

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UMA ANÁLISE BAYESIANA PARA DADOS  
COMPOSICIONAIS

Simone Cristina Obage

Dissertação apresentada ao Departamento  
de Estatística da Universidade Federal de São  
Carlos como parte dos requisitos necessários  
para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Jorge Alberto Achcar

São Carlos  
2005

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

O12ab

Obage, Simone Cristina.

Uma análise bayesiana para dados composicionais /  
Simone Cristina Obage. -- São Carlos : UFSCar, 2007.  
69 p.

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

1. Estatística - análise. 2. Inferência bayesiana. 3. Dados  
composicionais. 4. MCMC. I. Título.

CDD: 519.5 (20<sup>a</sup>)

*Agradeço,*

*à Deus pela força durante todos esses anos de estudo.*

*à minha mãe, meu irmão e ao meu namorado pela paciência apoio e compreensão durante todo esse processo.*

*ao professor Dr. Jorge Alberto Achcar pela orientação, pelas idéias e principalmente pelo exemplo de dedicação e disciplina ao trabalho.*

*ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela assistência financeira.*

*à todos os professores do Departamento de Estatística da UFSCar, em especial aqueles aos quais tive o privilégio de ser aluna.*

*às professoras Dra. Tereza Cristina Dias e Dra. Vera Lucia D. Tomazella pelas correções em meu exame de qualificação,*

*à todos meus colegas de pós-graduação, pela convivência, amizade e apoio durante todo o período de realização do mestrado.*

*aos meus amigos Flávia, Lia, Milton e Elíris, pelo companheirismo e por toda ajuda que recebi.*

# *Resumo*

Dados Composicionais são dados por vetores com elementos positivos cuja soma é um. Exemplos típicos de dados desta natureza são encontrados nas mais diversas áreas; como em geologia, biologia, economia entre outras. Neste trabalho, introduzimos uma análise Bayesiana para dados composicionais considerando as transformações razão log-aditiva e Box-Cox, assumindo a distribuição normal multivariada para erros correlacionados. Estes resultados generalizam uma abordagem bayesiana assumindo erros não correlacionados. Também consideramos o uso da distribuição potência exponencial para erros não correlacionados, assumindo a transformação razão log-aditiva. Nós ilustramos a metodologia proposta considerando um conjunto de dados reais.

**Palavras-chave:** Dados composicionais, Erros correlacionados, Inferência Bayesiana, MCMC.

# *Abstract*

Compositional data are given by vectors of positive numbers with sum equals to one. These kinds of data are common in many applications, as in geology, biology, economy among many others. In this paper, we introduce a Bayesian analysis for compositional data considering additive log-ratio (ALR) and Box-Cox transformations assuming a multivariate normal distribution for correlated errors. These results generalize some existing Bayesian approaches assuming uncorrelated errors. We also consider the use of exponential power distributions for uncorrelated errors considering additive log-ratio (ALR) transformation. We illustrate the proposed methodology considering a real data set.

**keywords:** Compositional data, Correlated errors, Bayesian Inference, MCMC.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Dados Composicionais Considerando as Transformações Razão Log-Aditiva e Box-Cox</b>	<b>5</b>
<b>3</b>	<b>Uma Introdução à Inferência Bayesiana</b>	<b>9</b>
3.1	Fórmula de Bayes . . . . .	10
3.2	Densidades <i>a priori</i> e <i>a posteriori</i> . . . . .	10
3.3	Seleção de Modelos . . . . .	12
3.3.1	Critério de Informação Bayesiana . . . . .	12
3.4	Estimador Pontual . . . . .	13
<b>4</b>	<b>Uma Análise Bayesiana para os Dados do Oceano Ártico considerando a Transformação ALR</b>	<b>15</b>
4.1	Uma análise Bayesiana considerando erros correlacionados . . . . .	17
4.1.1	Função de Verossimilhança . . . . .	17
4.1.2	Informação <i>a Priori</i> . . . . .	18
4.1.3	<i>Posteriori</i> Conjunta . . . . .	18
4.1.4	Distribuições Condicionais para o algoritmo Gibbs sampling . . . . .	19
4.2	Uma análise Bayesiana considerando erros não correlacionados . . . . .	21
4.2.1	Função de Verossimilhança . . . . .	21
4.2.2	Distribuições <i>a Priori</i> . . . . .	22
4.2.3	<i>Posteriori</i> Conjunta . . . . .	23
4.2.4	Distribuições Condicionais necessárias para o amostrador de Gibbs . . . . .	23
4.3	Aplicação da Metodologia . . . . .	24

4.3.1	Aplicação da metodologia considerando erros correlacionados . . . . .	24
4.3.2	Aplicação da metodologia considerando erros não correlacionados . . . . .	28
<b>5</b>	<b>Uma Análise Bayesiana para os Dados do Oceano Ártico considerando a Transformação de Box-Cox.</b>	<b>32</b>
5.1	Uma análise Bayesiana considerando erros correlacionados . . . . .	34
5.1.1	Função de Verossimilhança . . . . .	34
5.1.2	Informação <i>a Priori</i> . . . . .	36
5.1.3	<i>Posteriori</i> Conjunta . . . . .	36
5.1.4	Distribuições Condicionais necessárias para o amostrador de Gibbs . . . . .	36
5.2	Uma análise Bayesiana considerando erros não correlacionados . . . . .	39
5.2.1	Função de Verossimilhança . . . . .	39
5.2.2	Informação <i>a Priori</i> . . . . .	41
5.2.3	<i>Posteriori</i> Conjunta . . . . .	41
5.2.4	Distribuições Condicionais para o amostrador de Gibbs . . . . .	41
5.3	Aplicação da Metodologia . . . . .	42
5.3.1	Aplicação da metodologia considerando erros correlacionados . . . . .	42
5.3.2	Aplicação da metodologia considerando erros não correlacionados . . . . .	47
5.3.3	Seleção de Modelos . . . . .	50
<b>6</b>	<b>Uma Análise Bayesiana para Dados Composicionais Assumindo a Distribuição Potência Exponencial para os Erros</b>	<b>52</b>
6.1	Alguns casos especiais da distribuição potência exponencial . . . . .	53
6.1.1	(1) Distribuição Normal . . . . .	53
6.1.2	(2) Distribuição Exponencial Dupla . . . . .	53
6.1.3	(3) Distribuição Retangular . . . . .	53
6.2	Análise Bayesiana considerando a Distribuição Potência Exponencial . . . . .	54
6.2.1	Função de Verossimilhança . . . . .	54
6.2.2	Informação <i>a Priori</i> . . . . .	54
6.2.3	<i>Posteriori</i> Conjunta . . . . .	55
6.3	Análise dos Dados do Oceano Ártico . . . . .	55
6.3.1	Modelo M5 (Distribuição Potência Exponencial) . . . . .	56

---

6.3.2	Modelo M6 (Distribuição Normal) . . . . .	57
6.3.3	Modelo M7 (Distribuição Exponencial dupla) . . . . .	57
6.3.4	Modelo M8 (Distribuição Retangular) . . . . .	58
<b>7</b>	<b>Considerações finais</b>	<b>60</b>
	<b>Referências Bibliográficas</b>	<b>60</b>
<b>A</b>	<b>Programas</b>	<b>64</b>



# Capítulo 1

## Introdução

Em muitas aplicações com dados multivariados, temos um vetor de dados  $\underline{x}$  onde os componentes apresentam as restrições  $x_1 > 0, x_2 > 0, \dots, x_G > 0$  e  $x_1 + x_2 + \dots + x_G = 1$ . Esses dados são usualmente definidos como dados composicionais (ver, por exemplo, Aitchison, 1986).

Aplicações de dados composicionais são dadas nas mais diversas áreas, com destaque a estudos geológicos, estudos ambientais, entre outros.

Para a modelagem paramétrica de dados composicionais, a suposição usual de normalidade multivariada (ver, por exemplo, Johnson e Wichern, 1998) não é verificada para o vetor aleatório  $\underline{X}$ , devido às restrições existentes.

O Simplex é um espaço amostral natural para esses dados, porém provou ser um espaço amostral complicado para ser tratado estatisticamente devido à ausência de classes paramétricas satisfatórias. A distribuição de Dirichlet pode ser usada para modelar dados no simplex, mas às vezes sua estrutura de correlação não é adequada aos dados composicionais (ver, por exemplo, Aitchison, 1982).

Um primeiro passo para introdução de classes paramétricas dentro do simplex é estabelecermos uma notação apropriada para trabalharmos com dados composicionais.

Qualquer vetor no simplex é chamado uma composição; cada elemento deste vetor é chamado de componente, e qualquer coleção de tais vetores, dados composicionais.

Uma composição com  $G$  componentes, ou seja,  $\underline{x} = (x_1, \dots, x_G)$ , onde  $G = g + 1$ , esta completamente determinada por uma subcomposição com  $g$  componentes, ou seja,  $\underline{x} = (x_1, \dots, x_g)$ , pois se  $x_1 + \dots + x_g + x_{g+1} = 1$  (restrição de dados composicionais), temos

que  $x_{g+1} = 1 - x_1 - \dots - x_g$ .

Sendo assim para uma análise estatística com  $G$  componentes, consideramos um subvetor com  $g$  componentes, pois isso diminui a dimensão do espaço amostral.

**Espaços e vetores.** Seja  $R^g$  denotando o espaço real  $g$ -dimensional e  $S^g$  denotando o simplex  $g$ -dimensional,  $x$  e  $y$  são os vetores em  $S^g$  e  $R^g$  respectivamente, onde  $S^g$  é definido por,

$$S^g = \{(x_1, \dots, x_g) : x_i > 0 (i = 1, \dots, g), x_1 + \dots + x_g < 1\}.$$

Uma classe conhecida de distribuições em  $S^g$  é a classe de Dirichlet denotada por  $D^g(\alpha)$  tendo função de densidade dada por

$$\frac{\Gamma(\alpha_1 + \dots + \alpha_{g+1})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{g+1})} x_1^{\alpha_1-1} \dots x_g^{\alpha_g-1} (1 - x_1 - \dots - x_g)^{\alpha_{g+1}-1},$$

onde  $\alpha \in R_+^{g+1}$  e  $x = (x_1, \dots, x_g) \in S^g$ .

A classe de Dirichlet é uma classe bem estruturada no simplex; inicialmente houve uma grande utilização desta classe e de possíveis generalizações que são consideradas até hoje, mas na prática ela se torna inadequada para descrição da variabilidade de dados composicionais, pois sua estrutura de correlação é sempre negativa, sendo inapropriada aos dados com correlação positiva, o que geralmente ocorre com dados composicionais.

Aitchison (1982) propôs um caminho diferente, sair do espaço amostral restrito  $S^g$  e ir para o espaço bem definido  $R^g$  selecionando transformações adequadas entre  $S^g$  e  $R^g$ .

A idéia de induzir uma classe de distribuições bem estabelecida sobre um espaço amostral complicado é uma idéia antiga. McAlister (1879), ao tentar descrever padrões de variabilidade sobre os reais positivos, considerou  $y \in R$  com uma distribuição  $N(\mu, \sigma^2)$ , para induzir uma distribuição  $\Lambda(\mu, \sigma^2)$  nos  $R^+$ , ele expressou suas idéias em termos de inversa, ou seja,  $y = \log w$ , onde  $w \in R^+$ , a essa nova classe de distribuições atribuiu-se o nome de lognormal.

Aitchison e Shen (1985) introduziram a classe de distribuições normal-logística induzindo em  $S^g$  a classe de distribuições  $N^g(\mu, \Sigma)$  dos  $R^g$ , através da transformação razão log-aditiva (ALR). O primeiro uso desta classe para descrever padrões de variabilidade dos dados, foi dada por Aitchison e Shen (1985), que discutiram várias propriedades úteis

e demonstraram a simplicidade de sua aplicação em uma variedade de problemas com dados composicionais.

Na tabela 1 temos outras transformações elementares, a logística multiplicativa, e a híbrida logística. Essas transformações podem ser substituídas por outras que levem  $R^g$  para  $S^g$ , entretanto existem poucas transformações tão simples quanto essas.

