco e Dey, 2001). Implementaremos alguma rotina computacional para a estimação dos parâmetros e verificaremos a qualidade do ajuste através dos critérios AIC, BIC e HQ. Além disso, ilustraremos o estudo aplicando-o a um conjunto de dados reais. Na ótica bayesiana, faremos um estudo semelhante ao que foi feito no caso normal-assimétrico, utilizando o ADIC como medida de qualidade de ajuste. Branco e Dey (2001) foram os pioneiros no estudo de uma distribuição que acomodasse tanto a as-

simetria, quanto valores extremos, ou amostras pequenas, onde é necessário atribuir mais peso às caudas da distribuição. Nesse trabalho, Branco e Dey (2001) apresentaram uma distribuição que está um passo a frente do estudo feito por Azzalini e Dalla Valle (1996) e Azzalini e Capitanio (1999) para a modelagem. A seguir, Azzalini e Capitanio (2003) apresentaram os principais resultados inferenciais para a distribuição t-assimétrica multivariada. Nesse ensejo, Gomez, Venegas e Bolfarine (2007) apresentaram a distribuição t-normal-assimétrica que pode ser vista como um caso particular de modelos de mistura de escala normal como apresentou Ferreira (2008).

Muitos trabalhos recentes, como Godoy (2007) e Ferreira (2008), estudaram as propriedades dos modelos t-assimétricos e t-normal-assimétricos e algumas variações deles que já existem, devido às diferentes definições que são feitas para os esses modelos. Como iremos estender os resultados obtidos nos capítulos 1 e 2 para a distribuição normal-assimétrica, resolvemos aglutinar as abordagens freqüentista e bayesiana no mesmo capítulo, sendo que a segunda seção deste capítulo trará a investigação do modelo t-normal-assimétrico segundo a ótica freqüentista e a terceira seção sob a ótica bayesiana.

3.2 A distribuição t-normal-assimétrica

Uma variável aleatória X é dita t-normal-assimétrica se sua função densidade de probabilidade é dada por

$$f_X(x) = 2t_\nu(x|\mu; \sigma^2)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}, \quad (3.1)$$

onde μ é o parâmetro de locação, σ é o parâmetro de escala e λ , o parâmetro de forma. Além disso, $\Phi(\cdot)$ representa a função de distribuição acumulada da normal padrão e $t_\nu(\cdot|\mu; \sigma^2)$ representa a função densidade de probabilidade da distribuição

t-Student com parâmetro de locação μ , parâmetro de escala σ e ν graus de liberdade. Gomez, Venegas e Bolfarine (2007) apresentaram os principais resultados sobre esta distribuição. As idéias básicas para a implementação do modelo de regressão com a suposição dos erros t-normal-assimétricos distribuídos podem ser vistas em Ferreira (2008). Utilizando a mesma notação que a adotada no caso de regressão¹, indicaremos por $X \sim STN_\nu(\mu; \sigma^2; \lambda)$, a distribuição especificada em (3.1) e por $X \sim STN_\nu(\lambda)$ a distribuição t-normal-assimétrica padrão, que tem $\mu = 0$ e $\sigma^2 = 1$. A notação STN vem do inglês, *skew-t-normal*.

3.3 Enfoque freqüentista

3.3.1 Obtenção dos estimadores através do algoritmo EM

O modelo de calibração t-normal-assimétrico é dado por

$$\begin{aligned} y_i &= \alpha + \beta x_i + \varepsilon_i, & i = 1, \dots, n; \\ y_{0j} &= \alpha + \beta x_0 + \varepsilon_{0j}, & j = 1, \dots, k, \end{aligned} \quad (3.2)$$

supondo que ε_i e ε_{0j} são independentes e identicamente distribuídos segundo uma distribuição $STN_\nu(0; \sigma^2; \lambda)$. Os graus de liberdade, ν , serão considerados fixos.

Sob as condições acima, o logaritmo da função de verossimilhança pode ser escrito da seguinte maneira:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= n \log c - \frac{n}{2} \log \sigma^2 - \frac{\nu + 1}{2} \sum_{i=1}^n \log \left(1 + \frac{(y_i - \alpha - \beta x_i)^2}{\nu \sigma^2} \right) + \sum_{i=1}^n \log \Phi \left(\frac{\lambda(y_i - \alpha - \beta x_i)}{\sigma} \right) \\ &+ k \log c - \frac{k}{2} \log \sigma^2 - \frac{\nu + 1}{2} \sum_{j=1}^k \log \left(1 + \frac{(y_{0j} - \alpha - \beta x_0)^2}{\nu \sigma^2} \right) + \sum_{j=1}^k \log \Phi \left(\frac{\lambda(y_{0j} - \alpha - \beta x_0)}{\sigma} \right), \end{aligned}$$

¹Op. cit.

com a constante $c = \frac{2\Gamma(\frac{\nu+1}{2})}{\sqrt{(\pi\nu)}\Gamma(\frac{\nu}{2})}$ e $\boldsymbol{\theta} = (\alpha, \beta, x_0, \sigma^2, \lambda)^\top$. Para a obtenção dos estimadores de máxima verossimilhança, por maximização direta, precisaríamos encontrar as derivadas de primeira ordem em relação ao vetor $\boldsymbol{\theta} = (\alpha, \beta, x_0, \sigma^2, \lambda)^\top$.

Azzalini e Capitanio (2003) desenvolveram um processo de estimação dos parâmetros do modelo t-assimétrico através de maximização direta (vide Branco e Dey, 2001). O mesmo poderia ser tentado para o modelo t-normal-assimétrico, porém pode-se utilizar um algoritmo do tipo EM que possui expressões fechadas para os estimadores dos parâmetros e que evita a diferenciação do fator $\Phi()$, utilizando propriedades de misturas de variáveis bastante conhecidas. Nesse contexto de mistura de variáveis conhecidas, temos que a distribuição t de *Student* pode ser obtida através da mistura de uma normal e uma gama. Muitos autores chamam as misturas de representações hierárquicas.

Para exemplificar o processo de obtenção de uma função com densidade de probabilidade $t_\nu(\cdot|\mu; \sigma^2)$, sendo $\mu = \alpha + \beta x$, temos o seguinte resultado:

RESULTADO: Vide Fang, Kotz e Ng, 1990. Se $Y|W = w \sim N(\alpha + \beta x; \sigma^2/w)$ e $W \sim G(\nu/2; \nu/2)$, então $Y \sim t_\nu(\alpha + \beta x; \sigma^2)$, onde $G(a; b)$ é a notação utilizada para a distribuição Gama com parâmetro de forma $a > 0$ e de escala $b > 0$.

Demonstração: Observe que $f_{Y,W}(y, w)$

$$= \frac{1}{\sqrt{2\pi\sigma^2/w}} \exp\left\{-\frac{w}{2\sigma^2} [(y - \alpha - \beta x)^2]\right\} \times \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} w^{\nu/2-1} \exp\{-\nu/2w\}.$$

Após manipulações algébricas, temos que $f_{Y,W}(y, w)$

$$= \left[\frac{(y - \alpha - \beta x)^2}{2\sigma^2} + \nu/2 \right]^{-(\nu+1)/2} \frac{\Gamma((\nu+1)/2)}{\sqrt{2\pi\sigma^2}} \\ \times w^{(\nu+1)/2-1} \exp\left\{-w \left[\frac{(y - \alpha - \beta x)^2}{2\sigma^2} + \nu/2 \right]\right\} \frac{\left[\frac{(y - \alpha - \beta x)^2}{2\sigma^2} + \nu/2 \right]^{(\nu+1)/2}}{\Gamma((\nu+1)/2)}.$$

Basta integrarmos $f_{Y,W}(y, w)$ em relação a w que obteremos $f_Y(y)$ com distribuição t de *Student*. Notemos que o último fator da expressão acima é a distribuição

condicional $f_{W|Y}(w) \sim Gama((\nu + 1)/2); ((y - \alpha - \beta x)^2/2\sigma^2 + \nu/2)$, que integrada torna-se o elemento neutro da multiplicação e, portanto, o que sobra é a expressão de uma função com densidade t de *Student*. \square

Outras propriedades e detalhes podem ser encontrados em Fang, Kotz e Ng (1990). Com esta representação hierárquica combinada aos resultados obtidos por Ferreira (2008) para o caso da distribuição t-normal-assimétrica, temos que a verossimilhança aumentada do modelo (3.2), considerando z o vetor de dados aumentados, diferente do que havia sido definido em (1.8), agora acrescido das variáveis latentes $w = w_1, \dots, w_n$ e $w_0 = w_{01}, \dots, w_{0j}$ de modo que $z = (y, y_0, t, t_0, w, w_0)$, pode ser reescrita por

$$\begin{aligned} f(\mathbf{y}) &= \prod_{i=1}^n 2t_\nu(y_i|\alpha + \beta x_i; \sigma^2) \Phi\left(\lambda \frac{y_i - \alpha - \beta x_i}{\sigma}\right) \prod_{j=1}^k 2t_\nu(y_{0j}|\alpha + \beta x_0; \sigma^2) \Phi\left(\lambda \frac{y_{0j} - \alpha - \beta x_0}{\sigma}\right) \\ &= \prod_{i=1}^n 2 \int_0^\infty \int_0^\infty \phi(y_i|\alpha + \beta x_i; \sigma^2/w_i) G_{\frac{\nu}{2}; \frac{\nu}{2}}(w_i) \phi(t_i|\lambda y_i - \lambda(\alpha + \beta x_i); \sigma^2) dt_i dw_i \times \\ &\quad \prod_{j=1}^k 2 \int_0^\infty \int_0^\infty \phi(y_{0j}|\alpha + \beta x_0; \sigma^2/w_{0j}) G_{\frac{\nu}{2}; \frac{\nu}{2}}(w_{0j}) \phi(t_{0j}|\lambda y_{0j} - \lambda(\alpha + \beta x_0); \sigma^2) dt_{0j} dw_{0j} \end{aligned} \quad (3.3)$$

sendo $\phi(\cdot|a; b)$ a função densidade de probabilidade da $N_1(a; b)$ e $G_{\frac{\nu}{2}; \frac{\nu}{2}}(\cdot)$ a função densidade de probabilidade da distribuição $Gama(\frac{\nu}{2}; \frac{\nu}{2})$. De onde é visível a seguinte representação hierárquica para o modelo de calibração:

$$\begin{aligned} y_i|w_i; t_i; x_i &\sim N\left(\alpha + \beta x_i + \frac{\sigma\lambda}{w_i\sqrt{1 + \lambda^2/w_i}}t_i; \frac{\sigma^2}{w_i + \lambda^2}\right); \\ w_i &\sim G\left(\frac{\nu}{2}; \frac{\nu}{2}\right); \\ t_i &\sim HN(0; 1), \quad i = 1, \dots, n; \end{aligned}$$

e

$$\begin{aligned} y_{0j}|w_{0j}; t_{0j} &\sim N\left(\alpha + \beta x_0 + \frac{\sigma \lambda}{w_{0j} \sqrt{1 + \lambda^2/w_{0j}}} t_{0j}; \frac{\sigma^2}{w_{0j} + \lambda^2}\right); \\ w_{0j} &\sim G\left(\frac{\nu}{2}; \frac{\nu}{2}\right); \\ t_{0j} &\sim HN(0; 1), \quad j = 1, \dots, k. \end{aligned}$$

Note também que uma consequência direta do resultado anterior é que $W|Y = y_i \sim G\left(\frac{\nu+1}{2}; \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2} + \frac{\nu}{2}\right)$ e $W_0|Y_0 = y_{0j} \sim G\left(\frac{\nu+1}{2}; \frac{(y_{0j} - \alpha - \beta x_0)^2}{2\sigma^2} + \frac{\nu}{2}\right)$. Assim, podemos obter os dois primeiros momentos de $W|Y$ por

$$w_{1i} = E(w|y = y_i) = (\nu + 1) \left[\frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2} + \nu \right]^{-1}$$

e

$$\begin{aligned} w_{2i} = E(w^2|y = y_i) &= Var(w_i|y = y_i) + E^2(w_i|y = y_i) \\ &= \frac{\frac{\nu+1}{2}}{\left(\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2} + \frac{\nu}{2}\right)^2} \left[\frac{\nu + 1}{2} + 1 \right], \end{aligned}$$

para $i = 1, \dots, n$. Da mesma forma para a segunda parte do experimento, os momentos são dados por

$$w_{1j} = E(w|y = y_{0j}) = (\nu + 1) \left[\frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^2} + \nu \right]^{-1}$$

e

$$\begin{aligned} w_{2j} = E(w^2|y = y_{0j}) &= Var(w_j|y = y_{0j}) + E^2(w_j|y = y_{0j}) \\ &= \frac{\frac{\nu+1}{2}}{\left(\frac{(y_{0j} - \alpha - \beta x_0)^2}{2\sigma^2} + \frac{\nu}{2}\right)^2} \left[\frac{\nu + 1}{2} + 1 \right], \end{aligned}$$

para $j = 1, \dots, k$.