**Tabela 1:** Transformações logísticas elementares de  $R^g$  para  $S^g$

transformações	inversas
razão log-aditiva	$y_i = \log \frac{x_i}{x_{g+1}}$
logística multiplicativa	$y_i = \log \frac{x_i}{1 - \sum_{j=1}^i x_j}$
híbrida logística	$y_1 = \log \frac{x_1}{1-x_1}$ $y_i = \log \frac{x_1}{\left(1 - \sum_{j=1}^{i-1} x_j\right) \left(1 - \sum_{j=1} x_j\right)} i = 2, \dots, g$

Rayens e Srinivasan (1991a, 1991b) consideram uma generalização da transformação ALR incorporando a transformação Box-Cox (1964).

O objetivo principal deste trabalho é desenvolver métodos Bayesianos para modelar dados composicionais considerando as transformações ALR e Box-Cox. Para ilustrar a metodologia proposta consideramos um conjunto de dados reais.

Como parte do objetivo principal utilizamos Métodos de Monte Carlo baseados na simulação de Cadeias de Markov (MCMC), explorando o uso dos algoritmos Gibbs sampling e Metropolis-Hastings.

No capítulo 2 apresentamos o conjunto de dados reais e as transformações ALR e Box-Cox para esses dados.

No capítulo 3 apresentamos uma breve introdução a inferência Bayesiana incluindo algumas considerações sobre seleção de modelos e estimação pontual.

Desenvolvemos nos capítulos 4, 5 e 6, uma análise Bayesiana utilizando os algoritmos de Gibbs sampling e Metropolis-Hastings para cada um dos modelos, aplicando a teoria proposta a um conjunto de dados reais introduzido no capítulo 2.

No capítulo 4, consideramos a transformação ALR para os dados composicionais, considerando a distribuição normal para os erros, e erros correlacionados e não correla-

cionados.

No capítulo 5 transformamos os dados composicionais considerando a transformação Box-Cox, e modelamos os dados considerando a distribuição normal para os erros, e erros correlacionados e não correlacionados.

No capítulo 6 apresentamos a distribuição potência exponencial para os erros e alguns casos especiais, aplicando a dados composicionais.

## Capítulo 2

# Dados Composicionais Considerando as Transformações Razão

## Log-Aditiva e Box-Cox

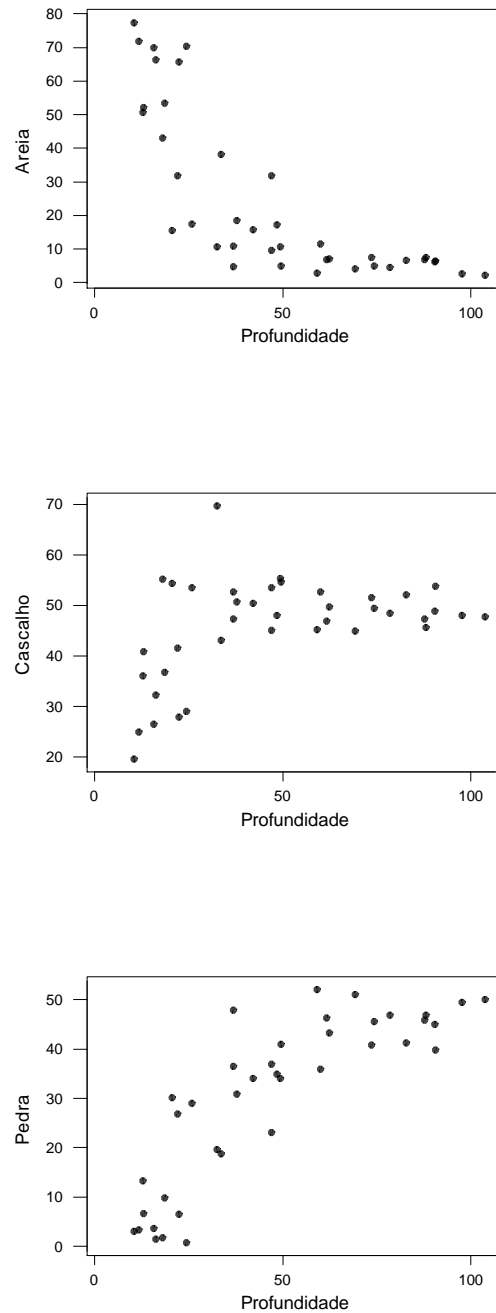
Para apresentar e ilustrar um processo de modelagem com dados composicionais, consideramos um conjunto de dados formado por três componentes na presença de uma covariável, ou seja,  $x_i = (x_{i1}, x_{i2}, x_{i3})$ ,  $i = 1, \dots, n$ . Este conjunto de dados trata da composição do solo do oceano Ártico, dada por areia, cascalho e pedra, onde  $x_{i1}$  denota a proporção de areia,  $x_{i2}$  a proporção de cascalho e  $x_{i3}$  a proporção de pedra para a  $i$ -ésima profundidade. Na tabela 2, temos a porcentagem dos componentes do solo para 39 diferentes profundidades.

Esses dados foram introduzidos originalmente por Coakley e Rust (1968) e foram discutidos por Aitchison [1986, página359].

**Tabela 2:** Composição de três componentes do solo do oceano  
Ártico para profundidades diferentes.

	Areia	Casc	Pedra	Prof		Areia	Casc	Pedra	Prof
<b>1</b>	77,5	19,5	3	10,4	<b>21</b>	9,5	53,5	37	47,1
<b>2</b>	71,9	24,9	3,2	11,7	<b>22</b>	17,1	48	34,9	48,4
<b>3</b>	50,7	36,1	13,2	12,8	<b>23</b>	10,5	55,4	34,1	49,4
<b>4</b>	52,2	40,9	6,6	13	<b>24</b>	4,8	54,7	41	49,5
<b>5</b>	70	26,5	3,5	15,7	<b>25</b>	2,6	45,2	52,2	59,2
<b>6</b>	66,5	32,2	1,3	16,3	<b>26</b>	11,4	52,7	35,9	60,1
<b>7</b>	43,1	55,3	1,6	18	<b>27</b>	6,7	46,9	46,4	61,7
<b>8</b>	53,4	36,8	9,8	18,7	<b>28</b>	6,9	49,7	43,4	62,4
<b>9</b>	15,5	54,4	30,1	20,7	<b>29</b>	4	44,9	51,1	69,3
<b>10</b>	31,7	41,5	26,8	22,1	<b>30</b>	7,4	51,6	40,9	73,6
<b>11</b>	65,7	27,8	6,5	22,4	<b>31</b>	4,8	49,5	45,7	74,4
<b>12</b>	70,4	29	0,6	24,4	<b>32</b>	4,5	48,5	47	78,5
<b>13</b>	17,4	53,6	29	25,8	<b>33</b>	6,6	52,1	41,3	82,9
<b>14</b>	10,6	69,8	19,6	32,5	<b>34</b>	6,7	47,3	45,9	87,7
<b>15</b>	38,2	43,1	18,7	33,6	<b>35</b>	7,4	45,6	46,9	88,1
<b>16</b>	10,8	52,7	36,5	36,8	<b>36</b>	6	48,9	45,1	90,4
<b>17</b>	18,4	50,7	30,9	37,8	<b>37</b>	6,3	53,8	39,9	90,6
<b>18</b>	4,6	47,4	48	36,9	<b>38</b>	2,5	48	49,5	97,7
<b>19</b>	15,6	50,4	34	42,2	<b>39</b>	2	47,8	50,2	103,7
<b>20</b>	31,9	45,1	23	47					

Considerando as proporções dos componentes em cada profundidade, gostaríamos de investigar se há alguma dependência entre estas proporções e as profundidades, ou seja, se a composição do solo é dependente da profundidade da água.



**Figura 1:** Proporções da composição versus covariável

Na figura 1, torna-se claro que há uma tendência decrescente nas proporções da areia conforme aumenta a profundidade, e quanto às proporções de pedra e cascalho, percebemos que elas seguem uma tendência crescente conforme aumenta a profundidades da água.

Na terminologia familiar estatística estamos tratando de um problema de regressão,

onde o vetor com as proporções dos componentes é a variável resposta, e a profundidade da água é a variável explicativa. O aspecto diferente do problema é a restrição de soma igual a um para os componentes do vetor da variável resposta.

Para modelar dados composicionais, podemos assumir o modelo de regressão (ver, por exemplo, Iyengar 1996a, 1996b) dado por,

$$\underline{y}_i = \underline{\alpha} + \underline{Z}_i \Theta + \varepsilon_i, \quad (2.1)$$

onde  $\underline{Z}_i$  é um vetor  $(1 \times p)$  de covariáveis associadas à  $i$ -ésima amostra;  $\underline{\alpha}$  é um vetor  $(1 \times g)$  de intercepto,  $\Theta$  é uma matriz  $(p \times g)$  de coeficientes de regressão,  $\varepsilon_i$  denota o vetor de erros,  $\underline{y}_i = (y_{i1}, \dots, y_{ig})$  é um vetor  $(g \times 1)$  onde  $g = G - 1$  e  $G$  é o número de componentes de dados composicionais, e  $y_{ik} = H(x_{ik}/x_{iG})$ ,  $i = 1, \dots, n$  e  $k = 1, \dots, g$ , sendo  $H(\cdot)$  a função de transformação escolhida para assegurar que o vetor resultante tenha componentes reais, onde  $x_{ik}$  representa a  $i$ -ésima observação para o  $k$ -ésimo componente, tal que  $x_{ik} > 0$  e  $\sum_{k=1}^G x_{ik} = 1$ ,  $i = 1, \dots, n$  e  $k = 1, \dots, G$ .

Para análise dos dados composicionais, consideramos inicialmente a transformação razão log-aditiva (ALR) dada por,

$$y_{ik} = H\left(\frac{x_{ik}}{x_{iG}}\right) = \log\left(\frac{x_{ik}}{x_{iG}}\right), \quad (2.2)$$

onde  $i = 1, \dots, n$  e  $k = 1, \dots, g$ .

Outra possibilidade é dada pela transformação Box-Cox (1964), dada por

$$y_{ik} = H\left(\frac{x_{ik}}{x_{iG}}\right) = \begin{cases} \frac{(x_{ik}/x_{iG})^{\lambda_k} - 1}{\lambda_k} & \text{se } \lambda_k \neq 0 \\ \log(x_{ik}/x_{iG}) & \text{se } \lambda_k = 0 \end{cases}, \quad (2.3)$$

onde  $\lambda_k$  é um parâmetro desconhecido  $i = 1, \dots, n$  e  $k = 1, \dots, g$ . Observar que a transformação Box-Cox inclui a transformação ALR como um caso especial.



# Capítulo 3

## Uma Introdução à Inferência Bayesiana

Métodos Bayesianos são métodos alternativos aos clássicos, que permitem a incorporação de informação *a priori* sobre os parâmetros do modelo, e ao contrário dos métodos clássicos os métodos Bayesianos consideram o parâmetro de interesse  $\theta$  uma variável aleatória, associando a ele uma distribuição de probabilidade.

A informação que se tem sobre uma quantidade de interesse é fundamental na Estatística. O verdadeiro valor de  $\theta$  é desconhecido e a idéia é reduzir este desconhecimento. Além disso, a intensidade da incerteza a respeito de  $\theta$  pode assumir diferentes graus. Do ponto de vista Bayesiano, estes diferentes graus de incerteza são representados através de modelos probabilísticos para  $\theta$ . Neste contexto, é natural que diferentes pesquisadores possam ter diferentes graus de incerteza sobre  $\theta$  (especificando modelos distintos).

A partir da fórmula de Bayes, que combina a informação vinda dos dados através da verossimilhança com a incorporação de conhecimentos *a priori*, ou seja, com as informações obtidas antes da coleta dos dados, obtém-se uma distribuição denominada distribuição *a posteriori*. A partir desta distribuição *a posteriori* é possível determinar as quantidades necessárias para o processo inferencial, tais como medidas de posição e dispersão.