Os primeiros momentos apresentados acima serão utilizados para a maximização do modelo através do algoritmo EM e obtenção dos estimadores do vetor $\theta = (\alpha, \beta, \sigma, \lambda, x_0)$. No passo M vamos maximizar $\ell(\theta|z)$ em relação aos parâmetros do vetor θ , sendo z o vetor de dados aumentados que surge através da representação hierárquica.

Passo E

Com o que já foi exposto e aplicando o logaritmo natural à verossimilhança aumentada, dada em (3.3), temos que

$$\begin{aligned}
\ell(\boldsymbol{\theta}|\mathbf{z}) &\propto -n \log \sigma^2 - \sum_{i=1}^n \frac{w_i}{2\sigma^2} (y_i - \alpha - \beta x_i)^2 - \frac{\nu}{2} \sum_{i=1}^n w_i \\
&\quad - \frac{1}{2\sigma^2} \sum_{i=1}^n (t_i^2 - 2t_i \lambda(y_i - \alpha - \beta x_i) + \lambda^2 (y_i - \alpha - \beta x_i)^2) \\
&\quad - k \log \sigma^2 - \sum_{j=1}^k \frac{w_{0j}}{2\sigma^2} (y_{0j} - \alpha - \beta x_0)^2 - \frac{\nu}{2} \sum_{j=1}^k w_{0j} \\
&\quad - \frac{1}{2\sigma^2} \sum_{j=1}^k (t_{0j}^2 - 2t_{0j} \lambda(y_{0j} - \alpha - \beta x_0) + \lambda^2 (y_{0j} - \alpha - \beta x_0)^2). \tag{3.4}
\end{aligned}$$

Com $\hat{t}_i = E[t|y = y_i]$, $\hat{t}_i^2 = E[t^2|y = y_i]$, $\hat{w}_i = E[w|y = y_i]$, $\hat{t}_{0j} = E[t|y = y_{0j}]$, $\hat{t}_{0j}^2 = E[t^2|y = y_{0j}]$ e $\hat{w}_{0j} = E[w|y = y_{0j}]$ obtemos, usando os momentos de uma distribuição normal truncada (vide Arellano-Valle, Ozan, Bolfarine e Lachos, 2005), que

$$\begin{aligned}
\hat{t}_i &= \hat{\lambda}(y_i - \hat{\alpha} - \hat{\beta}x_i) + \frac{\phi(\cdot)}{\Phi(\cdot)} \hat{\sigma}; \\
\hat{t}_i^2 &= \hat{\lambda}^2(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \hat{\sigma}^2 + \frac{\phi(\cdot)}{\Phi(\cdot)} \hat{\lambda}(y_i - \hat{\alpha} - \hat{\beta}x_i) \hat{\sigma}; \quad e \\
\hat{w}_i &= (\nu + 1) \left(\frac{(y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{\hat{\sigma}^2} + \nu \right)^{-1} \tag{3.5}
\end{aligned}$$

para as quantidades referentes à primeira parte do experimento, ou seja, com $(\cdot) = (\hat{\lambda}(y_i - \hat{\alpha} - \hat{\beta}x_i)/\hat{\sigma})$, e

$$\begin{aligned} \hat{t}_{0j} &= \hat{\lambda}(y_{0j} - \hat{\alpha} - \hat{\beta}x_0) + \frac{\phi(\cdot)}{\Phi(\cdot)}\hat{\sigma}; \\ \hat{t}_{0j}^2 &= \hat{\lambda}^2(y_{0j} - \hat{\alpha} - \hat{\beta}x_0)^2 + \hat{\sigma}^2 + \frac{\phi(\cdot)}{\Phi(\cdot)}\hat{\lambda}(y_{0j} - \hat{\alpha} - \hat{\beta}x_0)\hat{\sigma}; \quad e \\ \hat{w}_{0j} &= (\nu + 1)\left(\frac{(y_{0j} - \hat{\alpha} - \hat{\beta}x_0)^2}{\hat{\sigma}^2} + \nu\right)^{-1} \end{aligned} \quad (3.6)$$

para as quantidades referentes à segunda parte do experimento, ou seja, com $(\cdot) = (\hat{\lambda}(y_{0j} - \hat{\alpha} - \hat{\beta}x_0)/\hat{\sigma})$.

De onde segue que a esperança condicionada em y da função de verossimilhança completa com respeito a t e w , usando (3.5) e (3.6), tem a forma

$$\begin{aligned} E[\ell(\boldsymbol{\theta}|\mathbf{z})|\mathbf{y};\hat{\boldsymbol{\theta}}] &\propto -(n+k)\log\hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n\hat{w}_i(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 - \frac{\nu}{2}\sum_{i=1}^n\hat{w}_i \\ &\quad - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n(\hat{t}_i^2 - 2\hat{\lambda}\hat{t}_i(y_i - \hat{\alpha} - \hat{\beta}x_i) + \hat{\lambda}^2(y_i - \hat{\alpha} - \hat{\beta}x_i)^2). \\ &\quad - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^k\hat{w}_{0j}(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2 - \frac{\nu}{2}\sum_{j=1}^k\hat{w}_{0j} \\ &\quad - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^k(\hat{t}_{0j}^2 - 2\hat{\lambda}\hat{t}_{0j}(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0) + \hat{\lambda}^2(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2). \end{aligned} \quad (3.7)$$

Passo M

No passo M vamos maximizar $\ell(\boldsymbol{\theta}; \mathbf{z})$ em relação aos parâmetros do vetor $\boldsymbol{\theta}$. Assim, depois de algumas manipulações algébricas, temos que

$$\hat{\alpha} = \frac{\sum_{i=1}^n(\hat{w}_i + \hat{\lambda}^2)(y_i - \hat{\beta}x_i) - \hat{\lambda}\sum_{i=1}^n\hat{t}_i + \sum_{j=1}^k(\hat{w}_{0j} + \hat{\lambda}^2)(y_{0j} - \hat{\beta}\hat{x}_0) - \hat{\lambda}\sum_{j=1}^k\hat{t}_{0j}}{\sum_{i=1}^n(\hat{w}_i + \hat{\lambda}^2) + \sum_{j=1}^k(\hat{w}_{0j} + \hat{\lambda}^2)};$$

$$\hat{\beta} = \frac{\sum_{i=1}^n (\hat{w}_i x_i + \hat{\lambda}^2 x_i)(y_i - \hat{\alpha}) - \hat{\lambda} \sum_{i=1}^n \hat{t}_i x_i + \sum_{j=1}^k (\hat{w}_{0j} \hat{x}_0 + \hat{\lambda}^2 \hat{x}_0)(y_{0j} - \hat{\alpha}) - \hat{\lambda} \hat{x}_0 \sum_{j=1}^k \hat{t}_{0j}}{\sum_{i=1}^n (\hat{w}_i x_i^2 + \hat{\lambda}^2 x_i^2) + \sum_{j=1}^k (\hat{w}_{0j} x_0^2 + \hat{\lambda}^2 x_0^2)}; \quad e$$

$$\hat{x}_0 = \frac{\sum_{j=1}^k (\hat{w}_{0j} + \hat{\lambda}^2)(y_{0j} - \hat{\alpha}) - \hat{\lambda} \sum_{j=1}^k \hat{t}_{0j}}{\sum_{j=1}^k (\hat{w}_{0j} + \hat{\lambda}^2) \hat{\beta}}. \quad (3.8)$$

Note que substituindo \hat{x}_0 nas expressões de $\hat{\alpha}$ e $\hat{\beta}$, as expressões destes estimadores ficam reduzidas a

$$\hat{\alpha} = \frac{\sum_{i=1}^n (\hat{w}_i + \hat{\lambda}^2)(y_i - \hat{\beta} x_i) - \hat{\lambda} \sum_{i=1}^n \hat{t}_i}{\sum_{i=1}^n (\hat{w}_i + \hat{\lambda}^2)}; \quad e \quad (3.9)$$

$$\hat{\beta} = \frac{\sum_{i=1}^n (\hat{w}_i x_i + \hat{\lambda}^2 x_i)(y_i - \hat{\alpha}) - \hat{\lambda} \sum_{i=1}^n \hat{t}_i x_i}{\sum_{i=1}^n (\hat{w}_i x_i^2 + \hat{\lambda}^2 x_i^2)}. \quad (3.10)$$

Ou seja, $\hat{\alpha}$ e $\hat{\beta}$ não dependem da segunda parte do experimento. Para facilidade computacional podemos utilizar para a estimação de $\hat{\alpha}$ e $\hat{\beta}$ a forma matricial

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \left(\mathbf{diag}(w_i) \mathbf{y} - \hat{\lambda} (\hat{\mathbf{t}} - \hat{\lambda} \mathbf{y}) \right), \quad (3.11)$$

sendo \mathbf{X} a matrix de planejamento da primeira parte do experimento com dimensão $n \times 2$, \mathbf{y} o vetor de respostas associado a primeira parte do experimento, \mathbf{W} uma matriz diagonal dos elementos $(\hat{w}_i + \hat{\lambda}^2)$ e $\hat{\mathbf{t}}$ o vetor de valores esperados “ \hat{t}_i ” dados em (3.5).

Já os estimadores de $\hat{\sigma}$ e $\hat{\lambda}$ têm expressões que dependem diretamente das duas partes do experimento. São elas

$$\hat{\sigma} = \frac{1}{\sqrt{2(n+k)}} \left\{ \sum_{i=1}^n (\hat{w}_i (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 + \hat{t}_i^2 - 2\hat{\lambda} \hat{t}_i (y_i - \hat{\alpha} - \hat{\beta} x_i) + \hat{\lambda}^2 (y_i - \hat{\alpha} - \hat{\beta} x_i)^2) \right.$$

$$+ \left. \sum_{j=1}^k (\hat{w}_{0j}(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2 + \hat{t}_{0j}^2 - 2\hat{\lambda}\hat{t}_{0j}(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0) + \hat{\lambda}^2(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2) \right\}^{\frac{1}{2}} ; \quad \text{e} \quad (3.12)$$

$$\hat{\lambda} = \frac{\sum_{i=1}^n \hat{t}_i(y_i - \hat{\alpha} - \hat{\beta}x_i) + \sum_{j=1}^k \hat{t}_{0j}(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)}{\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum_{j=1}^k (y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2}. \quad (3.13)$$

Note ainda que quando $\hat{\lambda} = 0$ as expressões (3.8), (3.9), (3.10), (3.12) e (3.13) são as mesmas expressões do modelo de calibração linear com a suposição de erros distribuídos segundo uma *t* de *Student* (Lima *et al.*, 2008), dados por

$$\begin{aligned} \hat{\alpha} &= \bar{y}_w - \hat{\beta}\bar{x}_w, \\ \hat{\beta} &= \frac{\sum_{i=1}^n \hat{w}_i(y_i - \bar{y}_w)(x_i - \bar{x}_w)}{\sum_{i=1}^n \hat{w}_i(x_i - \bar{x}_w)^2}, \\ \hat{x}_0 &= \frac{\bar{y}_{0w} - \hat{\alpha}}{\hat{\beta}}, \\ \hat{\sigma} &= \frac{1}{n+k} \left[\sum_{i=1}^n \hat{w}_i(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum_{j=1}^k \hat{w}_{0j}(y_{0j} - \hat{\alpha} - \hat{\beta}\hat{x}_0)^2 \right]^{\frac{1}{2}}, \end{aligned} \quad (3.14)$$

onde

$$\bar{y}_w = \frac{\sum_{i=1}^n \hat{w}_i y_i}{\sum_{i=1}^n \hat{w}_i}, \quad \bar{x}_w = \frac{\sum_{i=1}^n \hat{w}_i x_i}{\sum_{i=1}^n \hat{w}_i}, \quad \bar{y}_{0w} = \frac{\sum_{j=1}^k \hat{w}_{0j} y_{0j}}{\sum_{j=1}^k \hat{w}_{0j}}.$$

No demais, o algoritmo segue da seguinte maneira

1. Atribua valores iniciais $\hat{\alpha}^0$, $\hat{\beta}^0$, $\hat{\sigma}^0$, $\hat{\lambda}^0$ e \hat{x}_0^0 para o vetor de parâmetros θ .
2. Calcule as esperanças condicionais, $\forall l$;
3. Obtenha novos valores para os estimadores $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, $\hat{\lambda}$ e \hat{x}_0 ;
4. Repita os passos 2 e 3 até a convergência.

Os valores obtidos após a convergência para os estimadores $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\lambda}$ e \hat{x}_0 são as estimativas de máxima verossimilhança do modelo de calibração.

3.3.2 Variância dos estimadores

Nesta seção vamos estudar o comportamento assintótico dos estimadores de máxima verossimilhança através de suas variâncias assintóticas.

Um problema do modelo normal-assimétrico era que as condições de regularidade necessárias para desenvolvermos uma inferência baseada na teoria da verossimilhança não estavam satisfeitas. As condições de regularidade que nos referimos podem ser encontradas em Severini (2000). Entre outros aspectos, havia também a necessidade de que a função escore relativa ao parâmetro de assimetria deveria convergir em distribuição para uma normal multivariada. No modelo normal-assimétrico esses problemas foram solucionados com a reparametrização apresentada nas proposições (1) e (2) do Capítulo 1. Arellano-Valle e Azzalini (2008) apresentam um estudo de simulação que mostra que com a utilização do processo de centralização, a distribuição assintótica da função escore relativa ao parâmetro de assimetria λ é normal.

O modelo t-normal-assimétrico não enfrenta as mesmas dificuldades que existiam no caso do modelo normal-assimétrico. Dadas as condições favoráveis do modelo t-normal-assimétrico, os estimadores de variância assintótica do modelo de calibração (3.2) podem ser obtidos através do inverso da informação de Fisher, $I_F(\theta) = E[S(\theta)S(\theta)^\top]$, com $S(\theta)$ representando o vetor escore de parâmetros θ e seu estimador consistente deveria ser $[S(\hat{\theta})^\top S(\hat{\theta})]$, tipicamente conhecido como a matriz de informação observada.

O modelo t-normal-assimétrico é regular. Um modelo é dito regular se a função de log-verossimilhança pode ser aproximada por um polinômio em θ . Vamos listar as três principais propriedades de modelos regulares que foram extraídas de Severini (2000).

Propriedade 1 Suponha que Θ é um subconjunto de \mathfrak{R}^d e que exista um aberto $\Theta_0 \in \Theta$ tal que para cada $\theta_0 \in \Theta_0$ seja possível expandi-lo em torno de θ até quarta

ordem de modo que o resto, $R_n(\theta)$, seja de ordem $O_p(1)$.

Propriedade 2 As derivadas de até quarta ordem da função de log-verossimilhança devem ter cumulantes conjuntos acima da quarta ordem e esses cumulantes devem ser de ordem $O(n)$. As funções escore devem seguir o teorema do limite central de modo que $n^{1/2}S_\theta(\theta)$ convirja em distribuição para uma variável aleatória com distribuição normal multivariada.

Propriedade 3 É a famosa condição que é válida, quando é possível permutar os sinais de integração com o de diferenciação. Essa condição acaba sendo suficiente, porque o interesse é mostrar que dados índices não negativos i_1, i_2, i_3 e i_4 , tal que $i_1 + i_2 + i_3 + i_4 \leq 4$ e dados índices j, k, ℓ e m assumindo valores no conjunto $\{0, \dots, d\}$, teremos que para $\theta_0 \in \Theta_0$,

$$\begin{aligned} & \frac{\partial^{i_1+i_2+i_3+i_4}}{\partial \theta_j^{i_1} \partial \theta_k^{i_2} \partial \theta_\ell^{i_3} \partial \theta_m^{i_4}} E\{exp\{\ell(\theta) - \ell(\theta_0)\}; \theta_0\} |_{\theta=\theta_0} \\ &= E \left\{ \frac{\partial^{i_1+i_2+i_3+i_4}}{\partial \theta_j^{i_1} \partial \theta_k^{i_2} \partial \theta_\ell^{i_3} \partial \theta_m^{i_4}} exp\{\ell(\theta) - \ell(\theta_0); \theta_0\} |_{\theta=\theta_0} \right\}. \end{aligned}$$

Os cálculos para a obtenção da matriz de informação de Fisher esperada aparecem no Apêndice. As estimativas de variâncias dos estimadores serão calculadas através da informação de Fisher observada.

3.3.3 Aplicação

Utilizaremos o mesmo conjunto de dados que modelamos no Capítulo 1 com a distribuição normal-assimétrica, que tratava das medidas do volume testicular de 42 adolescentes, em mililitros cúbicos. A idéia é discutir as semelhanças e diferenças obtidas nas duas modelagens. Da mesma forma que foi proposto no Capítulo 1, nós evitaremos a transformação de variáveis se utilizarmos essa segunda alternativa de metodologia, além de conseguirmos incorporar no modelo caudas mais pesadas do que a distribuição normal-assimétrica.

Tabela 3.1: Critérios de informação para diferentes valores de graus de liberdade ν .

	Modelo t-normal-assimétrico					Modelo SN
ν	4	5	10	20	30	
AIC	150,34	151,41	154,64	156,57	156,71	160,69
BIC	159,03	160,10	163,33	165,25	165,40	169,38
HQ	146,20	147,27	150,49	152,42	152,57	156,55

A Tabela 3.1 mostra os diferentes valores dos critérios AIC, BIC e HQ para o modelo t-normal-assimétrico, considerando diferentes valores de graus de liberdade e para o modelo normal-assimétrico. Na estimação dos parâmetros os graus de liberdade foram considerados fixos. Note que o modelo com os menores valores dos critérios de informação é o que possui $\nu = 4$, ou seja, o modelo que melhor se ajustou aos dados foi aquele que possui um valor de ν o menor possível ($n \geq 3$). Também foi observado que o modelo t-normal-assimétrico ajustou-se melhor aos dados, quando comparado com o modelo ajustado no Capítulo 1, que pode ter sido incapaz de incorporar parte do excesso de curtose presente nos dados, calculado através do estimador $\tilde{\gamma}_1 = n(n+1)/((n-1)(n-2)(n-3)) \sum((y-\bar{y})/DP(y))^4 - 3(n-1)^2/((n-2)(n-3))$ que resultou em $\tilde{\gamma}_1 = 0,40$.

Uma observação bastante relevante que foi notada é que, ao aumentarmos o valor dos graus de liberdade, observamos a convergência dos valores dos critérios para os obtidos com o modelo normal-assimétrico, ilustrando a possível convergência do modelo t-normal-assimétrico para o normal-assimétrico, quando $\nu \rightarrow \infty$.

Tabela 3.2: Estimativas dos parâmetros e desvios padrões para $\nu = 4$

Parâmetro	Estimativa	Desvio padrão
α	-0,27	0,50
β	0,86	0,06
σ	1,37	0,15
x_0	12,06	1,20
λ	1,81	1,70

A Tabela 3.2 mostra os valores das estimativas dos parâmetros, obtidas através

do modelo com $\nu = 4$, bem como os respectivos desvios desses estimadores obtidos através do procedimento que foi descrito na seção anterior. É importante notar que embora o valor dos critérios de informação tenham sido favoráveis à escolha desse modelo, o vício do estimador de x_0 foi maior no modelo t-normal-assimétrico, do que no modelo normal-assimétrico.

3.4 Enfoque bayesiano

A proposta desta seção não é apresentar grandes avanços na área de modelagem com o modelo t-normal-assimétrico em calibração, mas compará-lo com os modelos do Capítulo 2 e estender a idéia de comparação de modelos através de critérios de informação que levem em consideração a forma assimétrica das distribuições à posteriori. Desenvolveremos uma outra idéia, bastante coerente, quando há assimetria notória. Lançaremos mão do trabalho de Pereira e Stern (1999) que introduziu a noção de evidência em testes de hipóteses. Outros podem acreditar que se trata de trabalhar em um cenário em que há a função de perda 0-1. A junção do DIC de Spiegelhalter *et al.* (2002) com a noção de evidência, dará lugar ao que chamaremos de EDIC (*Evident Deviance Information Criterion*). Dessa maneira, o estimador de Bayes a ser considerado será a moda da distribuição à posteriori que se situa no interior da região com maior densidade à posteriori (HPD). Podemos também definir o estimador como foi definido em Pereira e Stern (1999) que consideraram \mathbf{Y} a variável aleatória que gerava os dados \mathbf{y} , onde o espaço de medida foi construído através da tripla $(\Omega; \mathfrak{S}; \Theta)$, sendo Ω um espaço amostral não vazio, \mathfrak{S} uma σ -álgebra de subconjuntos de Ω e Θ o espaço paramétrico. O contexto de modelagem era aquele em que ao atribuirmos distribuições à priori para Θ e tendo observado os dados, teríamos distribuições à posteriori bem definidas. A medida de evidência proposta por esses autores girava em torno da densidade à posteriori sob a hipótese nula, $\theta = \theta_0$, já que

o contexto era o de teste de hipóteses.

Assim, o EDIC será dado, calculando-se $Dbar$ como o máximo da densidade à posteriori de todas as desviâncias, onde a desviância é definida como $-2 * \log(vero.)$. A experiência mostrou que em geral a utilização da média, mediana ou moda não altera tanto o cálculo de $Dbar$, pois sua distribuição é praticamente simétrica. O que realmente pode ser perturbado em presença da assimetria é $Dhat$ que será dado pelo estimador pontual da desviância obtido pelas substituições das estimativas dos máximos das densidades à posteriori para os parâmetros do modelo e ρD , da mesma maneira que anteriormente, por $\rho D = Dbar - Dhat$. Quanto menor o valor do critério EDIC, melhor será o ajuste.

Foram ajustados modelos equivalentes aos modelos I e II do Capítulo anterior para ilustração e comparação dos três critérios (DIC, ADIC e EDIC), utilizando as mesmas prioris para os parâmetros α , β e σ . Ainda considerando $\nu = 4$. As especificações para os parâmetros x_0 e λ foram:

MODELO I: $x_0 \sim N(\mu_x; \sigma_x^2)$ e $\lambda \sim t(0; 1/2; 2)$ com parametrização λ ; e

MODELO II: $x_0 \sim N(\mu_x; \sigma_x^2)$ e $\lambda \sim t(0; \pi/4; 1/2)$ com parametrização λ .

As Figuras 3.1 e 3.3 deixam claro a grande presença de assimetria nas distribuições à posteriori dos parâmetros do modelo. Como foi adotado os mesmos procedimentos de MCMC que no modelo normal-assimétrico, vimos nas Figuras 3.2 e 3.4 que os saltos de defasagem 100, não foram suficientes para eliminar a autocorrelação induzida pelo modelo t-normal-assimétrico.

Referir-nos-emos a esses modelos por Uniforme (Modelo I) e de Jeffreys (Modelo II), já que a sutileza deles é a especificação à priori que foi utilizada para o parâmetro λ . Para a manutenção do mesmo padrão, foram geradas 5005000 iterações como no Capítulo 2 e descartadas as 5000 primeiras no processo de aquecimento das cadeias (*burn-in*) com saltos em defasagem 100.

A Tabela 3.3 mostra os valores do EDIC, ADIC e DIC para o estudo realizado

Figura 3.1: Densidades à posteriori dos parâmetros do modelo t-normal-assimétrico com priori uniforme para o parâmetro de assimetria.

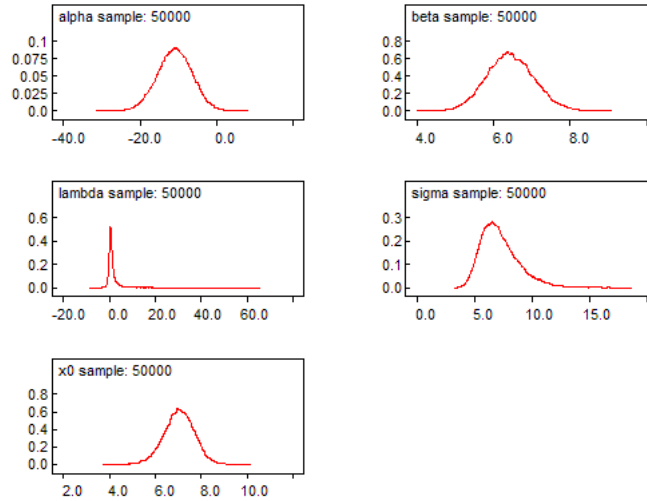


Tabela 3.3: Critérios de Informação

	Uniforme	Jeffreys
DIC	166,79	-1,07
ADIC	167,66	164,41
EDIC	170,61	156,86

Figura 3.2: Autocorrelações das cadeias de Markov para o modelo t-normal-assimétrico (uniforme).

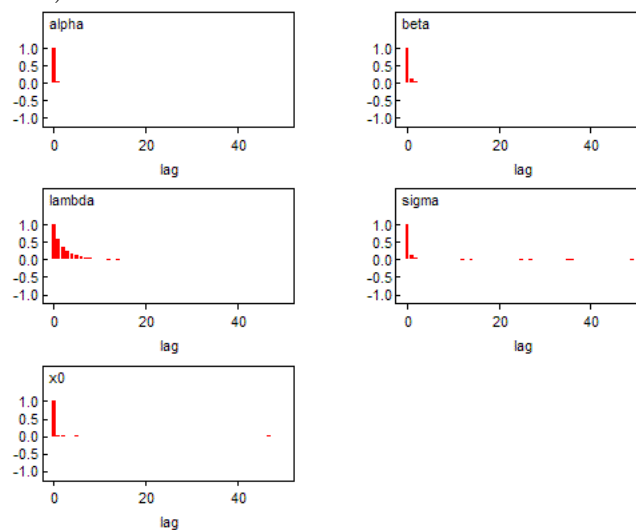
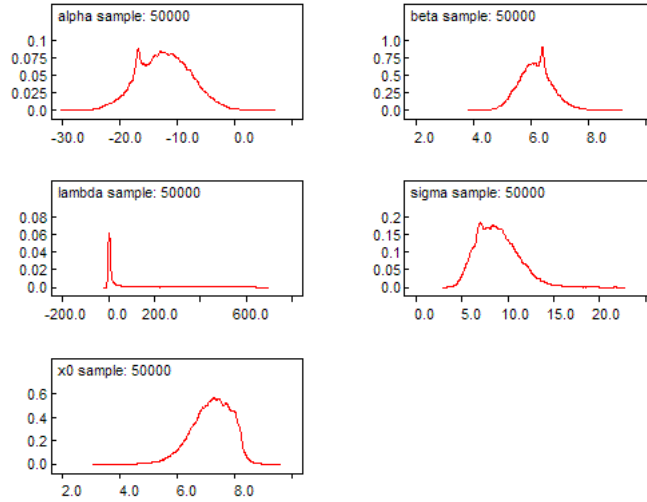


Figura 3.3: Densidades à posteriori dos parâmetros do modelo t-normal-assimétrico com priori de Jeffreys para o parâmetro de assimetria.



no Capítulo 2 para as vigas de concreto. É interessante notar que no caso do modelo Uniforme, tivemos os valores todos muito próximos e o grande impasse continua sendo quando temos posterioris bastante assimétricas como pode ser visto no caso do modelo com priori de Jeffreys.