### 3.1 Fórmula de Bayes

Considere uma quantidade de interesse desconhecida  $\theta$  (tipicamente não observável). A informação de que dispomos sobre  $\theta$ , resumida probabilisticamente através de  $p(\theta)$ , chamada distribuição *a priori*, pode ser aumentada observando-se uma quantidade aleatória  $X$  relacionada com  $\theta$ . A distribuição amostral  $p(x|\theta)$  define esta relação. A idéia de que após observar  $X = x$  a quantidade de informação sobre  $\theta$  aumenta é bastante intuitiva e o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação,

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)}, \quad (3.1)$$

onde,  $p(\theta|x)$  é chamada densidade *a posteriori*, e  $p(x) = \int p(x|\theta)p(\theta) d\theta$  recebe o nome de densidade preditiva.

Para um valor fixo de  $x$ , temos que  $L(\theta; x) = p(x|\theta)$  é a verossimilhança de cada um dos possíveis valores de  $\theta$ . Como  $p(\theta|x)$  é uma densidade para  $\theta$ , a observação  $x$  é apenas uma constante bem como  $p(x)$ , temos que a forma usual do teorema de Bayes é dada por,

$$p(\theta|D) \propto L(\theta)p(\theta), \quad (3.2)$$

onde  $D$  denota o conjunto de observações de  $x$ .

Mudando a notação para uma linguagem Bayesiana, denotamos as distribuições *a priori* e *a posteriori* dadas respectivamente por  $\pi(\theta)$ ,  $\pi(\theta|D)$ , desta forma a expressão dada em (3.2) é reescrita como,

$$\pi(\theta|D) \propto L(\theta)\pi(\theta). \quad (3.3)$$

### 3.2 Densidades *a priori* e *a posteriori*

Na abordagem Bayesiana, os parâmetros são variáveis aleatórias onde podemos incorporar no modelo uma informação adicional sobre a variabilidade dos erros; essa é uma informação *a priori* do pesquisador sobre os parâmetros, que pode ser expressa utilizando

distribuições *a priori* informativas. Caso não haja opinião sólida sobre os parâmetros em estudo, então expressamos o não conhecimento *a priori* através de distribuições *a priori* não informativas.

Definido o modelo para os dados e a distribuição *a priori*, combinamos a informação contida nos dados com a informação prévia sobre o parâmetro  $\theta$  e obtemos assim a distribuição *a posteriori* para  $\theta$ .

A partir da distribuição *a posteriori* de  $\theta$ , podemos determinar as quantidades de interesse necessárias para o processo inferencial, tais como média, moda, variância, etc.

Uma das limitações dos métodos Bayesianos nas aplicações em geral eram relacionadas à complexidade das densidades conjuntas *a posteriori*. Uma alternativa muito usada recentemente são dadas por técnicas de simulação. Para simular as quantidades de interesse, utilizamos os Métodos de Monte Carlo baseados na simulação de Cadeias de Markov (MCMC).

MCMC é uma metodologia baseada na substituição da expressão analítica da densidade por uma amostra gerada a partir dessa densidade. A inferência é realizada basicamente através desse procedimento. A grande vantagem desse método é a possibilidade de gerar uma amostra de uma densidade sem a necessidade de conhecê-la. Em inferência Bayesiana o procedimento MCMC é extremamente importante devido à complexidade das distribuições *a posteriori*.

Quando as densidades condicionais da função da qual queremos gerar uma amostra são conhecidas, podemos usar o algoritmo de Gibbs sampling que é um método de simulação via cadeias de Markov que oferece uma alternativa para se obter uma amostra da distribuição de interesse baseada em sucessivas gerações das distribuições condicionais.

Em modelos mais complexos, geralmente não temos formas conhecidas nas distribuições condicionais; nesse caso usamos o algoritmo Metropolis-Hastings (ver por exemplo Gelfand & Smith, 1990). Com esses métodos encontramos estimativas pontuais e por intervalo dos parâmetros de interesse do modelo.

A convergência das amostras de interesse pode ser verificada, usando o critério proposto por Gelman & Rubin (1992).

### 3.3 Seleção de Modelos

Quando mais de um modelo é considerado para analisar um conjunto de dados, surge uma questão: Qual ou quais modelos são mais apropriados para a análise dos dados em questão?

O processo de seleção de modelos se torna um problema importante em estatística. Sob um enfoque Bayesiano vários métodos de seleção de modelos são propostos na literatura, entre os quais destacamos o fator de Bayes. A seguir apresentamos alguns critérios mais usados na discriminação de modelos.

#### 3.3.1 Critério de Informação Bayesiana

A comparação de modelos sob a teoria Bayesiana pode ser feita a partir de medidas de adequabilidade, como o Deviance Information Criterion (DIC) (ver Spiegelhalter et al, 2002) ou Bayesian Information Criterion (BIC) (Schwarz, 1978), os quais são aproximações do fator de Bayes.

##### Critério BIC

Carlin e Louis (2000) introduziram uma modificação do critério BIC dado por

$$BIC_i = -2E(\ln L(\theta_i|x, M_i) + p_i \ln(n)), \quad (3.4)$$

onde  $i$  indexa o modelo;  $n$  é o número da amostra e  $p_i$  é o número de parâmetros sob o modelo  $M_i$ . Maiores valores do BIC indicam o melhor ajuste do modelo (ver por exemplo, Paulino et al, 2003).

##### Critério DIC

Spiegelhalter et al (2002) generalizam o critério AIC (Akaike, 1973) baseado na distribuição *a posteriori* dada por

$$D_i(\theta) = -2 \log \left( \frac{f\left(\underline{x}|\theta, M_i\right)}{h\left(\underline{x}\right)} \right), \quad (3.5)$$

onde  $h\left(\begin{smallmatrix} x \\ - \end{smallmatrix}\right)$  é uma função só dos dados e não tem impacto na escolha do modelo.

A função genérica que calcula esta informação é dada por

$$DIC_i = \bar{D}_i + p_{D_i}, \quad (3.6)$$

onde  $\bar{D}_i = E_{\theta_i|x, M_i} \left[ D_i | \theta_i \right]$  a média *a posteriori* da função deviance e  $p_{D_i}$  é o número efetivo de parâmetros do modelo  $i$  em estudo, dado por

$$p_{D_i} = E(\theta_i|x, M_i) [D_i(\theta_i)] - D_i [E(\theta_i|x, M_i) [D_i(\theta_i)]] . \quad (3.7)$$

Menores valores do DIC indicam o melhor ajuste do modelo.

### 3.4 Estimador Pontual

Um estimador Bayesiano de  $\theta$  é dado por  $\hat{\theta} = d(x)$ , que minimiza o erro esperado (função de risco) com respeito à distribuição *a posteriori* para  $\theta$ .

Considerando a função de perda quadrática dada por

$$L(\hat{\theta}, \theta) = \left( \hat{\theta} - \theta \right)^2, \quad (3.8)$$

devemos achar  $d(x)$  que minimiza o risco Bayesiano (valor esperado da perda em relação à distribuição *a posteriori*), dado por

$$R(d, \theta) = E_{\theta|x} \{L(d, \theta)\} = \int_{\theta} (d(x) - \theta)^2 \pi(\theta|x) d\theta. \quad (3.9)$$

Sendo  $d(x)$  uma função diferenciável, o estimador de Bayes com respeito à função perda quadrática é dado por

$$\frac{dR(d, \theta)}{d(d)} = 2 \int_{\theta} (d(x) - \theta) \pi(\theta|x) d\theta = 0, \quad (3.10)$$

isto é,  $\hat{\theta} = d(x) = E(\theta|x)$ .

Outras funções de perda também podem ser utilizadas na estimação pontual (ver por

exemplo Paulino et al, 2003).

## Capítulo 4

# Uma Análise Bayesiana para os Dados do Oceano Ártico considerando a Transformação ALR

Considerando os dados composicionais da tabela 2,  $x_{i1}$ ,  $x_{i2}$ ,  $x_{i3}$ ,  $i = 1, \dots, n$ , e assumindo a transformação ALR temos,

$$y_{i1} = \log\left(\frac{x_{i1}}{x_{i3}}\right) \text{ e } y_{i2} = \log\left(\frac{x_{i2}}{x_{i3}}\right).$$

Considere o modelo (2.1) para os dados transformados  $y_{i1}$  e  $y_{i2}$ , ou seja,

$$\begin{aligned} y_{i1} &= \alpha_1 + \theta_1 Z_i + \varepsilon_{i1}, \\ y_{i2} &= \alpha_2 + \theta_2 Z_i + \varepsilon_{i2}, \end{aligned} \tag{4.1}$$

onde  $i = 1, \dots, n$ ,  $Z_i$  é a covariável profundidade,  $y_i = (y_{i1}, y_{i2})$  é o vetor de variáveis resposta e  $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2})$  é o vetor de erros. Assumiremos em uma primeira análise que  $\varepsilon_i$  tem uma distribuição normal bivariada  $N\{\underline{0}, \Sigma\}$  onde  $\underline{0}$  é o vetor de zeros e  $\Sigma$  é a matriz de variância-covariância dada por,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{4.2}$$

onde  $\rho$  é o parâmetro de correlação entre  $\varepsilon_{i1}$  e  $\varepsilon_{i2}$ ; denotamos este modelo por Modelo M1.

Observar que considerando as proporções verdadeiras dos componentes,  $\beta_{i1}$ ,  $\beta_{i2}$  e  $\beta_{i3}$ , onde  $\beta_{i1} + \beta_{i2} + \beta_{i3} = 1$  e  $\beta_{i1} > 0$ ,  $\beta_{i2} > 0$  e  $\beta_{i3} > 0$ , e a transformação ALR, temos de (4.1),  $\log(\beta_{i1}/\beta_{i3}) = \alpha_1 + \theta_1 Z_i$  e  $\log(\beta_{i2}/\beta_{i3}) = \alpha_2 + \theta_2 Z_i$ , onde aplicando a função exponencial obtemos,

$$\begin{aligned} e^{\alpha_1 + \theta_1 Z_i} &= \frac{\beta_{i1}}{\beta_{i3}}, \\ e^{\alpha_2 + \theta_2 Z_i} &= \frac{\beta_{i2}}{\beta_{i3}}, \end{aligned}$$

ou seja,

$$\begin{aligned} \beta_{i1} &= \beta_{i3} [e^{\alpha_1 + \theta_1 Z_i}], \\ \beta_{i2} &= \beta_{i3} [e^{\alpha_2 + \theta_2 Z_i}]. \end{aligned} \tag{4.3}$$

Como  $\beta_{i1} + \beta_{i2} + \beta_{i3} = 1$ , temos  $\beta_{i3} = 1 - (\beta_{i1} + \beta_{i2})$ , e de (4.3) temos,  $\beta_{i1} + \beta_{i2} = \beta_{i3} [e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}]$  e  $\beta_{i1}/\beta_{i2} = e^{\alpha_1 + \theta_1 Z_i} / e^{\alpha_2 + \theta_2 Z_i}$ .

Assim  $\beta_{i3} = 1 - (\beta_{i1} + \beta_{i2}) = 1 - \beta_{i3} [e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}]$ , isto é,  $\beta_{i3} [1 + e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}] = 1$ , ou,  $\beta_{i3} = 1 / [1 + e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}]$  e  $\beta_{i1} + \beta_{i2} = e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i} / [1 + e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}]$ .

Portanto, obtemos as seguintes relações:

$$\begin{aligned} \beta_{i1} &= \frac{e^{\alpha_1 + \theta_1 Z_i}}{1 + e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}}, \\ \beta_{i2} &= \frac{e^{\alpha_2 + \theta_2 Z_i}}{1 + e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}} \end{aligned} \tag{4.4}$$

e



$$\beta_{i3} = \frac{1}{1 + e^{\alpha_1 + \theta_1 Z_i} + e^{\alpha_2 + \theta_2 Z_i}},$$

onde  $i = 1, \dots, n$ .