Figura 3.4: Autocorrelações das cadeias de Markov para o modelo t-normal-assimétrico (Jeffreys).

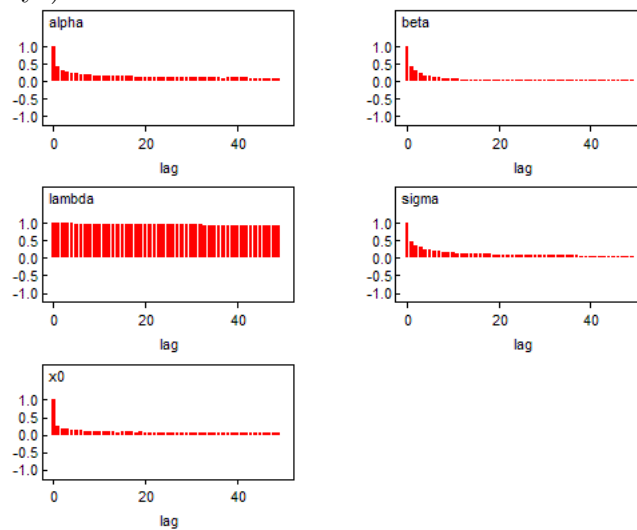


Tabela 3.4: Médias, Medianas e Modas

Modelo Uniforme			
	Média	Mediana	Moda
α	-11,13	-11,02	-10,74
β	6,30	6,27	6,39
σ	7,29	6,94	6,60
x_0	7,07	7,05	7,09
λ	2,03	0,66	0,50
Modelo de Jeffreys			
	Média	Mediana	Moda
α	-12,30	-12,36	-12,7
β	6,17	6,17	6,12
σ	8,83	8,63	7,32
x_0	7,19	7,25	7,34
λ	30,54	3,76	0,83

A Tabela 3.4 mostra que os valores das médias, medianas e modas para o modelo uniforme estão muito próximas e este foi o motivo pelo qual os critérios estiveram todos muito próximos. Já no modelo de Jeffreys a distribuição a posteriori para o parâmetro λ , com considerável assimetria, leva a valores de média, mediana e moda bastante diferentes. Como no Capítulo anterior, o DIC foi bastante irrealístico para esse modelo e o ADIC mostrou muito mais a informação presente na desviância do modelo. Vale ressaltar aqui, que como a assimetria era notável, o EDIC apresenta um valor mais coerente para a comparação dos modelos. Ou seja, a especificação com priori de Jeffreys foi a que melhor se ajustou aos dados.

Conclusões Finais

No Capítulo 1, desenvolvemos o algoritmo EM para a estimação dos parâmetros do modelo normal-assimétrico e observamos que os estimadores obtidos são uma generalização do modelo usual normal de calibração, quando $\hat{b} = 0$. Quando isto ocorrer, os estimadores serão os mesmos do modelo usual normal de calibração. Como um destaque da abordagem que foi desenvolvida, encontramos uma transformação para deixar a informação de Fisher, na vizinhança de $\lambda = 0$, não-singular que deixou o parâmetro de interesse x_0 invariante a ela. Além disto, nosso propósito foi satisfeito, pois quando avaliamos os modelos normal-assimétrico e normal, utilizando os três critérios de informação (AIC, BIC e HQ) notamos que o modelo normal-assimétrico produz um melhor ajuste para o conjunto de dados em questão.

Já no Capítulo 2, desenvolvemos uma modelagem bayesiana relacionada ao modelo de calibração normal-assimétrico. Além disso, introduzimos e avaliamos o critério assimétrico de informação da desviância (ADIC) e pudemos ver que o modelo de calibração normal-assimétrico superou o modelo usual por apresentar um valor mais baixo desta medida, especialmente, os modelos com a utilização da priori de Jeffreys. Também é possível salientar que quando a assimetria é forte, o estimador de Bayes sob a perda 0-1, ou seja, a moda, pode aferir mais fidedignamente qual o melhor ajuste e, uma outra modificação no DIC que utiliza a moda como o estimador

de Bayes, foi estudada no Capítulo 3.

No Capítulo 3, continuamos com a investigação relativa a encontrar um critério bayesiano que fosse o mais fiel possível para a escolha do modelo assimétrico que melhor descrevesse o comportamento dos dados. A metodologia bayesiana também foi investigada, onde propusemos uma segunda modificação do DIC usual para modelos assimétricos, o critério evidente de infomção da desviância (EDIC), que parece apresentar comportamento mais adequado que o DIC usual. Em suma, é necessário lembrar que o DIC usual apresenta, em geral, resultados insatisfatórios na presença de variáveis latentes.

Informação esperada e observada do modelo normal-assimétrico

Neste apêndice, as componentes da informação de Fisher esperada são apresentadas para o modelo de calibração normal-assimétrico descrito em (1.3). Elas foram obtidas através do cálculo da matriz de segundas derivadas da função de log-verossimilhança dada pela expressão (1.4)

Para o cálculo dessas expressões, precisaremos do cálculo de duas derivadas importantes e que aparecerão ao longo do texto. A primeira delas é:

$$\begin{aligned}\phi(.) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\lambda \frac{(y_i - \theta_i)}{\sigma} \right)^2 \right\} \\ \frac{\partial \phi(.)}{\partial \theta_i} &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\lambda \frac{(y_i - \theta_i)}{\sigma} \right)^2 \right\} \left(\lambda \frac{(y_i - \theta_i)}{\sigma} \right) \left(\frac{\lambda}{\sigma} \right) \\ &= \phi(.) \left(\lambda \frac{(y_i - \theta_i)}{\sigma} \right) \left(\frac{\lambda}{\sigma} \right)\end{aligned}$$

aqui $(.) = \lambda \frac{(y_i - \theta_i)}{\sigma}$.

A outra derivada importante é uma aplicação direta do teorema fundamental do cálculo¹ que diz se f for contínua em $[a, b]$, então a função g definida por

¹STEWART, J. **Cálculo I**, 5 ed., vol. I. São Paulo: Pioneira Thomsom Learning. 2006. Pp 393-400

$g(h(x)) = \int_a^{h(x)} f(t)dt$, $a \leq h(x) \leq b$ é contínua e diferenciável em (a, b) e $g'(h(x)) = f(h(x))h'(x)$.

Assim com $g(h(x)) =$

$$\begin{aligned}\Phi(\cdot) &= \int_{-\infty}^{\lambda \frac{(y_i - \theta_i)}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (y)^2 \right\} dy, \quad \text{vem que} \\ \frac{\partial \Phi}{\partial \theta_i} &= \phi(\cdot) \left(-\frac{\lambda}{\sigma} \right).\end{aligned}$$

Com o que foi exposto acima, teremos

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \alpha^2} &= - \frac{n}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{i=1}^n \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \\ &\quad - \frac{k}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \beta \partial \alpha} &= - \frac{\sum_{i=1}^n x_i}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{i=1}^n x_i \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \\ &\quad - \frac{k x_0}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) x_0 \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k x_0 \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \sigma \partial \alpha} &= - \frac{2 \sum_{i=1}^n (y_i - \alpha - \beta x_i)}{\sigma^3} + \sum_{i=1}^n \frac{\lambda}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &\quad - \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \frac{\lambda^3}{\sigma^4} \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{i=1}^n (y_i - \alpha - \beta x_i) \frac{\lambda^2}{\sigma^3} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \\ &\quad - \frac{2 \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)}{\sigma^3} + \sum_{j=1}^k \frac{\lambda}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &\quad - \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^3}{\sigma^4} \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\lambda^2}{\sigma^3} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda \partial \alpha} = & - \sum_{i=1}^n \frac{1}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \frac{\lambda^2}{\sigma^3} \frac{\phi(\cdot)}{\Phi(\cdot)} \\
& + \sum_{i=1}^n (y_i - \alpha - \beta x_i) \frac{\lambda}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \\
& - \sum_{j=1}^k \frac{1}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^2}{\sigma^3} \frac{\phi(\cdot)}{\Phi(\cdot)} \\
& + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\lambda}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};
\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial x_0 \partial \alpha} = - \frac{k\beta}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \beta \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k \beta \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta^2} = & - \frac{\sum_{i=1}^n x_i^2}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i^2 \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{i=1}^n x_i^2 \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \\
& - \frac{kx_0^2}{\sigma^2} - \frac{\lambda^3}{\sigma^3} \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) x_0^2 \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k x_0^2 \frac{\lambda^2}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma \partial \beta} = & - \frac{2 \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i)}{\sigma^3} + \sum_{i=1}^n x_i \frac{\lambda}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} \\
& - \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i)^2 \frac{\lambda^3}{\sigma^4} \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) \frac{\lambda^2}{\sigma^3} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \\
& - \frac{2 \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0)}{\sigma^3} + \sum_{j=1}^k x_0 \frac{\lambda}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} \\
& - \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^3}{\sigma^4} \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0) \frac{\lambda^2}{\sigma^3} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda \partial \beta} = & - \sum_{i=1}^n \frac{x_i \phi(\cdot)}{\sigma \Phi(\cdot)} + \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i)^2 \frac{\lambda^2 \phi(\cdot)}{\sigma^3 \Phi(\cdot)} \\
& + \sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) \frac{\lambda \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)} \\
& - \sum_{j=1}^k \frac{x_0 \phi(\cdot)}{\sigma \Phi(\cdot)} + \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^2 \phi(\cdot)}{\sigma^3 \Phi(\cdot)} \\
& + \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0) \frac{\lambda \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial x_0 \partial \beta} = & - \sum_{j=1}^k \frac{(y_{0j} - \alpha - 2\beta x_0)}{\sigma^2} - \sum_{j=1}^k \frac{\lambda \phi(\cdot)}{\sigma \Phi(\cdot)} \\
& - \frac{\lambda^3}{\sigma^3} \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0) \beta \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k \beta \frac{\lambda^2 x_0 \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma^2} = & - \frac{n}{\sigma^2} - \frac{3 \sum_{i=1}^n (y_i - \alpha - \beta x_i)}{\sigma^4} + 2 \sum_{i=1}^n \frac{\lambda}{\sigma^3} (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} \\
& - \sum_{i=1}^n (y_i - \alpha - \beta x_i)^3 \frac{\lambda^3 \phi(\cdot)}{\sigma^5 \Phi(\cdot)} - \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \frac{\lambda^2 \phi^2(\cdot)}{\sigma^4 \Phi^2(\cdot)} \\
& - \frac{k}{\sigma^2} - \frac{3 \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)}{\sigma^4} + 2 \sum_{j=1}^k \frac{\lambda}{\sigma^3} (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} \\
& - \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^3 \frac{\lambda^3 \phi(\cdot)}{\sigma^5 \Phi(\cdot)} - \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^2 \phi^2(\cdot)}{\sigma^4 \Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda \partial \sigma} = & - \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i) \phi(\cdot)}{\sigma^2} \Phi(\cdot) + \sum_{i=1}^n (y_i - \alpha - \beta x_i)^3 \frac{\lambda^2 \phi(\cdot)}{\sigma^4 \Phi(\cdot)} \\
& + \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \frac{\lambda \phi^2(\cdot)}{\sigma^3 \Phi^2(\cdot)} \\
& - \sum_{j=1}^k \frac{(y_{0j} - \alpha - \beta x_0) \phi(\cdot)}{\sigma^2} \Phi(\cdot) + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^3 \frac{\lambda^2 \phi(\cdot)}{\sigma^4 \Phi(\cdot)} \\
& + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda \phi^2(\cdot)}{\sigma^3 \Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial x_0 \partial \sigma} = & - \sum_{j=1}^k \frac{2(y_{0j} - \alpha - \beta x_0) \beta}{\sigma^3} + \sum_{j=1}^k \frac{\lambda \beta \phi(\cdot)}{\sigma^2 \Phi(\cdot)} \\
& - \sum_{j=1}^k \beta (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^3 \phi(\cdot)}{\sigma^4 \Phi(\cdot)} - \sum_{j=1}^k \beta (y_{0j} - \alpha - \beta x_0) \frac{\lambda^2 \phi^2(\cdot)}{\sigma^3 \Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda^2} = & - \sum_{i=1}^n (y_i - \alpha - \beta x_i)^3 \frac{\lambda \phi(\cdot)}{\sigma^3 \Phi(\cdot)} - \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \frac{1 \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)} \\
& - \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^3 \frac{\lambda \phi(\cdot)}{\sigma^3 \Phi(\cdot)} - \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^2 \frac{1 \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial x_0 \partial \lambda} = & - \sum_{j=1}^k \frac{\beta \phi(\cdot)}{\sigma \Phi(\cdot)} + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)^2 \frac{\lambda^2 \beta \phi(\cdot)}{\sigma^3 \Phi(\cdot)} \\
& + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\lambda \beta \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)};
\end{aligned}$$