## 4.1 Uma análise Bayesiana considerando erros correlacionados

### 4.1.1 Função de Verossimilhança

Considerando  $\underline{y}_i = (y_{i1}, y_{i2})$ , dados observados transformados pela função ALR, o modelo dado em (2.1) e assumindo que o vetor de erros  $\underline{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2})$  tem uma distribuição normal bivariada com matriz de variância-covariância dada em (4.2), a função de verossimilhança para  $\underline{v}_1 = (\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho)$  é dada por,

$$L(\underline{v}_1) = \prod_{i=1}^n f(\underline{\varepsilon}_i | \underline{v}_1), \quad (4.5)$$

onde  $f(\underline{\varepsilon}_i | \underline{v}_1) \sim N_2\{0, \Sigma\}$ , ou seja,

$$f(\underline{\varepsilon}_i | \underline{v}_1) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{\varepsilon_{i1}^2}{\sigma_1^2} - \frac{2\rho\varepsilon_{i1}\varepsilon_{i2}}{\sigma_1\sigma_2} + \frac{\varepsilon_{i2}^2}{\sigma_2^2}\right]\right\}.$$

Assim (4.5) pode ser reescrita como,

$$L(\underline{v}_1) \propto (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} (1-\rho^2)^{-n/2} \times \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n \varepsilon_{i1}\varepsilon_{i2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2\right]\right\}, \quad (4.6)$$

onde,

$$\varepsilon_{i1} = y_{i1} - \alpha_1 - \theta_1 Z_i;$$

$$\begin{aligned}
 \varepsilon_{i2} &= y_{i2} - \alpha_2 - \theta_2 Z_i; \\
 \sum_{i=1}^n \varepsilon_{i1}^2 &= \sum_{i=1}^n (y_{i1} - \alpha_1 - \theta_1 Z_i)^2 = \sum_{i=1}^n y_{i1}^2 - 2\alpha_1 \sum_{i=1}^n y_{i1} - 2\theta_1 \sum_{i=1}^n Z_i y_{i1} + n\alpha_1^2 + 2\theta_1 \alpha_1 \sum_{i=1}^n Z_i + \\
 &\theta_1^2 \sum_{i=1}^n Z_i^2; \\
 \sum_{i=1}^n \varepsilon_{i2}^2 &= \sum_{i=1}^n (y_{i2} - \alpha_2 - \theta_2 Z_i)^2 = \sum_{i=1}^n y_{i2}^2 - 2\alpha_2 \sum_{i=1}^n y_{i2} - 2\theta_2 \sum_{i=1}^n Z_i y_{i2} + n\alpha_2^2 + 2\theta_2 \alpha_2 \sum_{i=1}^n Z_i + \\
 &\theta_2^2 \sum_{i=1}^n Z_i^2; \\
 \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} &= \sum_{i=1}^n (y_{i1} - \alpha_1 - \theta_1 Z_i)(y_{i2} - \alpha_2 - \theta_2 Z_i) = \sum_{i=1}^n y_{i1} y_{i2} - \alpha_2 \sum_{i=1}^n y_{i1} - \theta_2 \sum_{i=1}^n Z_i y_{i1} - \\
 &\alpha_1 \sum_{i=1}^n y_{i2} + n\alpha_1 \alpha_2 + \theta_2 \alpha_1 \sum_{i=1}^n Z_i + \\
 &\theta_1 \sum_{i=1}^n Z_i y_{i2} + \theta_1 \alpha_2 \sum_{i=1}^n Z_i + \theta_1 \theta_2 \sum_{i=1}^n Z_i^2.
 \end{aligned}$$

### 4.1.2 Informação *a Priori*

Para uma análise Bayesiana dos dados composicionais da tabela 2, considerando a verossimilhança (4.6), assumimos as seguintes distribuições *a priori* para os parâmetros:

$$\begin{aligned}
 \text{i) } \theta_k &\sim N\{a_k, b_k^2\}; \quad k = 1, 2; \quad a_k, b_k^2 \text{ conhecidos,} \\
 \text{ii) } \alpha_k &\sim N\{c_k, d_k^2\}; \quad k = 1, 2; \quad c_k, d_k^2 \text{ conhecidos,} \\
 \text{iii) } \sigma_k^2 &\sim IG[e_k, f_k]; \quad k = 1, 2; \quad e_k, f_k \text{ conhecidos,} \\
 \text{iv) } \rho &\sim U[-1, 1],
 \end{aligned} \tag{4.7}$$

onde  $N\{\mu, \sigma^2\}$  denota uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ ,  $IG[a, b]$  denota uma distribuição gama inversa com média  $b/(a-1)$  e variância  $b^2/[(a-1)^2(a-2)]$ ,  $a > 2$  e  $U[c, d]$ , denota uma distribuição uniforme no intervalo  $[c, d]$ .

### 4.1.3 *Posteriori Conjunta*

Considerando a função de verossimilhança (4.6) e as distribuições *a priori* (4.7) assumindo independência *a priori* entre os parâmetros, obtemos a partir do teorema de Bayes a distribuição *a posteriori* conjunta para  $v_1 = (\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho)$  dada por,

$$\begin{aligned} \pi \left( \underline{v}_1 | \text{dados} \right) &\propto \left\{ \prod_{k=1}^2 \exp \left[ -\frac{1}{2b_k^2} (\theta_k - a_k)^2 \right] \right\} \times \\ &\times \left\{ \prod_{k=1}^2 \exp \left[ -\frac{1}{2d_k^2} (\alpha_k - d_k)^2 \right] \right\} \times \\ &\times \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(e_k+1)} e^{-f_k/\sigma_k^2} \right\} (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} (1 - \rho^2)^{-n/2} \times \\ &\times \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n \varepsilon_{i1}\varepsilon_{i2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right] \right\}. \end{aligned} \quad (4.8)$$

#### 4.1.4 Distribuições Condicionais para o algoritmo Gibbs sampling

As distribuições condicionais necessárias para o algoritmo Gibbs sampling são dadas por,

$$\text{i) } \theta_1 | v_1(\theta_1), D \sim N \left\{ \frac{a_1\sigma_1^2\sigma_2(1 - \rho^2) + \sigma_2 b_1^2 \sum_{i=1}^n Z_i u_i^{(1)} - \sigma_1 b_1^2 \rho \sum_{i=1}^n Z_i B_i^{(1)}}{\sigma_2 \left[ \sigma_1^2(1 - \rho^2) + b_1^2 \sum_{i=1}^n Z_i^2 \right]}; \frac{b_1^2 \sigma_1^2(1 - \rho^2)}{\sigma_1^2(1 - \rho^2) + b_1^2 \sum_{i=1}^n Z_i^2} \right\},$$

onde  $u_i^{(1)} = y_{i1} - \alpha_1$ ,  $B_i^{(1)} = y_{i2} - \alpha_2 - \theta_2 Z_i$  e  $v_1(\theta_1)$  é o vetor de todos os parâmetros exceto  $\theta_1$ .

$$\text{ii) } \theta_2 | v_1(\theta_2), D \sim N \left\{ \frac{a_2\sigma_2^2\sigma_1(1 - \rho^2) + \sigma_1 b_2^2 \sum_{i=1}^n Z_i u_i^{(2)} - \sigma_2 b_2^2 \rho \sum_{i=1}^n Z_i B_i^{(2)}}{\sigma_1 \left[ \sigma_2^2(1 - \rho^2) + b_2^2 \sum_{i=1}^n Z_i^2 \right]}; \frac{b_2^2 \sigma_2^2(1 - \rho^2)}{\sigma_2^2(1 - \rho^2) + b_2^2 \sum_{i=1}^n Z_i^2} \right\},$$

onde  $u_i^{(2)} = y_{i2} - \alpha_2$ ,  $B_i^{(2)} = y_{i1} - \alpha_1 - \theta_1 Z_i$ ;  $i = 1, \dots, n$  e  $v_1(\theta_2)$  é o vetor de todos os parâmetros exceto  $\theta_2$ .

$$\text{iii) } \alpha_1 | v_1(\alpha_1), D \sim N \left\{ \frac{c_1\sigma_1^2\sigma_2(1 - \rho^2) + \sigma_2 d_1^2 \sum_{i=1}^n \xi_i^{(1)} - \sigma_1 d_1^2 \rho \sum_{i=1}^n C_i^{(1)}}{\sigma_2 \left[ \sigma_1^2(1 - \rho^2) + n d_1^2 \right]}; \frac{d_1^2 \sigma_1^2(1 - \rho^2)}{\sigma_1^2(1 - \rho^2) + n d_1^2} \right\}, \quad (4.9)$$

onde  $\xi_i^{(1)} = y_{i1} - \theta_1 Z_i$ ,  $C_i^{(1)} = y_{i2} - \alpha_2 - \theta_2 Z_i$ ;  $i = 1, \dots, n$  e  $v_{1(\alpha_1)}$  é o vetor de todos os parâmetros exceto  $\alpha_1$ .

$$\text{iv) } \alpha_2 | v_{1(\alpha_2)}, D \sim N \left\{ \frac{c_2 \sigma_2^2 \sigma_1 (1 - \rho^2) + \sigma_1 d_2^2 \sum_{i=1}^n \xi_i^{(2)} - \sigma_2 d_2^2 \rho \sum_{i=1}^n C_i^{(2)}}{\sigma_2 [\sigma_1^2 (1 - \rho^2) + n d_2^2]}; \frac{d_2^2 \sigma_2^2 (1 - \rho^2)}{\sigma_2^2 (1 - \rho^2) + n d_2^2} \right\},$$

onde  $\xi_i^{(2)} = y_{i2} - \theta_2 Z_i$ ,  $C_i^{(2)} = y_{i1} - \alpha_1 - \theta_1 Z_i$ ;  $i = 1, \dots, n$  e  $v_{1(\alpha_2)}$  é o vetor de todos os parâmetros exceto  $\alpha_2$ .

$$\text{v) } \pi \left( \sigma_1^2 | v_{1(\sigma_1^2)}, D \right) \propto (\sigma_1^2)^{-(e_1+1)} e^{-f_1/\sigma_1^2} \Psi_1 \left( v_1 \right),$$

$$\text{onde } \Psi_1 \left( v_1 \right) = \exp \left\{ -\frac{n}{2} \ln(\sigma_1^2) - \frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1 \sigma_2} \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} \right] \right\},$$

$\varepsilon_{i1} = y_{i1} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2} - \alpha_2 - \theta_2 Z_i$ ,  $i = 1, \dots, n$  e  $v_{1(\sigma_1^2)}$  é o vetor de todos os parâmetros exceto  $\sigma_1^2$ .

$$\text{vi) } \pi \left( \sigma_2^2 | v_{1(\sigma_2^2)}, D \right) \propto (\sigma_2^2)^{-(e_2+1)} e^{-f_2/\sigma_2^2} \Psi_2 \left( v_1 \right),$$

$$\text{onde } \Psi_2 \left( v_1 \right) = \exp \left\{ -\frac{n}{2} \ln(\sigma_2^2) - \frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 - \frac{2\rho}{\sigma_1 \sigma_2} \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} \right] \right\},$$

$\varepsilon_{i1} = y_{i1} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2} - \alpha_2 - \theta_2 Z_i$ ,  $i = 1, \dots, n$  e  $v_{1(\sigma_2^2)}$  é o vetor de todos os parâmetros exceto  $\sigma_2^2$ .

$$\text{vii) } \pi \left( \rho | v_{1(\rho)}, D \right) \propto (1 - \rho^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1 \sigma_2} \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right] \right\},$$

onde  $\varepsilon_{i1} = y_{i1} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2} - \alpha_2 - \theta_2 Z_i$ ,  $i = 1, \dots, n$  e  $v_{1(\rho)}$  é o vetor de todos os parâmetros exceto  $\rho$ .

A partir das distribuições condicionais (4.9), podemos utilizar o algoritmo computa-

cional Gibbs sampling ou simplesmente amostrador de Gibbs, para gerar amostras para  $\alpha_1, \alpha_2, \theta_1, \theta_2$ , pois as distribuições condicionais apresentam forma conhecida.

Observar que para gerar  $\sigma_1^2, \sigma_2^2$  e  $\rho$ , precisamos utilizar o algoritmo de Metropolis-Hastings (ver por exemplo, Roberts e Smith, 1993), pois as distribuições condicionais apresentam forma desconhecida.

## 4.2 Uma análise Bayesiana considerando erros não correlacionados

Considerando os dados observados transformados pela função ALR e o modelo dado em (4.1), agora assumimos que as variáveis aleatórias  $\varepsilon_{i1}$  e  $\varepsilon_{i2}$  são independentes com distribuições normais com médias iguais a zero e variâncias  $\sigma_1^2$  e  $\sigma_2^2$ , respectivamente; neste caso observar que temos um caso particular do Modelo M1, com  $\rho = 0$  (erros não correlacionados), que denotamos por Modelo M2. Assim,

$$\varepsilon_i \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right\}. \quad (4.10)$$

De (4.1) temos,

$$\begin{aligned} \varepsilon_{i1} &= y_{i1} - \alpha_1 - \theta_1 Z_i \\ \varepsilon_{i2} &= y_{i2} - \alpha_2 - \theta_2 Z_i. \end{aligned} \quad (4.11)$$

### 4.2.1 Função de Verossimilhança

Considerando as suposições de normalidade e independência observamos que  $y_{i1} \sim N\{\alpha_1 + \theta_1 Z_i; \sigma_1^2\}$  e  $y_{i2} \sim N\{\alpha_2 + \theta_2 Z_i; \sigma_2^2\}$ , logo a função densidade de probabilidade conjunta de  $y_{i1}$  e  $y_{i2}$  é dada por

$$\begin{aligned} f(y_{i1}, y_{i2}) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left\{ -\frac{1}{2\sigma_1^2} (y_{i1} - \alpha_1 - \theta_1 Z_i)^2 \right\} \times \\ &\times \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \left\{ -\frac{1}{2\sigma_2^2} (y_{i2} - \alpha_2 - \theta_2 Z_i)^2 \right\}. \end{aligned} \quad (4.12)$$

A função de verossimilhança para os parâmetros  $\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2$  e  $\sigma_2^2$  do modelo ( 4.1) assumindo erros não correlacionados é dada por:

$$L\left(v_2\right)=\left(\sigma_1^2\right)^{-n / 2}\left(\sigma_2^2\right)^{-n / 2} \exp \left\{-\frac{1}{2 \sigma_1^2} \sum_{i=1}^n \varepsilon_{i 1}^2-\frac{1}{2 \sigma_2^2} \sum_{i=1}^n \varepsilon_{i 2}^2\right\}, \quad (4.13)$$

onde  $v_2=\left(\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2\right)$ ,

$$\begin{aligned} \sum_{i=1}^n \varepsilon_{i 1}^2 & =\sum_{i=1}^n\left(y_{i 1}-\alpha_1-\theta_1 Z_i\right)^2=\sum_{i=1}^n y_{i 1}^2-2 \alpha_1 \sum_{i=1}^n y_{i 1}-2 \theta_1 \sum_{i=1}^n Z_i y_{i 1}+n \alpha_1^2+2 \theta_1 \alpha_1 \sum_{i=1}^n Z_i+ \\ & \theta_1^2 \sum_{i=1}^n Z_i^2 ; e \\ \sum_{i=1}^n \varepsilon_{i 2}^2 & =\sum_{i=1}^n\left(y_{i 2}-\alpha_2-\theta_2 Z_i\right)^2=\sum_{i=1}^n y_{i 2}^2-2 \alpha_2 \sum_{i=1}^n y_{i 2}-2 \theta_2 \sum_{i=1}^n Z_i y_{i 2}+n \alpha_2^2+2 \theta_2 \alpha_2 \sum_{i=1}^n Z_i+ \\ & \theta_2^2 \sum_{i=1}^n Z_i^2 . \end{aligned}$$

#### 4.2.2 Distribuições *a Priori*

Para estimar os parâmetros de interesse assumimos as distribuições *a priori* para  $\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2$  dadas por

$$\begin{aligned} \text{i) } \theta_k & \sim N\left\{a_k, b_k^2\right\} ; k=1,2 ; a_k, b_k^2 \text{ conhecidos,} \\ \text{ii) } \alpha_k & \sim N\left\{c_k, d_k^2\right\} ; k=1,2 ; c_k, d_k^2 \text{ conhecidos,} \\ \text{iii) } \sigma_k^2 & \sim IG\left[e_k, f_k\right] ; k=1,2 ; e_k, f_k \text{ conhecidos.} \end{aligned} \quad (4.14)$$

Também assumimos independência *a priori* entre os parâmetros.

### 4.2.3 Posteriori Conjunta

Considerando a verossimilhança (4.13) e as distribuições *a priori* (4.14), a *posteriori* conjunta para  $v_2 = (\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$  é dada por,

$$\begin{aligned} \pi \left( v_2 | \text{dados} \right) &\propto \left\{ \prod_{k=1}^2 \exp \left[ -\frac{1}{2b_k^2} (\theta_k - a_k)^2 \right] \right\} \times \\ &\times \left\{ \prod_{k=1}^2 \exp \left[ -\frac{1}{2d_k^2} (\alpha_k - c_k)^2 \right] \right\} \times \\ &\times \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(e_k+1)} e^{-f_k/\sigma_k^2} \right\} (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} \times \\ &\times \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right\}. \end{aligned} \quad (4.15)$$

### 4.2.4 Distribuições Condicionais necessárias para o amostrador de Gibbs

As distribuições *a posteriori* condicionais para o amostrador de Gibbs são dadas por,

$$\text{i) } \theta_k | v_2(\theta_k), D \sim N \left\{ \frac{a_k \sigma_k^2 + b_k^2 \sum_{i=1}^n Z_i u_i^{(k)}}{\sigma_k^2 + b_k^2 \sum_{i=1}^n Z_i^2}; \frac{b_k^2 \sigma_k^2}{\sigma_k^2 + b_k^2 \sum_{i=1}^n Z_i^2} \right\},$$

onde  $k = 1, 2$  e  $u_i^{(k)} = y_{ik} - \alpha_k$ ;  $i = 1, \dots, n$ .

$$\text{ii) } \alpha_k | v_2(\alpha_k), D \sim N \left\{ \frac{c_k \sigma_k^2 + d_k^2 \sum_{i=1}^n \xi_i^{(k)}}{\sigma_k^2 + n d_k^2}; \frac{d_k^2 \sigma_k^2}{\sigma_k^2 + n d_k^2} \right\}, \quad (4.16)$$

onde  $k = 1, 2$  e  $\xi_i^{(k)} = y_{ik} - \theta_k Z_i$ ;  $i = 1, \dots, n$ .

$$\text{iii) } \sigma_k^2 | v_2(\sigma_k^2), D \sim IG \left[ e_k + \frac{n}{2}; f_k + \frac{1}{2} \sum_{i=1}^n \varepsilon_{ik}^2 \right],$$

onde  $k = 1, 2$  e  $\varepsilon_{ik} = y_{ik} - \alpha_k - \theta_k Z_i$ ;  $i = 1, \dots, n$ .

Observar que para simular as quantidades de interesse, utilizamos o algoritmo de Gibbs

sampling (Geman & Geman, 1984; Gelfand & Smith, 1990), pois todas as distribuições condicionais apresentam formas conhecidas.

Uma grande simplificação para gerar amostras da distribuição *a posteriori* conjunta é dada pelo *software* Winbugs (Spiegelhalter et al, 1999) que não necessita a especificação das distribuições condicionais.

## 4.3 Aplicação da Metodologia

### 4.3.1 Aplicação da metodologia considerando erros correlacionados

A metodologia é ilustrada considerando os dados da tabela 2. Para análise dos dados, inicialmente consideramos o modelo M1, ou seja, transformação ALR com erros normais correlacionados para o modelo (4.1).

Assumindo as distribuições *a priori* (4.7) com  $a_1 = a_2 = c_1 = c_2 = 0$ ,  $b_1^2 = b_2^2 = d_1^2 = d_2^2 = 100000$ ,  $e_1 = e_2 = f_1 = f_2 = 1000$  (*prioris* não informativas para  $\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2$ ), geramos duas cadeias cada uma com 50000 iterações partindo de pontos iniciais diferentes e descartamos as primeiras 5000 iterações (“burn-in”). Além disso, consideramos um salto de tamanho 20 para eliminar a correlação entre as amostras. Na geração das amostras, utilizamos o *software* Winbugs (ver Spiegelhalter, 1999). A convergência do algoritmo foi verificada usando o método de Gelman e Rubin (1992). Na Tabela 3 temos os sumários *a posteriori* e os valores do fator de redução de escala  $\hat{R}$  (critério Gelman e Rubin). Para todos os parâmetros observamos a convergência do algoritmo Gibbs *sampling* pois  $\sqrt{\hat{R}} < 1.1$  para todos os casos.



**Tabela 3:** Sumários *a posteriori* do modelo M1

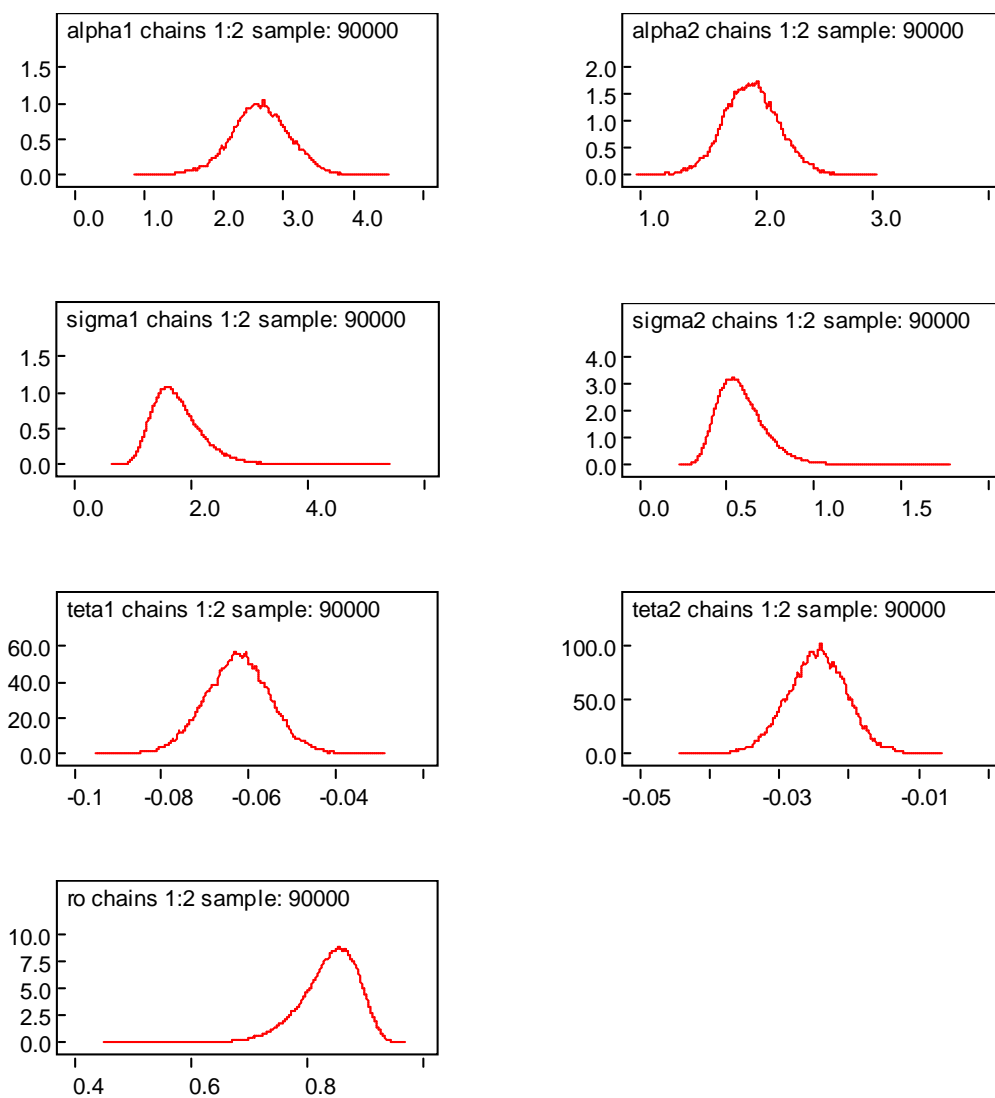
Parâmetros	Média	Int.Cred.95%	$\hat{R}$
$\alpha_1$	2.6690	(1.8350; 3.4830)	1.001
$\alpha_2$	1.9600	(1.4950; 2.4420)	1.003
$\theta_1$	-0.0624	(-0.0772; -0.0474)	1.001
$\theta_2$	-0.0245	(-0.0330; -0.0162)	1.002
$\sigma_1^2$	1.7730	(1.1270; 2.7800)	0.999
$\sigma_2^2$	0.5849	(0.3710; 0.9175)	1.000
$\rho$	0.8380	(0.7220; 0.9160)	1.010

A partir dos resultados da tabela 3 observamos que a profundidade apresenta um efeito significativo para as proporções de componentes na mistura (zero não está incluído nos intervalos de credibilidade para  $\theta_1$  e  $\theta_2$ ).