e

$$\frac{\partial^2 \ell}{\partial x_0^2} = - \frac{k \beta^2}{\sigma^2} - \sum_{j=1}^k \frac{\lambda^3 \beta^2}{\sigma^3} (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} - \sum_{j=1}^k \frac{\lambda^2 \beta^2 \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)}.$$

com $(.) = \lambda(y_i - \alpha - \beta x_i)/\sigma$ e $(..) = \lambda(y_{0j} - \alpha - \beta x_0)/\sigma$. Finalmente, tomando as esperanças destas quantidades e multiplicando-as por menos um, temos as seguintes componentes da informação de Fisher esperada, designadas por

$$\begin{aligned}
I_{\alpha\alpha} &= (n+k)\frac{1+\lambda^2 a_0}{\sigma^2}; \quad I_{\alpha\beta} = \left(\sum_{i=1}^n x_i + kx_0\right)\frac{1+\lambda^2 a_0}{\sigma^2}; \\
I_{\alpha\sigma} &= (n+k)\frac{\sqrt{\frac{2}{\pi}}\frac{\lambda(1+2\lambda^2)}{(1+\lambda^2)^{3/2}} + \lambda^2 a_1}{\sigma^2}; \quad I_{\alpha\lambda} = (n+k)\frac{\sqrt{\frac{2}{\pi}}(1+\lambda^2)^{-3/2} - \lambda a_1}{\sigma}; \\
I_{\alpha x_0} &= k\beta\frac{(1+\lambda^2 a_0)}{\sigma^2}; \quad I_{\beta\beta} = \left(\sum_{i=1}^n x_i^2 + kx_0^2\right)\frac{1+\lambda^2 a_0}{\sigma^2}; \\
I_{\beta\sigma} &= \left(\sum_{i=1}^n x_i + kx_0\right)\frac{\sqrt{\frac{2}{\pi}}\frac{\lambda(1+2\lambda^2)}{(1+\lambda^2)^{3/2}} + \lambda^2 a_1}{\sigma^2}; \quad I_{\beta\lambda} = \left(\sum_{i=1}^n x_i + kx_0\right)\frac{\sqrt{\frac{2}{\pi}}(1+\lambda^2)^{-3/2} - \lambda a_1}{\sigma}; \\
I_{\beta x_0} &= k\beta x_0\frac{1+\lambda^2 a_0}{\sigma^2} + k\frac{\sqrt{\frac{2}{\pi}}\lambda(1+\lambda^2)^{1/2}}{\sigma}; \quad I_{\sigma\sigma} = (n+k)\frac{2+\lambda^2 a_2}{\sigma^2}; \\
I_{\sigma\lambda} &= -(n+k)\frac{\lambda}{\sigma} a_2; \quad I_{\sigma x_0} = k\beta\frac{\sqrt{\frac{2}{\pi}}\frac{\lambda(1+2\lambda^2)}{(1+\lambda^2)^{3/2}} + \lambda^2 a_1}{\sigma^2}; \\
I_{\lambda\lambda} &= (n+k)a_2; \quad I_{\lambda x_0} = k\beta\frac{\sqrt{\frac{2}{\pi}}(1+\lambda^2)^{-3/2} - \lambda a_1}{\sigma}; \\
\text{e } I_{x_0 x_0} &= k\beta^2\frac{1+\lambda^2 a_0}{\sigma^2}
\end{aligned}$$

onde $a_m = E_Z \left[Z^m \left(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)} \right)^2 \right]$, $m = 0, 1, 2$, não depende da primeira ou segunda parte do experimento de calibração, levando a seguinte matriz de informação de Fisher esperada:

$$\begin{bmatrix}
I_{\alpha\alpha} & I_{\alpha\beta} & I_{\alpha\lambda} & I_{\alpha\sigma} & I_{\alpha x_0} \\
I_{\beta\alpha} & I_{\beta\beta} & I_{\beta\lambda} & I_{\beta\sigma} & I_{\beta x_0} \\
I_{\sigma\alpha} & I_{\sigma\beta} & I_{\sigma\lambda} & I_{\sigma\sigma} & I_{\sigma x_0} \\
I_{\lambda\alpha} & I_{\lambda\beta} & I_{\lambda\lambda} & I_{\lambda\sigma} & I_{\lambda x_0} \\
I_{x_0\alpha} & I_{x_0\beta} & I_{x_0\lambda} & I_{x_0\sigma} & I_{x_0 x_0}
\end{bmatrix}$$

Note que quando o expoente de Z , digamos r , em $a_{rs} = E_Z \left[Z^r \left(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)} \right)^s \right]$, é ímpar e o expoente de $\frac{\phi(\lambda Z)}{\Phi(\lambda Z)}$, digamos s , é igual a um, a esperança do produto de ambas é sempre zero, e será maior que zero quando ambos os expoentes forem pares. Rodriguez (2005) desenvolveu aproximações para calcular estas esperanças.

Informação observada do modelo t-normal-assimétrico

No caso do modelo t-normal-assimétrico, sendo $(\cdot) = (\lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma})$ e $(\cdot\cdot) = (\lambda \frac{(y_{0j} - \alpha - \beta x_0)})$. Calculamos as primeiras e segundas derivadas da função de log-verossimilhança, que resulta nas expressões abaixo

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \left(\frac{\nu + 1}{\nu} \right) \sum_{i=1}^n \left(1 + \frac{(y_i - \alpha - \beta x_i)^2}{\nu \sigma^2} \right)^{-1} \frac{(y_i - \alpha - \beta x_i)}{\sigma^2} - \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &+ \left(\frac{\nu + 1}{\nu} \right) \sum_{j=1}^k \left(1 + \frac{(y_{0j} - \alpha - \beta x_0)^2}{\nu \sigma^2} \right)^{-1} \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^2} - \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} \end{aligned} \quad (\text{B.1})$$

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \left(\frac{\nu + 1}{\nu} \right) \sum_{i=1}^n \left(1 + \frac{(y_i - \alpha - \beta x_i)^2}{\nu \sigma^2} \right)^{-1} x_i \frac{(y_i - \alpha - \beta x_i)}{\sigma^2} - \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n x_i \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &+ \left(\frac{\nu + 1}{\nu} \right) \sum_{j=1}^k \left(1 + \frac{(y_{0j} - \alpha - \beta x_0)^2}{\nu \sigma^2} \right)^{-1} x_0 \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^2} - \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k x_0 \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} \end{aligned} \quad (\text{B.2})$$

$$\frac{\partial \ell}{\partial x_0} = \left(\frac{\nu + 1}{\nu} \right) \sum_{j=1}^k \left(1 + \frac{(y_{0j} - \alpha - \beta x_0)^2}{\nu \sigma^2} \right)^{-1} \beta \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^2} - \left(\frac{\lambda \beta}{\sigma} \right) \sum_{j=1}^k \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} \quad (\text{B.3})$$

$$\frac{\partial \ell}{\partial \lambda} = \left(\frac{1}{\sigma} \right) \sum_{i=1}^n (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} + \left(\frac{1}{\sigma} \right) \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} \quad (\text{B.4})$$

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma} = & -\left(\frac{n}{\sigma}\right) + \left(\frac{\nu+1}{\nu}\right) \sum_{i=1}^n \left(1 + \frac{(y_i - \alpha - \beta x_i)^2}{\nu \sigma^2}\right)^{-1} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^3} - \left(\frac{\lambda}{\sigma^2}\right) \sum_{i=1}^n (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} \\ & - \left(\frac{k}{\sigma}\right) + \left(\frac{\nu+1}{\nu}\right) \sum_{j=1}^k \left(1 + \frac{(y_{0j} - \alpha - \beta x_0)^2}{\nu \sigma^2}\right)^{-1} \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^3} - \left(\frac{\lambda}{\sigma^2}\right) \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} \end{aligned} \quad (\text{B.5})$$

Note que para a obtenção dos estimadores foi desenvolvido um algoritmo numérico, porém as variâncias dos estimadores dependerão do inverso da informação de Fisher. Portanto, o cálculo das segundas derivadas será necessário. Vamos calcular, então, as segundas derivadas do vetor $\boldsymbol{\theta} = (\alpha, \beta, x_0, \sigma^2, \lambda)^\top$, onde $(\cdot) = \left(\lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma}\right)$ e $(\cdot\cdot) = \left(\lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma}\right)$. Teremos:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha^2} = & -\frac{n}{\sigma^2} \left(\frac{\nu+1}{\nu}\right) - \left(\frac{\lambda}{\sigma}\right)^2 \left\{ \sum_{i=1}^n \lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{i=1}^n \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\ & - \frac{k}{\sigma^2} \left(\frac{\nu+1}{\nu}\right) - \left(\frac{\lambda}{\sigma}\right)^2 \left\{ \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} + \sum_{j=1}^k \frac{\phi^2(\cdot\cdot)}{\Phi^2(\cdot\cdot)} \right\}; \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta^2} = & -\frac{\sum_{i=1}^n x_i^2}{\sigma^2} \left(\frac{\nu+1}{\nu}\right) - \left(\frac{\lambda}{\sigma}\right)^2 \left\{ \sum_{i=1}^n x_i^2 \lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{i=1}^n x_i^2 \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\ & - \frac{k x_0^2}{\sigma^2} \left(\frac{\nu+1}{\nu}\right) - \left(\frac{\lambda}{\sigma}\right)^2 x_0^2 \left\{ \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} + \sum_{j=1}^k \frac{\phi^2(\cdot\cdot)}{\Phi^2(\cdot\cdot)} \right\}; \end{aligned}$$