A Figura 3 mostra os traços *a posteriori* de  $\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2$  e  $\rho$ . Aparentemente os parâmetros apresentam convergência, pois há uniformidade nos traços *a posteriori*.

Nas Figuras 2, e 4 também temos as representações gráficas dos sumários *a posteriori*, gráficos das distribuições *a posteriori* condicionais, gráficos de convergência das amostras geradas e o teste gráfico de Gelman e Rubin.

Através do teste gráfico do critério de Gelman e Rubin, Figura 4, verificamos a convergência das cadeias, ou seja, os valores estão próximos de um, como vimos na tabela 3.



**Figura 2:** Gráficos das densidades *a posteriori*

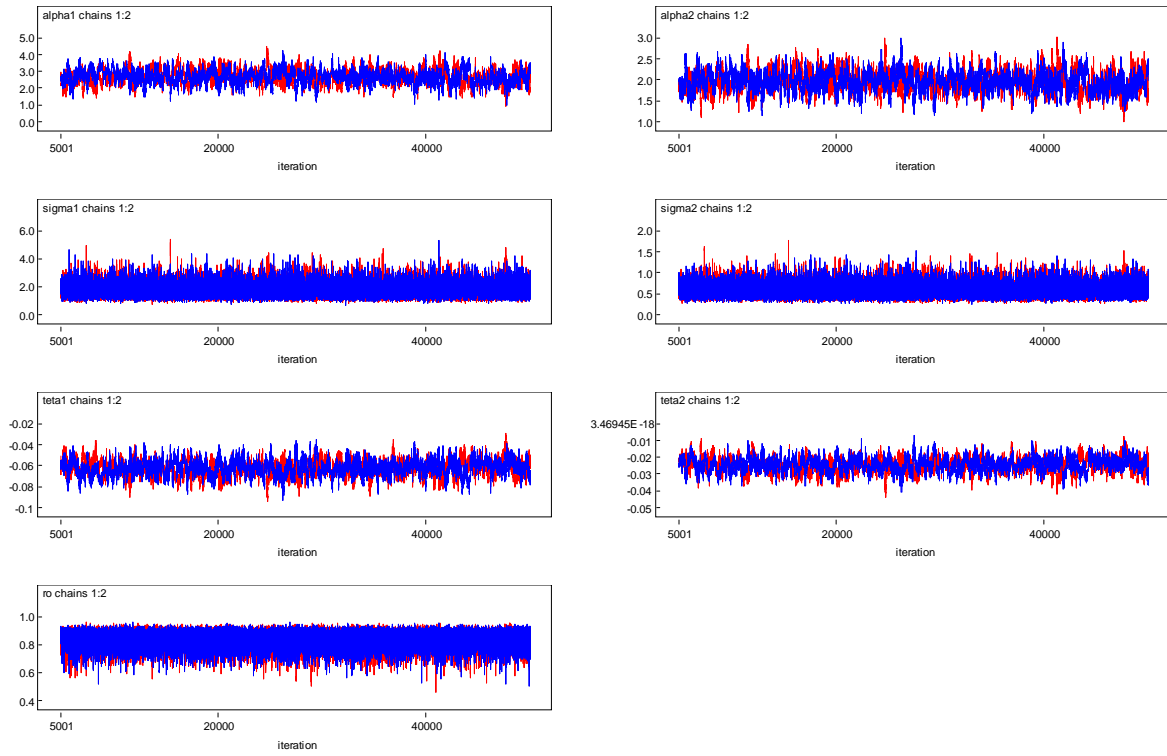


Figura 3: Traços *a posteriori*

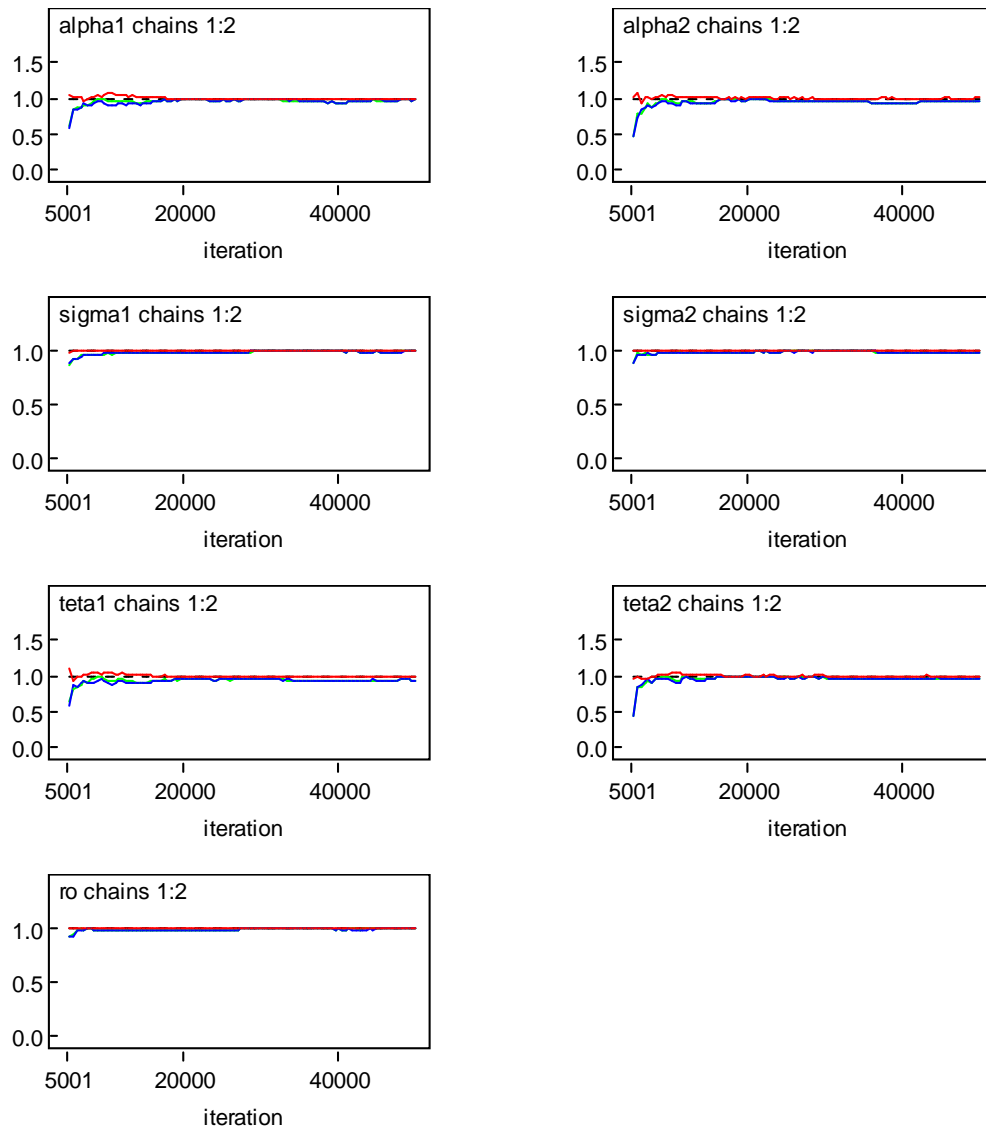


Figura 4: Teste gráfico de Gelman e Rubin

### 4.3.2 Aplicação da metodologia considerando erros não correlacionados

Assumindo  $\rho = 0$ , ou seja modelo M2, e os mesmos valores dos hiperparâmetros para as distribuições *a priori* consideradas anteriormente para o caso  $\rho \neq 0$ , temos na tabela 4 os sumários *a posteriori* para os parâmetros do modelo, onde também foram geradas duas cadeias paralelas de tamanho 50000 e foram descartadas as 5000 primeiras amostras.

Tabela 4: Sumários *a posteriori* modelo M2

Parâmetros	Média	Int.Cred 95%	$\hat{R}$
$\alpha_1$	2.6750	(1.8160; 3.5130)	1.001
$\alpha_2$	1.9620	(1.4750; 2.4470)	0.998
$\theta_1$	-0.0625	(-0.0775; -0.0471)	0.993
$\theta_2$	-0.0246	(-0.0335; -0.0158)	0.998
$\sigma_1^2$	1.8280	(1.1470; 2.8970)	1.003
$\sigma_2^2$	0.6031	(0.3790; 0.9520)	0.997

A partir dos resultados da tabela 4, também observamos o efeito significativo da profundidade dado pelos parâmetros  $\theta_1$  e  $\theta_2$  (os respectivos IC's não contém o valor zero).

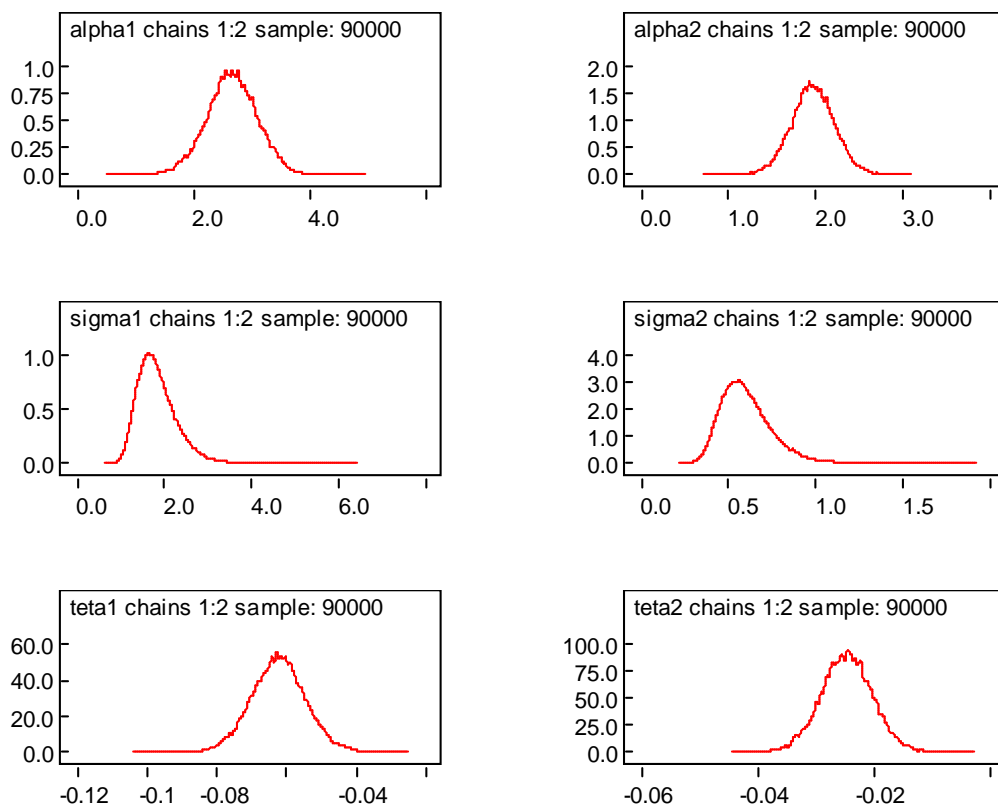


Figura 5: Gráficos das densidades geradas pelo amostrador de Gibbs

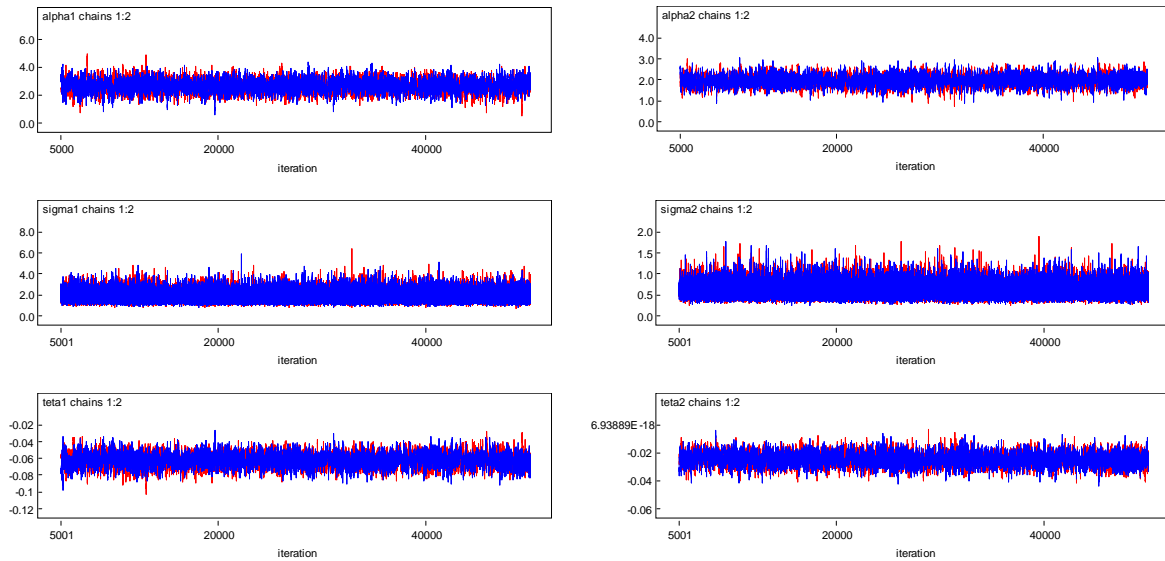


Figura 6: Gráficos dos traços *a posteriori*

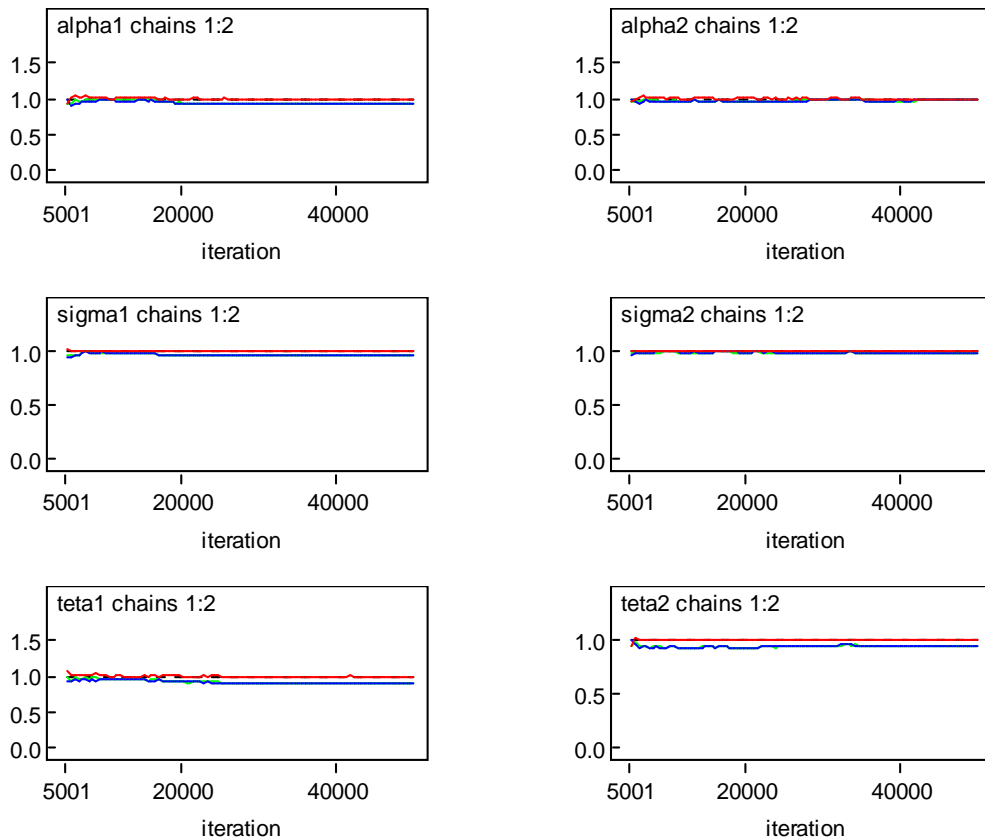


Figura 7: Teste gráfico de Gelman e Rubin

Nas Figuras 5, 6 e 7 temos as representações gráficas dos sumários *a posteriori*, gráficos das distribuições *a posteriori* condicionais aproximadas pelas amostras geradas via amostrador de Gibbs, gráficos de convergência das amostras geradas e o teste gráfico de Gelman e Rubin.

Observando a Figura 6, verificamos uniformidade dos gráficos das amostras geradas, indicando convergência para as cadeias.

Através da Figura 7, ou seja, o teste gráfico do critério de Gelman e Rubin, verificamos a convergência para todas as cadeias.

Para a geração das amostras da distribuição *a posteriori* (4.16) foi utilizado o *software* WinBugs.

Para a seleção de modelos, utilizamos as medidas de adequabilidade; Deviance Information Criterion (DIC) (ver Spiegelhalter et al. 2002) e Bayesian Information Criterion (BIC) dado em (3.4). Para cada modelo ajustado, precedeu-se o calculo desses critérios.

Na Tabela 5, temos estimativas Monte Carlo para o DIC e BIC para os modelos M1 (transformação ALR com erros correlacionados); M2 (transformação ALR com erros não correlacionados)

**Tabela 5:** Medidas de adequabilidade para os modelos

<b>Modelos</b>	<b>DIC</b>	<b>BIC</b>
M1	178.43	26.99437
M2	229.492	24.64422

Observamos que o modelo com erros correlacionados se adequou melhor aos dados da Tabela 2.

## Capítulo 5

# Uma Análise Bayesiana para os Dados do Oceano Ártico considerando a Transformação de Box-Cox.

Para análise dos dados da Tabela 2, ou seja, o caso particular, onde temos três componentes na presença de uma covariável,  $x_{i1}, x_{i2}, x_{i3}$ ,  $i = 1, \dots, n$  consideraremos agora a transformação Box-Cox (2.3).

Neste capítulo assumimos o modelo (2.1) e uma análise Bayesiana via MCMC é desenvolvida para determinar as quantidades *a posteriori*.

Consideremos a transformação de Box-Cox (2.3), para os dados introduzidos na Tabela 2, temos

$$y_{i1}^{(\lambda_1)} = \frac{\left(\frac{x_{i1}}{x_{i3}}\right)^{\lambda_1} - 1}{\lambda_1} \text{ se } \lambda_1 \neq 0 \text{ e } y_{i1}^{(\lambda_1)} = \log\left(\frac{x_{i1}}{x_{i3}}\right) \text{ se } \lambda_1 = 0,$$
$$y_{i2}^{(\lambda_2)} = \frac{\left(\frac{x_{i2}}{x_{i3}}\right)^{\lambda_2} - 1}{\lambda_2} \text{ se } \lambda_2 \neq 0 \text{ e } y_{i2}^{(\lambda_2)} = \log\left(\frac{x_{i2}}{x_{i3}}\right) \text{ se } \lambda_2 = 0.$$

Observar que a transformação ALR é um caso particular da transformação Box-Cox.



Assumir o modelo (2.1) para os dados transformados  $y_{i1}^{(\lambda_1)}$  e  $y_{i2}^{(\lambda_2)}$ , ou seja,

$$\begin{aligned} y_{i1}^{(\lambda_1)} &= \alpha_1 + \theta_1 Z_i + \varepsilon_{i1} \\ y_{i2}^{(\lambda_2)} &= \alpha_2 + \theta_2 Z_i + \varepsilon_{i2}, \end{aligned} \quad (5.1)$$

onde  $i = 1, \dots, n$  e  $\varepsilon_{i1}$  e  $\varepsilon_{i2}$  são os erros. Consideremos erros dependentes com uma distribuição normal bivariada  $N\left\{0, \Sigma\right\}$ , sendo  $\Sigma$  a matriz de variância-covariância dada em (4.2), ou seja,

$$\varepsilon_i \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right\}$$

onde

$$\begin{aligned} \varepsilon_{i1} &= y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i \\ \varepsilon_{i2} &= y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i. \end{aligned}$$

Denotamos este modelo por Modelo M3.

Considerando o modelo (5.1) e a transformação Box-Cox, as proporções verdadeiras dos componentes são dadas por  $\beta_{i1}, \beta_{i2}$  e  $\beta_{i3}$ ,  $i = 1, \dots, n$ , onde  $\beta_{i1} + \beta_{i2} + \beta_{i3} = 1$  e  $\beta_{i1} > 0, \beta_{i2} > 0$  e  $\beta_{i3} > 0$ . Dai temos,

$$\frac{\left(\frac{\beta_{i1}}{\beta_{i3}}\right)^{\lambda_1} - 1}{\lambda_1} = \theta_1 Z_i + \alpha_1 \quad \text{e} \quad \frac{\left(\frac{\beta_{i2}}{\beta_{i3}}\right)^{\lambda_2} - 1}{\lambda_2} = \theta_2 Z_i + \alpha_2, \quad (5.2)$$

ou

$$\frac{\beta_{i1}}{\beta_{i3}} = [\lambda_1 (\theta_1 Z_i + \alpha_1) + 1]^{1/\lambda_1} \quad \text{e} \quad \frac{\beta_{i2}}{\beta_{i3}} = [\lambda_2 (\theta_2 Z_i + \alpha_2) + 1]^{1/\lambda_2}. \quad (5.3)$$

Como  $\beta_{i1} + \beta_{i2} + \beta_{i3} = 1$ , dividindo a igualdade por  $\beta_{i3}$  encontramos,

$$\frac{\beta_{i1}}{\beta_{i3}} + \frac{\beta_{i2}}{\beta_{i3}} + 1 = \frac{1}{\beta_{i3}}. \quad (5.4)$$

Substituindo (5.3) em (5.4) temos,

$$1/\beta_{i3} = [\lambda_1 (\theta_1 Z_i + \alpha_1) + 1]^{1/\lambda_1} + [\lambda_2 (\theta_2 Z_i + \alpha_2) + 1]^{1/\lambda_2} + 1.$$

Portanto, temos as seguintes relações:

$$\begin{aligned} \beta_{i1} &= \frac{[\lambda_1 (\theta_1 Z_i + \alpha_1) + 1]^{1/\lambda_1}}{[\lambda_1 (\theta_1 Z_i + \alpha_1) + 1]^{1/\lambda_1} + [\lambda_2 (\theta_2 Z_i + \alpha_2) + 1]^{1/\lambda_2} + 1}, \\ \beta_{i2} &= \frac{[\lambda_2 (\theta_2 Z_i + \alpha_2) + 1]^{1/\lambda_2}}{[\lambda_1 (\theta_1 Z_i + \alpha_1) + 1]^{1/\lambda_1} + [\lambda_2 (\theta_2 Z_i + \alpha_2) + 1]^{1/\lambda_2} + 1} \end{aligned} \quad (5.5)$$

e

$$\beta_{i3} = \frac{1}{[\lambda_1 (\theta_1 Z_i + \alpha_1) + 1]^{1/\lambda_1} + [\lambda_2 (\theta_2 Z_i + \alpha_2) + 1]^{1/\lambda_2} + 1},$$

onde  $i = 1, \dots, n$ .