$$\frac{\partial^2 \ell}{\partial x_0^2} = -\frac{k \beta^2}{\sigma^2} \left(\frac{\nu+1}{\nu}\right) - \left(\frac{\lambda \beta}{\sigma}\right)^2 \left\{ \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} + \sum_{j=1}^k \frac{\phi^2(\cdot\cdot)}{\Phi^2(\cdot\cdot)} \right\};$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \lambda^2} = & -\left(\frac{1}{\sigma}\right) \sum_{i=1}^n (y_i - \alpha - \beta x_i) \left\{ \lambda \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} + \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\ & - \left(\frac{1}{\sigma}\right) \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \left\{ \lambda \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^2} \frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} + \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi^2(\cdot\cdot)}{\Phi^2(\cdot\cdot)} \right\} \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma^2} &= \binom{n}{\sigma^2} - 3 \left(\frac{\nu+1}{\nu} \right) \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^4} + 2 \sum_{i=1}^n \lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma^3} \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&- \sum_{i=1}^n \lambda^2 \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^4} \left\{ \lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\
&+ \binom{k}{\sigma^2} - 3 \left(\frac{\nu+1}{\nu} \right) \sum_{j=1}^k \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^4} + 2 \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^3} \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&- \sum_{j=1}^k \lambda^2 \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^4} \left\{ \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta \partial \alpha} &= -\frac{\sum_{i=1}^n x_i}{\sigma^2} \left(\frac{\nu+1}{\nu} \right) - \left(\frac{\lambda}{\sigma} \right)^2 \left\{ \sum_{i=1}^n x_i \lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{i=1}^n x_i \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\
&- \frac{kx_0}{\sigma^2} \left(\frac{\nu+1}{\nu} \right) - \left(\frac{\lambda}{\sigma} \right)^2 x_0 \left\{ \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{j=1}^k \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\};
\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial x_0 \partial \alpha^2} = -\frac{k\beta}{\sigma^2} \left(\frac{\nu+1}{\nu} \right) - \left(\frac{\lambda}{\sigma} \right)^2 \beta \left\{ \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{j=1}^k \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\};$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda \partial \alpha} &= -\left(\frac{1}{\sigma} \right) \sum_{i=1}^n \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n \left\{ \lambda \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\
&- \left(\frac{1}{\sigma} \right) \sum_{i=1}^n \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n \left\{ \lambda \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^2} \frac{\phi(\cdot)}{\Phi(\cdot)} + \sum_{i=1}^n \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma \partial \alpha} &= -2 \left(\frac{\nu+1}{\nu} \right) \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)}{\sigma^3} + \left(\frac{\lambda}{\sigma^2} \right) \sum_{i=1}^n \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n \left\{ \lambda^2 \frac{(y_i - \alpha - \beta x_i)^2 \phi(\cdot)}{\sigma^3 \Phi(\cdot)} - \lambda \frac{(y_i - \alpha - \beta x_i) \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)} \right\} \\
&- 2 \left(\frac{\nu+1}{\nu} \right) \sum_{j=1}^k \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^3} + \left(\frac{\lambda}{\sigma^2} \right) \sum_{j=1}^k \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k \left\{ \lambda^2 \frac{(y_{0j} - \alpha - \beta x_0)^2 \phi(\cdot)}{\sigma^3 \Phi(\cdot)} - \lambda \frac{(y_{0j} - \alpha - \beta x_0) \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)} \right\};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial x_0 \partial \beta} &= \left(\frac{\nu+1}{\nu} \right) \sum_{j=1}^k \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^2} - \left(\frac{\nu+1}{\nu} \right) \frac{k \beta x_0}{\sigma^2} - \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&- \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k x_0 \left(\frac{\lambda \beta}{\sigma} \right) \left\{ \lambda \frac{(y_{0j} - \alpha - \beta x_0) \phi(\cdot)}{\sigma \Phi(\cdot)} + \sum_{j=1}^k \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda \partial \beta} &= - \left(\frac{1}{\sigma} \right) \sum_{i=1}^n x_i \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n x_i \left\{ \frac{(y_i - \alpha - \beta x_i)^2 \phi(\cdot)}{\sigma^2 \Phi(\cdot)} + \frac{(y_i - \alpha - \beta x_i) \phi^2(\cdot)}{\sigma \Phi^2(\cdot)} \right\} \\
&- \left(\frac{1}{\sigma} \right) \sum_{j=1}^k x_0 \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k x_0 \left\{ \frac{(y_{0j} - \alpha - \beta x_0)^2 \phi(\cdot)}{\sigma^2 \Phi(\cdot)} + \frac{(y_{0j} - \alpha - \beta x_0) \phi^2(\cdot)}{\sigma \Phi^2(\cdot)} \right\};
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma \partial \beta} &= -2 \left(\frac{\nu+1}{\nu} \right) \sum_{i=1}^n x_i \frac{(y_i - \alpha - \beta x_i)}{\sigma^3} + \left(\frac{\lambda}{\sigma^2} \right) \sum_{i=1}^n x_i \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{i=1}^n x_i \left\{ \lambda^2 \frac{(y_i - \alpha - \beta x_i)^2 \phi(\cdot)}{\sigma^3 \Phi(\cdot)} - \lambda \frac{(y_i - \alpha - \beta x_i) \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)} \right\} \\
&- 2 \left(\frac{\nu+1}{\nu} \right) \sum_{j=1}^k x_0 \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^3} + \left(\frac{\lambda}{\sigma^2} \right) \sum_{j=1}^k x_0 \frac{\phi(\cdot)}{\Phi(\cdot)} \\
&+ \left(\frac{\lambda}{\sigma} \right) \sum_{j=1}^k x_0 \left\{ \lambda^2 \frac{(y_{0j} - \alpha - \beta x_0)^2 \phi(\cdot)}{\sigma^3 \Phi(\cdot)} - \lambda \frac{(y_{0j} - \alpha - \beta x_0) \phi^2(\cdot)}{\sigma^2 \Phi^2(\cdot)} \right\};
\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial \lambda \partial x_0} = -\frac{\beta}{\sigma} \sum_{j=1}^k \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} - \left(\frac{\lambda \beta}{\sigma}\right)^2 \sum_{j=1}^k \left\{ \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\};$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \sigma \partial \beta} &= -2 \left(\frac{\nu+1}{\nu}\right) \sum_{j=1}^k \beta \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^3} + \left(\frac{\lambda \beta}{\sigma^2}\right) \sum_{j=1}^k x_0 \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &+ \left(\frac{\lambda \beta}{\sigma}\right) \sum_{j=1}^k \left\{ \lambda^2 \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^3} \frac{\phi(\cdot)}{\Phi(\cdot)} - \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma^2} \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\}; \quad e \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \sigma \partial \lambda} &= -\left(\frac{1}{\sigma^2}\right) \sum_{i=1}^n (y_i - \alpha - \beta x_i) \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &+ \left(\frac{1}{\sigma}\right) \sum_{i=1}^n \lambda \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2} \left\{ \lambda \frac{(y_i - \alpha - \beta x_i)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\} \\ &- \left(\frac{1}{\sigma^2}\right) \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0) \frac{\phi(\cdot)}{\Phi(\cdot)} \\ &+ \left(\frac{1}{\sigma}\right) \sum_{j=1}^k \lambda \frac{(y_{0j} - \alpha - \beta x_0)^2}{\sigma^2} \left\{ \lambda \frac{(y_{0j} - \alpha - \beta x_0)}{\sigma} \frac{\phi(\cdot)}{\Phi(\cdot)} + \frac{\phi^2(\cdot)}{\Phi^2(\cdot)} \right\}. \end{aligned}$$

As esperanças dessas quantidades formarão a matriz de informação esperada

$$\begin{bmatrix} I_{\alpha\alpha} & I_{\alpha\beta} & I_{\alpha\lambda} & I_{\alpha\sigma} & I_{\alpha x_0} \\ I_{\beta\alpha} & I_{\beta\beta} & I_{\beta\lambda} & I_{\beta\sigma} & I_{\beta x_0} \\ I_{\sigma\alpha} & I_{\sigma\beta} & I_{\sigma\lambda} & I_{\sigma\sigma} & I_{\sigma x_0} \\ I_{\lambda\alpha} & I_{\lambda\beta} & I_{\lambda\lambda} & I_{\lambda\sigma} & I_{\lambda x_0} \\ I_{x_0\alpha} & I_{x_0\beta} & I_{x_0\lambda} & I_{x_0\sigma} & I_{x_0 x_0} \end{bmatrix}$$

que poderemos simplificar utilizando a notação matricial para o vetor de parâmetros da regressão $\beta = (\alpha; \beta)$. Essa simplicidade favorece a comparação da informação de Fisher do modelo de calibração com a do modelo de regressão dado em Gómez et al. (2007)

Comandos: modelo t-normal-assimétrico, abordagem clássica

Seguem os comandos que foram utilizados, na linguagem R, para a obtenção dos estimadores do modelo de calibração t-normal-assimétrico, segundo a abordagem freqüentista, a fim de facilitar o manuseio dos conceitos desenvolvidos nesta tese.

```
##### APLICAÇÃO TESTIS #####
```

```
rm(list=ls(all=TRUE))
```

```
#-----
```

```
# Entrada dos dados
```

```
#-----
```

```
"testis.right" <- structure(list(us =  
c(5,5.7,7.4,2.6,5.7,6.1,6.2,10.4,9.1,14.8,9.6,15.7,3,17.6,10,4.1,  
12.7,2.7,10.2,16.5,4.5,5.6,11,9.2,8.5,5.4,6.7,5.3,20,18.8,13.9,
```



```

9.4,9.1,14.1,9.3,20.9,11.5,9.7,13.7,8.9),
      ii = c(5.9,4.8,6.8,3.1,5,4.4,6,8.8,7.9,13,8.2,
             19.8,2,17.3,7.9,4.4,11.4,4.1,11.1,15.3,
             3.9,4.5,9.7,11.3,8.8,6.1,7.2,5.9,16.3,
             16.3,12.2,10.3,10.8,13,8.4,22.1,10.6,
             9.7,11.6,8.1)
    )
  )
attach(testis.right)
x<-us y<-ii
y0<-c(10.3,17.3)

#-----
# Gerar Matrizes
#-----

X<-cbind(1,x)
y<-cbind(y)
y0<-cbind(y0)
n<-length(y)
k<-length(y0)

#-----
# Valores Iniciais
#-----

b0<-(solve(t(X)%*%X))%*%(t(X)%*%(y))

```

```

x0ini<-(mean(y)-b0[1,1])/b0[2,1]
X0<-matrix(rep(cbind(1,x0ini),k),k,2,byrow=T)
beta0<- b0
sigma0<-sqrt(sum((y-mean(y))^2)/(n-2))
lambda0<-n/((n-1)*(n-2))*sum((y-mean(y))/sd(y))^3

##### E-step #####

#----- # Quociente dos phi's #-----

phi<-cbind(dnorm(lambda0*(y-X%*%beta0)/sigma0, 0, 1))
Phi<-cbind(pnorm(lambda0*(y-X%*%beta0)/sigma0, 0, 1))

phij<-cbind(dnorm(lambda0*(y0-X0%*%beta0)/sigma0, 0, 1))
Phij<-cbind(pnorm(lambda0*(y0-X0%*%beta0)/sigma0, 0, 1))

#-----
# Valores Iniciais para os vetores de atualizações
#-----

beta1<-0
alpha1<-0
sigma1<-0
lambda1<-0
x01 <-0
nu<-4 # variar os graus de liberdade, quando for preciso

```

```

#----- # Outros valores iniciais #-----

j<-0
cleber<-0

#----- # Critério de parada #-----

crit<-0.03

#----- # Esperanças condicionais iniciais #-----

t <- lambda0*(y-X%*%beta0)+ (phi/Phi)*sigma0

t2 <- lambda0^2*(y-X%*%beta0)^2+ sigma0^2
      + (phi/Phi)* lambda0*(y-X%*%beta0)*sigma0

w <- (nu +1)*1/(((y-X%*%beta0)/sigma0)^2+nu)

t0 <- lambda0*(y0-X0%*%beta0)+ (phij/Phij)*sigma0

t02 <- lambda0^2*(y0-X0%*%beta0)^2+ sigma0^2
      + (phij/Phij)* lambda0*(y0-X0%*%beta0)*sigma0

w0 <- (nu +1)*1/(((y0-X0%*%beta0)/sigma0)^2+nu)

#----- # Eis o programa! #-----

```

```

while(cleber==0){

    if(j!=0){

        beta0<-beta1

        sigma0<-sigma1

        lambda0<-lambda1

        x0ini <- x01

        t <- t.atualizado
        t2 <- t2.atualizado
        w <- w.atualizado
        t0 <- t0.atualizado
        t02 <- t02.atualizado
        w0 <- w0.atualizado

    }

##### M-step #####

#-----
# Atualizações dos parâmetros
#-----

W <- diag(as.vector(w+lambda0^2), n, n )
U <- diag(as.vector(w), n, n )
U0 <- diag(as.vector(w0), k, k )
X0<-matrix(rep(cbind(1,x0ini),k),k,2,byrow=T)

```

```
beta1 <- solve(t(X)%*%W%*%X)%*%t(X)%*%(w*y-lambda0*(t-lambda0*y))
```

```
sigma1 <- as.vector(1/sqrt(2*(n+k))*sqrt((t(y-X%*%beta0)%*%U%*%
(y-X%*%beta0) + sum(t2) -2*lambda0*t(t)%*%(y-X%*%beta0)
+ lambda0^2*t(y-X%*%beta0)%*%(y-X%*%beta0))
+ t(y0-X0%*%beta0)%*%U0%*%(y0-X0%*%beta0) + sum(t02)
- 2*lambda0*t(t0)%*%(y0-X0%*%beta0)
+ lambda0^2*t(y0-X0%*%beta0)%*%(y0-X0%*%beta0)))
```

```
lambda1 <- as.vector((t(t)%*%(y-X%*%beta0) + t(t0)%*%
(y0-X0%*%beta0))%*%solve(t(y-X%*%beta0)%*%(y-X%*%beta0)
+ t(y0-X0%*%beta0)%*%(y0-X0%*%beta0)))
```

```
x01<- (sum((w0 + lambda0^2)*(y0 - beta0[1,]))
- lambda0*sum(t0))/sum((w0 + lambda0^2)* beta0[2,])
```

```
X01<-matrix(rep(cbind(1,x01),k),k,2,byrow=T)
```

```
##### E-step #####
```

```
#-----
```

```
# Atualizações das esperanças
```

```
#-----
```

```
phi<-cbind(dnorm(lambda1*(y-X%*%beta1)/sigma1, 0, 1))
```

```
Phi<-cbind(pnorm(lambda1*(y-X%*%beta1)/sigma1, 0, 1))
```

```

phij<-cbind(dnorm(lambda1*(y0-X01%%beta1)/sigma1, 0, 1))
Phi j<-cbind(pnorm(lambda1*(y0-X01%%beta1)/sigma1, 0, 1))

t.atualizado <- lambda1*(y-X%%beta1)+ (phi/Phi)*sigma1

t2.atualizado <- lambda1^2*(y-X%%beta1)^2+ sigma1^2
                + (phi/Phi)* lambda1*(y-X%%beta1)*sigma1

w.atualizado <- (nu +1)*1/(((y-X%%beta1)/sigma1)^2+nu)

t0.atualizado <- lambda1*(y0-X01%%beta1)+ (phij/Phi j)*sigma1

t02.atualizado <- lambda1^2*(y0-X01%%beta1)^2+ sigma1^2
                + (phij/Phi j)* lambda1*(y0-X01%%beta1)*sigma1

w0.atualizado <- (nu +1)*1/(((y0-X01%%beta1)/sigma1)^2+nu)

convergiu<-(Mod(beta1[1,]-beta0[1,])<crit
            & Mod(beta1[2,]-beta0[2,])<crit
            & Mod(sigma1-sigma0)<crit
            & Mod(lambda1-lambda0)<crit
            & Mod(x01-x0ini)<crit)

print(cbind(convergiu,j,
            Mod(beta1[1,]-beta0[1,]),
            Mod(beta1[2,]-beta0[2,]),
            Mod(sigma1-sigma0),

```

```

Mod(lambda1-lambda0),
Mod(x01-x0ini))

if(convergiu == "TRUE")
  {cleber<-1}
  j<-j+1
}

resultado<- cbind(j,beta1[1,],beta1[2,],sigma1,lambda1, x01)
t(resultado)

##### CRITÉRIOS DE INFORMAÇÃO #####

cte <- 2*gamma(nu/2+0.5)/(sqrt(pi*nu)*gamma(nu/2))
log.likelihood.1<- n*log(cte) - n/2*log(sigma1^2)
  -(nu+1)/2*sum(log(1 + ((y-X%%beta1)^2)/(nu*sigma1^2)))
  + sum(log( pnorm(lambda1/sigma1*(y-X%%beta1), 0, 1) ))
log.likelihood.2<-k*log(cte) - k/2*log(sigma1^2)
  -(nu+1)/2*sum(log(1 + ((y0-X01%%beta1)^2)/(nu*sigma1^2)))
  + sum(log( pnorm(lambda1/sigma1*(y0-X01%%beta1), 0, 1) ))
log.like<-log.likelihood.1+log.likelihood.2

crit.AIC<- -2*log.like + 2*5
crit.BIC<- -2*log.like + log(n+k)*5
crit.HQ<- -2*log.like + 2*log(log(n+k)*5)

```

Comandos: modelo

t-normal-assimétrico, abordagem bayesiana

```
# UNIFORME

model
{
  # Verossimilhanca aumentada
  for( i in 1 : N ) {
    y[i] ~ dnorm(a[i],b[i])
    a[i]<- alpha+beta*x[i]+(1/(w[i]*sqrt(tau)))*
      (lambda/sqrt(1+pow(lambda,2)/w[i]))*t[i]
    b[i]<- (w[i]+pow(lambda,2))*tau
    t[i] ~ dnorm(0, 1 ) I(0,)
    w[i]~dgamma(vv,vv)
  }
  #
  centralx[i]<- x[i]-x.bar
}
```



```

centrally[i]<- y[i]-y.bar
Numxy[i]<- centralx[i]*centrally[i]
Denxx[i]<-pow(centralx[i],2)
Res[i]<-y[i] - mua - mub*x[i]
Res2[i]<-pow(Res[i],2)
#
}
for( j in 1 : J ) {
y0[j] ~ dnorm(a0[j],b0[j])
a0[j]<- alpha+beta*x0+(1/(w0[j]*sqrt(tau)))*
      (lambda/sqrt(1+pow(lambda,2)/w0[j]))*t0[j]
b0[j]<- (w0[j]+pow(lambda,2))*tau
t0[j] ~ dnorm(0, 1 ) I(0,)
w0[j]~dgamma(vv,vv)
}
# Preparações para os hiperparâmetros fixos
x.bar <- mean(x[])
y.bar <- mean(y[])
y0.bar <- mean(y0[])
Sxy <- sum(Numxy[1:N])
Sxx <- sum(Denxx[1:N])
mub<-Sxy/Sxx
mua<-y.bar - x.bar*mub
QMR<-sum(Res2[1:N])/(N-2)
vara<-1/(QMR*(1/N+pow(x.bar,2)/Sxx))
varb<-1/(QMR/Sxx)
x1<-(y0.bar - mua)/mub

```

```

x2<-1/(pow(sd(x[]),2))
#
alpha ~ dnorm(mua, vara)
beta ~ dnorm(mub, varb)
x0 ~ dnorm(x1,x2)
tau ~ dgamma(0.05,1)
#
sigma<-sqrt(1/tau)
lambda ~ dnorm(0,zz)
zz<-2*z
z ~ dgamma(1,1)
vv<-2
}

# JEFFREYS

model
{
# Verossimilhança aumentada utilizando
# a representação hierárquica da normal assimétrica
for( i in 1 : N ) {
y[i] ~ dnorm(a[i],b[i])
a[i]<- alpha+beta*x[i]+(1/(w[i]*sqrt(tau)))*
      (lambda/sqrt(1+pow(lambda,2)/w[i]))*t[i]
b[i]<- (w[i]+pow(lambda,2))*tau
t[i] ~ dnorm(0, 1 ) I(0,)
w[i]~dgamma(g,g)
}
}

```

```

#
centralx[i]<- x[i]-x.bar
centrally[i]<- y[i]-y.bar
Numxy[i]<- centralx[i]*centrally[i]
Denxx[i]<-pow(centralx[i],2)
Res[i]<-y[i] - mua - mub*x[i]
Res2[i]<-pow(Res[i],2)
#
}
for( j in 1 : J ) {
y0[j] ~ dnorm(a0[j],b0[j])
a0[j]<- alpha+beta*x0+(1/sqrt(tau))*
      (lambda/sqrt(1+pow(lambda,2)))*t0[j]
b0[j]<- (1+pow(lambda,2))*tau
t0[j] ~ dnorm(0, 1 ) I(0,)
w0[j]~dgamma(g,g)
}
# Preparações para os hiperparâmetros fixos
x.bar <- mean(x[])
y.bar <- mean(y[])
y0.bar <- mean(y0[])
Sxy <- sum(Numxy[1:N])
Sxx <- sum(Denxx[1:N])
mub<-Sxy/Sxx
mua<-y.bar - x.bar*mub
QMR<-sum(Res2[1:N])/(N-2)
vara<-1/(QMR*(1/N+pow(x.bar,2)/Sxx))

```

```

varb<-1/(QMR/Sxx)
x1<-(y0.bar - mua)/mub
x2<-1/(pow(sd(x[]),2))
#
alpha ~ dnorm(mua, vara)
beta ~ dnorm(mub, varb)
x0 ~ dnorm(x1,x2)
tau ~ dgamma(0.05,1)
#
sigma<-sqrt(1/tau)
# Especificação a priori (JEFFREYS)
# para o parâmetro de assimetria
lambda ~ dnorm(0,ww)
ww<-0.4052847*h
h ~ dgamma(0.25,0.25)
g<-2
}

list(
y=c(29.8, 37.7, 46, 33.2, 38.7, 46.9, 33.7, 38.8, 48, 35.3, 39.6, 49.3,
35.5, 41, 51.7, 42.8, 62.6, 36.2, 42.8, 69.8, 36.3, 43.5, 79.5, 45.6, 80),
x = c(5.9, 6.5, 7.4, 7.2, 7, 7.7, 7.3, 6.3, 9.7, 6.3, 7.9, 7.8,
8.1, 9, 7.7, 8.2, 11.6, 7, 8.7, 11.3, 7.6, 7.8, 11.8, 9.7, 10.7) ,
y0 =c(36.1, 37.5) , N=25, J=2)

list(alpha = 0, beta = 0, tau = 1, lambda = 0)

```

DIC, ADIC e EDIC

Para o cálculo dos critérios DIC, ADIC e EDIC é necessário realizar uma leitura das saídas do WinBugs. Supondo que, após clicar no botão CODA do WinBugs, você tenha guardado cada saída em um arquivo .dat, então, para a leitura dos dados você deve fazer:

```
##### Leitura das Cadeias de Markov #####
```

```
lambda1<-scan("lambdauni.dat", what = list(codigo1 = 0, lambda =  
0))  
attach(lambda1)  
alpha1<-scan("alfauni.dat", what = list(codigo2 = 0, alpha = 0))  
attach(alpha1)  
beta1<-scan("betauni.dat", what = list(codigo3 = 0, beta = 0))  
attach(beta1)  
sigma1<-scan("sigmauni.dat", what = list(codigo4 = 0, sigma = 0))  
attach(sigma1)  
x01<-scan("x0uni.dat", what = list(codigo5 = 0, x0 = 0))  
attach(x01)
```

```
##### Leitura dos Dados #####

y<-c(29.8, 37.7, 46, 33.2, 38.7, 46.9, 33.7, 38.8, 48, 35.3, 39.6,
49.3, 35.5, 41, 51.7, 42.8, 62.6, 36.2, 42.8, 69.8, 36.3, 43.5,
79.5, 45.6, 80)

x<-c(5.9, 6.5, 7.4, 7.2, 7, 7.7, 7.3, 6.3, 9.7, 6.3, 7.9, 7.8,
8.1, 9, 7.7, 8.2, 11.6, 7, 8.7, 11.3, 7.6, 7.8, 11.8, 9.7, 10.7)

y0<-c(36.1, 37.5)

##### Comprimento dos vetores #####

p<-length(y) mc<-length(alpha)

##### Cálculo de D barra #####

nu<-2 cte <- 2*gamma(nu/2+0.5)/(sqrt(pi*nu)*gamma(nu/2))
veros<-matrix(0,mc,p) for(j in 1:p ){
  for(i in 1:mc ){
    if(j == 1){
      veros[i,j]<-1/sigma[i]*(1+ ((y[j]-alpha[i]-beta[i]*x[j])
/sigma[i])^2/nu)^(-(nu/2+0.5))*1/sigma[i]*(1+ ((y0[1]-
alpha[i]-beta[i]*x0[i])/sigma[i])^2/nu)^(-(nu/2+0.5))*
1/sigma[i]*(1+ ((y0[2]-alpha[i]-beta[i]*x0[i])/sigma[i])^2/nu)
^(-(nu/2+0.5))*pnorm((lambda[i])*(y[j]-alpha[i]-beta[i]*x[j])
/sigma[i], 0, 1)*pnorm((lambda[i])*(y0[1]-alpha[i]-beta[i]
```

```

        *x0[i])/sigma[i], 0,1)*pnorm((lambda[i])*(y0[2]-alpha[i]
        -beta[i]*x0[i])/sigma[i], 0, 1)
    }
if(j != 1){
    veros[i,j]<-1/sigma[i]*(1+ ((y[j]-alpha[i]-beta[i]*x[j])
    /sigma[i])^2/nu)^(-(nu/2+0.5))*pnorm((lambda[i])
    *(y[j]-alpha[i]-beta[i]*x[j])/sigma[i], 0, 1)
    }
}
}
log.like<-log(veros) deviance<-matrix(0,mc,1)

for(k in 1:mc){
    deviance[k,1]<--2*sum(log.like[k,1:p])
}
Deviance.barra<-mean(deviance)

##### Estimadores de Bayes sob perda quadrática #####

alpha.bar<-mean(alpha)
beta.bar<-mean(beta)
sigma.bar<-mean(sigma)
x0.bar<-mean(x0)
lambda.bar<-mean(lambda)

##### Cálculo de D( theta barra ) #####

```

```

parcela<-array(0,p) for(j in 1:p ){
  if(j == 1){
    parcela[j]<-1/sigma.bar*(1+ ((y[j]-alpha.bar-beta.bar*x[j])
    /sigma.bar)^2/nu)^(-(nu/2+0.5))*1/sigma.bar*(1+ ((y0[1]
    -alpha.bar-beta.bar*x0.bar)/sigma.bar)^2/nu)^(-(nu/2+0.5))*1
    /sigma.bar*(1+ ((y0[2]-alpha.bar-beta.bar*x0.bar)/sigma.bar)
    ^2/nu)^(-(nu/2+0.5))*pnorm((lambda.bar)*(y[j]-alpha.bar
    -beta.bar*x[j])/sigma.bar, 0, 1)*pnorm((lambda.bar)*(y0[1]
    -alpha.bar-beta.bar*x0.bar)/sigma.bar, 0, 1)
    *pnorm((lambda.bar)*(y0[2]-alpha.bar-beta.bar*x0.bar)
    /sigma.bar, 0, 1)
  }
  if(j != 1){
    parcela[j]<-1/sigma.bar*(1+ ((y[j]-alpha.bar
    -beta.bar*x[j])/sigma.bar)^2/nu)^(-(nu/2+0.5))
    *pnorm((lambda.bar)*(y[j]-alpha.bar-beta.bar*x[j])
    /sigma.bar, 0, 1)
  }
} Deviance.theta.barra<--2*sum(log(parcela))

```

```
##### Cálculo de pD e DIC #####
```

```
pD<-Deviance.barra - Deviance.theta.barra
```

```
DIC<-2*Deviance.barra - Deviance.theta.barra
```

```
#####
```

```
##### Cálculos para o ADIC #####
```



```
#####

Deviance.med<-median(deviance)

##### Estimadores de Bayes sob perda absoluta #####

alpha.med<-median(alpha)
beta.med<-median(beta)
sigma.med<-median(sigma)
x0.med<-median(x0)
lambda.med<-median(lambda)

##### Cálculo de D( theta median ) #####

parcela<-array(0,p) for(j in 1:p ){
  if(j == 1){
    parcela[j]<-1/sigma.med*(1+ ((y[j]-alpha.med-beta.med*x[j])
    /sigma.med)^2/nu)^(-(nu/2+0.5))*1/sigma.med*(1+ ((y0[1]
    -alpha.med-beta.med*x0.med)/sigma.med)^2/nu)^(-(nu/2+0.5))
    *1/sigma.med*(1+ ((y0[2]-alpha.med-beta.med*x0.med)/sigma.med)
    ^2/nu)^(-(nu/2+0.5))*pnorm((lambda.med)*(y[j]-alpha.med
    -beta.med*x[j])/sigma.med, 0, 1)*pnorm((lambda.med)*(y0[1]
    -alpha.med-beta.med*x0.med)/sigma.med, 0, 1)
    *pnorm((lambda.med)*(y0[2]-alpha.med-beta.med*x0.med)
    /sigma.med, 0, 1)
  }
}
```

```

if(j != 1){
  parcela[j]<-1/sigma.med*(1+ ((y[j]-alpha.med-beta.med*x[j])
  /sigma.med)^2/nu)^(-(nu/2+0.5))*pnorm((lambda.med)*(y[j]
  -alpha.med-beta.med*x[j])/sigma.med, 0, 1)
}
} Deviance.theta.med<--2*sum(log(parcela))

##### Cálculo de pD e ADIC #####

pD<-Deviance.med - Deviance.theta.med
ADIC<-2*Deviance.med - Deviance.theta.med

#####
##### Cálculos para o EDIC #####
#####

Deviance.mod<-mlv(deviance, method = "naive", bw = 1/2)$M

##### Estimadores de Bayes sob perda 0-1 #####

alpha.mod<-mlv(alpha, method = "naive", bw = 1/2)$M
beta.mod<-mlv(beta, method = "naive", bw = 1/2)$M
sigma.mod<-mlv(sigma, method = "naive", bw = 1/2)$M
x0.mod<-mlv(x0, method = "naive", bw = 1/2)$M
lambda.mod<-mlv(lambda, method = "naive", bw = 1/2)$M

##### Cálculo de D( theta mode ) #####

```

```

parcela<-array(0,p) for(j in 1:p ){
  if(j == 1){
    parcela[j]<-1/sigma.mod*(1+ ((y[j]-alpha.mod-beta.mod*x[j])
    /sigma.mod)^2/nu)^(-(nu/2+0.5))*1/sigma.mod*(1+ ((y0[1]
    -alpha.mod-beta.mod*x0.mod)/sigma.mod)^2/nu)^(-(nu/2+0.5))
    *1/sigma.mod*(1+ ((y0[2]-alpha.mod-beta.mod*x0.mod)
    /sigma.mod)^2/nu)^(-(nu/2+0.5))*pnorm((lambda.mod)*(y[j]
    -alpha.mod-beta.mod*x[j])/sigma.mod, 0, 1)*pnorm((lambda.mod)
    *(y0[1]-alpha.mod-beta.mod*x0.mod)/sigma.mod, 0, 1)
    *pnorm((lambda.mod)*(y0[2]-alpha.mod-beta.mod*x0.mod)
    /sigma.mod, 0, 1)

  }
  if(j != 1){
    parcela[j]<-1/sigma.mod*(1+ ((y[j]-alpha.mod-beta.mod*x[j])
    /sigma.mod)^2/nu)^(-(nu/2+0.5))*pnorm((lambda.mod)*(y[j]
    -alpha.mod-beta.mod*x[j])/sigma.mod, 0, 1)
  }
}
Deviance.theta.mod<--2*sum(log(parcela))

##### Cálculo de pD e EDIC #####

pD<-Deviance.mod - Deviance.theta.mod
EDIC<-2*Deviance.mod - Deviance.theta.mod

```

Referências Bibliográficas

- [1] ACHIC, B. G. B. *Calibração controlada aplicada à química analítica*. São Paulo, 2005. Dissertação (Mestrado) - Departamento de Estatística, IME-USP.
- [2] ANDEL, D. J.; NETUKA, I.; ZVARA, K. On threshold autoregressive process. *Kybernetika*, 20, p. 89-106. 1984.
- [3] ARELLANO-VALLE, R. B.; AZZALINI, A. The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99, Issue 7, p. 1362-1382. 2008.
- [4] ARELLANO-VALLE, R. B.; OZAN, S.; BOLFARINE, H.; LACHOS, V. H. Skew-normal measurement error models. *Technical Report*. IME-USP: 2003-10. 2003.
- [5] ARELLANO-VALLE, R. B.; BOLFARINE, H.; LACHOS, V. H. Asymetric mixed linear models. *RT-MAE*, 2004-04.
- [6] ARELLANO-VALLE, R. B.; BRANCO, M. D. Distribuições Elípticas Assimétricas. Caxambu. XVI SINAPE, 2004.
- [7] ARELLANO-VALLE, R. B.; OZAN, S.; BOLFARINE, H.; LACHOS, V. H. Skew normal measurement error models. *Journal of Multivariate Analysis*, 96, p. 265-281. 2005.
- [8] ARELLANO-VALLE, R. B.; BOLFARINE, H.; LACHOS, V. H. Skew-normal linear mixed models. *Journal of Data Science*, 3, 4, p. 415-438. 2005.
- [9] AZZALINI, A. A class of distributions which includes the normal one. *Scandinavian Journal of Statistics*, 12, p. 171-178. 1985.

- [10] AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew-normal distribution, *Journal of the Royal Statistical Society, Series B*, 61(3): p. 579-602. 1999.
- [11] AZZALINI, A.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution, *Journal of the Royal Statistical Society: Series B* 65(2): p. 367-389. 2003.
- [12] AZZALINI, A.; DALLA VALLE, A. The multivariate skew-normal distribution. *Biometrika*, 83, 4, p. 715-726. 1996.
- [13] BAYES, C. L.; BRANCO, M. D. Bayesian inference for the skewness parameter on skew normal distribution. *RT-MAE*, 2005-18. 2005.
- [14] BAYES, C. L.; BRANCO, M. D. Bayesian inference for the skewness parameter on skew normal distribution. *Brazilian Journal of Probability and Statistics*, 21, 2, p. 141-163. 2007.
- [15] BAZÁN, J. L.; BRANCO, M. D.; BOLFARINE, H. A skew item response model. *Bayesian Analysis*, 1, 4, p. 861-892. 2006.
- [16] BRANCO, M. D. *Calibração: uma abordagem bayesiana*. São Paulo, 1997. Tese (Doutorado) - Departamento de Estatística, IME-USP.
- [17] BRANCO, M. D.; BOLFARINE, H.; IGLESIAS, P.; ARELLANO-VALLE, R. B. Bayesian analysis of the calibration problem under elliptical distributions. *Journal of Statistical Planning and Inference*, 90, 1: p. 69-85. 2000.
- [18] BRANCO, M. D.; DEY, D. K. A general class of multivariate skewelliptical distributions, *Journal of Multivariate Analysis*, 79: p. 99-113. 2001.
- [19] BROWN, P. *Measurement, regression and calibration*. Oxford University Press, 1993.
- [20] CHAIBUB NETO, E.; BRANCO, M. D. Bayesian reference analysis for binomial calibration problem. *RT - MAE*, 2003-12: IME - USP. 2003.
- [21] CHIOGNA, M. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods & Applications*, 14, p. 331-341. 2005.

- [22] CHIPKEVITCH, E; NISHIMURA, R. T; TU, D. G. S.; GALEA-ROJAS, M. Clinical measurement of testicular volume in adolescents: comparison of the reliability of 5 methods. *The Journal of Urology*, 156, p. 2050-2053. 1996.
- [23] CONOVER, W. J. *Practical nonparametric statistic*. 3. ed. Nova Iorque: Wiley, 1999.
- [24] COSTA, A. M.; CALDAS, L. V. E. Intercomparação e calibração de medidores de atividades utilizados em serviços de medicina nuclear. *Radiol Bras*, 36, (5): p. 293-297. 2003.
- [25] DEVORE, J. L. *Probabilidade e estatística para engenharia e ciências*. São Paulo: Thomson Learning, 2006.
- [26] DEMPSTER, A. P; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, p. 1-38. 1977.
- [27] EISENHART, C. The interpretation of certain regression methods, and their use in biological and industrial research. *Annals of Mathematical Statistics*, 10: p. 162-186. 1939.
- [28] FANG, K. T; KOTZ, S.; NG, K. W. *Symmetric multivariate and related distributions*. Nova Iorque: Chapman and Hall, 1990.
- [29] FERREIRA, C. S. *Inferência e diagnósticos em modelos assimétricos*. São Paulo, 2008. Tese (Doutorado) - Departamento de Estatística, IME-USP.
- [30] FIGUEIREDO, C. C; SANDOVAL, M. C.; BOTTER, D. A. Análise de regressão incorporando o esquema amostral. São Paulo, 2004. Tese (Mestrado) - Departamento de Estatística, IME-USP.
- [31] FIGUEIREDO, C. C; SANDOVAL, M. C; BOLFARINE, H.; LIMA, C. R. O. Skew-normal linear calibration: a Bayesian perspective. *Journal of Chemometrics*, 22, p. 472-480. 2008.
- [32] FIGUEIREDO, C. C; BOLFARINE, H; SANDOVAL, M. C.; Lima, C. R. O. On the skew-normal calibration model. *Journal of Applied Statistics*. (in press). 2009.

- [33] FIRTH, D. Bias reduction of maximum likelihood estimates, *Biometrika*, 80, p. 27-38. 1993.
- [34] GODOY, L. G. *A distribuição t-assimétrica univariada: propriedades e inferência*. São Paulo, 2007. Dissertação (Mestrado) - Departamento de Estatística, IME-USP.
- [35] GÓMEZ, H.W; VENEGAS, O.; BOLFARINE, H. Skew-symmetric distributions generated by the normal distribution function. *Environmetrics*, 18, p. 359-363. 2007.
- [36] GUIMARÃES-FILHO, Z,O; HELENE, O. One step self-calibration procedure in gamma-ray energy measurements. *Brazilian Journal of Physics*, 33, 2, p. 280-281. 2003.
- [37] HENZE, N. A probabilistic representation of the skew-normal distribution, *Scandinavian Journal of Statistics*, 13, p. 271-275. 1986.
- [38] HOADLEY, B. A Bayesian look at linear regression. *JASA*, 65: p. 356-369. 1970.
- [39] JOHNSON, N. L; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*. Vol. 1, 2. ed. New York: John Wiley, 1994.
- [40] KRUTCHKOFF, R. Classical and inverse regression methods of calibration. *Technometrics*, 9: p. 425-439. 1967.
- [41] KRUTCHKOFF, R. Classical and inverse regression methods of calibration in extrapolation. *Technometrics*, 11: p. 605-608. 1969.
- [42] LACHOS, V. H. *Modelos lineares mistos assimétricos*. São Paulo, 2004. Tese (Doutorado) - Departamento de Estatística, IME-USP.
- [43] LIMA, C. R. O. *Calibração absoluta com erros nas variáveis*. São Paulo, 1996 Tese (Doutorado) - Departamento de Estatística, IME-USP.
- [44] LIMA, C. R. O. P; SANDOVAL, M. C; BOLFARINE, H.; SOUSA, S. O. Robust Estimation in calibration models using the Student-t distribution. *JASS*. 2008.
- [45] LISEO, B.; LOPERFIDO, N. A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference*. 136(2): p. 373-389. 2006.

- [46] McDONALD J. F. Calibration of a monocentric city model with mixed land use and congestion. *Regional Science and Urban Economics*, 39, 1, p. 90-96. 2009.
- [47] MORETTIN, P. A.; HAZZAN, S.; BUSSAB, W. O. *Cálculo: funções de uma e várias variáveis*. São Paulo: Saraiva, 2003.
- [48] NETER, J; KUTNER, M. H; NACHTSHEIN, C. J.; LI, W. *Applied linear statistical models*. 4. ed. Boston: McGraw-Hill Irwin, 1996.
- [49] PEREIRA, C. A. B.; STERN, J. M. Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy*, 1, p. 99-110. 1999.
- [50] PILLAI, K. G.; HOFACKER, C. Calibration of consumer knowledge of the web. *International Journal of Research in Marketing*, 24, 3, p. 254-267. 2007.
- [51] RODRIGUEZ, C. L. B. *Inferência bayesiana no modelo normal assimétrico*. São Paulo, 2005. Dissertação (Mestrado) - Departamento de Estatística, IME-USP.
- [52] ROTNITZKY, A; COX, D. R; BOTTAI, M.; ROBINS, J. Likelihood-based inference with singular information matrix. *Bernoulli*, 6: p. 243-284. 2000.
- [53] SARTORI, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. Dipartimento di Statistica. Università "Ca'Foscari" di Venezia. Technical Report. 2003.
- [54] SEN, P.K.; SINGER, J. *Large sample methods in statistics: an introduction with application*. Nova Iorque: Chapman Hall, 1993.
- [55] SEVERINI, T. A. *Likelihood methods in statistics*. Nova Iorque: Oxford University Press, 2000.
- [56] SPIEGELHALTER, D. J; BEST, N. G; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. *Journal Royal Statistical Society, Series B*, 64, 3, p. 583-639. 2002.
- [57] STEWART, J. *Cálculo I*. Vol. 1, 5. ed. São Paulo: Pioneira Thomsom Learning, 2006.
- [58] THOMAS, E. V.; GE, N. Development of Robust Multivariate Calibration Models. *Technometrics*, 42, 2, p. 168-177. 2000.

- [59] VENABLES, W. N.; SMITH, D. M. *An Introduction to R. R: A Programming Environment for Data Analysis and Graphics*. Version 2.3.1. 2006.

©Capítulo 1. Direito de publicação reservado ao autor e ao *Journal of Applied Statistics*.

©Capítulo 2. Direito de publicação reservado ao autor e ao *Journal of Chemometrics*.

©Capítulo 3. Todos os direitos reservados ao autor.