## 5.1 Uma análise Bayesiana considerando erros correlacionados

### 5.1.1 Função de Verossimilhança

Considerando  $y_{i1}^{(\lambda_1)}$  e  $y_{i2}^{(\lambda_2)}$  dados observados transformados pela função Box Cox, o modelo (5.1) e assumindo que as variáveis aleatórias  $\varepsilon_{i1}$  e  $\varepsilon_{i2}$  têm distribuição normal bivariada, a função densidade de probabilidade conjunta de  $y_{i1}^{(\lambda_1)}$  e  $y_{i2}^{(\lambda_2)}$  é dada por

$$f \left( y_{i1}^{(\lambda_1)}, y_{i2}^{(\lambda_2)} \right) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{\varepsilon_{i1}^2}{\sigma_1^2} - \frac{2\rho\varepsilon_{i1}\varepsilon_{i2}}{\sigma_1\sigma_2} + \frac{\varepsilon_{i2}^2}{\sigma_2^2} \right] \right\}, \quad (5.6)$$

onde

$$\begin{aligned} \varepsilon_{i1} &= y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i; \\ \varepsilon_{i2} &= y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i; \\ \sum_{i=1}^n \varepsilon_{i1}^2 &= \sum_{i=1}^n (y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i)^2 = \sum_{i=1}^n (y_{i1}^{(\lambda_1)})^2 - 2\alpha_1 \sum_{i=1}^n y_{i1}^{(\lambda_1)} - 2\theta_1 \sum_{i=1}^n Z_i y_{i1}^{(\lambda_1)} + n\alpha_1^2 + \\ &2\theta_1\alpha_1 \sum_{i=1}^n Z_i + \theta_1^2 \sum_{i=1}^n Z_i^2; \\ \sum_{i=1}^n \varepsilon_{i2}^2 &= \sum_{i=1}^n (y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i)^2 = \sum_{i=1}^n (y_{i2}^{(\lambda_2)})^2 - 2\alpha_2 \sum_{i=1}^n y_{i2}^{(\lambda_2)} - 2\theta_2 \sum_{i=1}^n Z_i y_{i2}^{(\lambda_2)} + n\alpha_2^2 + \\ &2\theta_2\alpha_2 \sum_{i=1}^n Z_i + \theta_2^2 \sum_{i=1}^n Z_i^2; \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} &= \sum_{i=1}^n (y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i)(y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i) = \sum_{i=1}^n y_{i1}^{(\lambda_1)} y_{i2}^{(\lambda_2)} - \alpha_2 \sum_{i=1}^n y_{i1}^{(\lambda_1)} - \\ \theta_2 \sum_{i=1}^n Z_i y_{i1}^{(\lambda_1)} - \alpha_1 \sum_{i=1}^n y_{i2}^{(\lambda_2)} + n\alpha_1 \alpha_2 + \theta_2 \alpha_1 \sum_{i=1}^n Z_i + \\ \theta_1 \sum_{i=1}^n Z_i y_{i2}^{(\lambda_2)} + \theta_1 \alpha_2 \sum_{i=1}^n Z_i + \theta_1 \theta_2 \sum_{i=1}^n Z_i^2. \end{aligned}$$

Considerando  $y_{i1} = x_{i1}/x_{i3}$  e  $y_{i2} = x_{i2}/x_{i3}$ , razão dos dados composicionais sem considerar a transformação Box-Cox, a função densidade de probabilidade de  $y_{i1}$  e  $y_{i2}$  não transformados (variáveis originais) é dada por

$$\begin{aligned} f(y_{i1}, y_{i2}) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{\varepsilon_{i1}^2}{\sigma_1^2} - \frac{2\rho\varepsilon_{i1}\varepsilon_{i2}}{\sigma_1\sigma_2} + \frac{\varepsilon_{i2}^2}{\sigma_2^2} \right] \right\} \times \\ &\times \left( \frac{dy_{i1}^{(\lambda_1)}}{dy_{i1}} \right) \left( \frac{dy_{i2}^{(\lambda_2)}}{dy_{i2}} \right), \end{aligned}$$

onde  $\left( \frac{dy_{i1}^{(\lambda_1)}}{dy_{i1}} \right)$  e  $\left( \frac{dy_{i2}^{(\lambda_2)}}{dy_{i2}} \right)$  são os elementos diagonais do jacobiano da transformação, com  $\frac{dy_{i1}^{(\lambda_1)}}{dy_{i1}} = \frac{\lambda_1 y_{i1}^{\lambda_1-1}}{\lambda_1} = y_{i1}^{\lambda_1-1}$  e  $\frac{dy_{i2}^{(\lambda_2)}}{dy_{i2}} = \frac{\lambda_2 y_{i2}^{\lambda_2-1}}{\lambda_2} = y_{i2}^{\lambda_2-1}$

Sendo assim, a função de verossimilhança para  $v_3 = (\lambda_1, \lambda_2, \alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho)$  é dada por,

$$\begin{aligned} L(v_3) &\propto (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} (1-\rho^2)^{-n/2} \left( \prod_{i=1}^n y_{i1}^{\lambda_1} \right) \left( \prod_{i=1}^n y_{i2}^{\lambda_2} \right) \times \\ &\times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n \varepsilon_{i1}\varepsilon_{i2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right] \right\}, \end{aligned} \quad (5.7)$$

onde  $\left( \prod_{i=1}^n y_{i1}^{\lambda_1} \right)$  e  $\left( \prod_{i=1}^n y_{i2}^{\lambda_2} \right)$  são os produtos do jacobiano.

### 5.1.2 Informação *a Priori*

Para uma análise Bayesiana, assumir as distribuições *a priori* para os parâmetros  $(\lambda_k, \theta_k, \alpha_k, \sigma_k^2, \text{ e } \rho, k = 1, 2)$  dadas por,

$$\begin{aligned}
 \text{i) } \lambda_k &\sim N(g_k, h_k^2); k = 1, 2; g_k, h_k^2 \text{ conhecidos,} \\
 \text{ii) } \theta_k &\sim N\{a_k, b_k^2\}; k = 1, 2; a_k, b_k^2 \text{ conhecidos,} \\
 \text{iii) } \alpha_k &\sim N\{c_k, d_k^2\}; k = 1, 2; c_k, d_k^2 \text{ conhecidos,} \\
 \text{iv) } \sigma_k^2 &\sim IG[e_k, f_k]; k = 1, 2; e_k, f_k \text{ conhecidos,} \\
 \text{v) } \rho &\sim U[-1, 1],
 \end{aligned} \tag{5.8}$$

onde  $N\{\mu, \sigma^2\}$  denota uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ ,  $IG[a, b]$  denota uma distribuição gama inversa com média  $b/(a-1)$  e variância  $b^2/[(a-1)^2(a-2)]$ ,  $a > 2$  e  $U[c, d]$ , denota uma distribuição uniforme no intervalo  $[c, d]$ .

Também assumir independência *a priori* entre os parâmetros.

### 5.1.3 *Posteriori* Conjunta

Considerando a verossimilhança (5.7) e as distribuições *a priori* (5.8), a *posteriori* conjunta para  $v_3 = (\lambda_1, \lambda_2, \alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho)$  é dada por

$$\begin{aligned}
 \Pi(v_3 \mid \text{dados}) &\propto \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2h_k^2}(\lambda_k - g_k)^2\right] \right\} \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2d_k^2}(\alpha_k - c_k)^2\right] \right\} \times \\
 &\times \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2b_k^2}(\theta_k - a_k)^2\right] \right\} \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(e_k+1)} e^{-f_k/\sigma_k^2} \right\} \times \\
 &\times (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} (1 - \rho^2)^{-n/2} \left( \prod_{i=1}^n y_{i1}^{\lambda_1} \right) \left( \prod_{i=1}^n y_{i2}^{\lambda_2} \right) \times \\
 &\times \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n \varepsilon_{i1}\varepsilon_{i2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right] \right\}.
 \end{aligned} \tag{5.9}$$

### 5.1.4 Distribuições Condicionais necessárias para o amostrador de Gibbs

As distribuições necessárias para os algoritmos de Gibbs sampling e Metropolis-Hastings são dadas por,

$$\text{i) } \pi \left( \lambda_1 | v_{3(\lambda_1)}, D \right) \propto \exp \left\{ -\frac{1}{2h_1^2} (\lambda_1 - g_1)^2 \right\} \Psi_1 \left( v_3 \right),$$

$$\text{onde } \Psi_1 \left( v_3 \right) = \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_1^2} \left( \sum_{i=1}^n y_{i1}^{(\lambda_1)^2} - 2\alpha_1 \sum_{i=1}^n y_{i1}^{(\lambda_1)} - 2\theta_1 \sum_{i=1}^n Z_i y_{i1}^{(\lambda_1)} \right) - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n y_{i1}^{(\lambda_1)} D_i^{(2)} \right] + \lambda_1 \sum_{i=1}^n \ln y_{i1} \right\}$$

$$\text{onde } D_i^{(2)} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i, \quad i = 1, \dots, n;$$

$$\text{ii) } \pi \left( \lambda_2 | v_{3(\lambda_2)}, D \right) \propto \exp \left\{ -\frac{1}{2h_2^2} (\lambda_2 - g_2)^2 \right\} \Psi_2 \left( v_3 \right), \quad (5.10)$$

$$\text{onde } \Psi_2 \left( v_3 \right) = \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_2^2} \left( \sum_{i=1}^n y_{i2}^{(\lambda_2)^2} - 2\alpha_2 \sum_{i=1}^n y_{i2}^{(\lambda_2)} - 2\theta_2 \sum_{i=1}^n Z_i y_{i2}^{(\lambda_2)} \right) - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n y_{i2}^{(\lambda_2)} D_i^{(1)} \right] + \lambda_2 \sum_{i=1}^n \ln y_{i2} \right\}$$

$$\text{onde } D_i^{(1)} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i, \quad i = 1, \dots, n;$$

$$\text{iii) } \theta_1 | v_{3(\theta_1)}, D \sim N \left\{ \frac{a_1 \sigma_1^2 \sigma_2 (1 - \rho^2) + \sigma_2 b_1^2 \sum_{i=1}^n Z_i u_i^{(1)} - \sigma_1 b_1^2 \rho \sum_{i=1}^n Z_i B_i^{(1)}}{\sigma_2 \left[ \sigma_1^2 (1 - \rho^2) + b_1^2 \sum_{i=1}^n Z_i^2 \right]}, \frac{b_1^2 \sigma_1^2 (1 - \rho^2)}{\sigma_1^2 (1 - \rho^2) + b_1^2 \sum_{i=1}^n Z_i^2} \right\},$$

onde  $u_i^{(1)} = y_{i1}^{(\lambda_1)} - \alpha_1$ ,  $B_i^{(1)} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i$ ;  $i = 1, \dots, n$  e  $v_{3(\theta_1)}$  é o vetor de todos os parâmetros exceto  $\theta_1$ .

$$\text{iv) } \theta_2 | v_{3(\theta_2)}, D \sim N \left\{ \frac{a_2 \sigma_2^2 \sigma_1 (1 - \rho^2) + \sigma_1 b_2^2 \sum_{i=1}^n Z_i u_i^{(2)} - \sigma_2 b_2^2 \rho \sum_{i=1}^n Z_i B_i^{(2)}}{\sigma_1 \left[ \sigma_2^2 (1 - \rho^2) + b_2^2 \sum_{i=1}^n Z_i^2 \right]}, \frac{b_2^2 \sigma_2^2 (1 - \rho^2)}{\sigma_2^2 (1 - \rho^2) + b_2^2 \sum_{i=1}^n Z_i^2} \right\},$$

onde  $u_i^{(2)} = y_{i2}^{(\lambda_2)} - \alpha_2$ ,  $B_i^{(2)} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i$ ;  $i = 1, \dots, n$  e  $v_{3(\theta_2)}$  é o vetor de todos os parâmetros exceto  $\theta_2$ .

$$v) \alpha_1 | v_{3(\alpha_1)}, D \sim N \left\{ \frac{c_1 \sigma_1^2 \sigma_2 (1 - \rho^2) + \sigma_2 d_1^2 \sum_{i=1}^n \xi_i^{(1)} - \sigma_1 d_1^2 \rho \sum_{i=1}^n C_i^{(1)}}{\sigma_2 [\sigma_1^2 (1 - \rho^2) + n d_1^2]}, \frac{d_1^2 \sigma_1^2 (1 - \rho^2)}{\sigma_1^2 (1 - \rho^2) + n d_1^2} \right\},$$

onde  $\xi_i^{(1)} = y_{i1}^{(\lambda_1)} - \theta_1 Z_i$ ,  $C_i^{(1)} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i$ ;  $i = 1, \dots, n$  e  $v_{3(\alpha_1)}$  é o vetor de todos os parâmetros exceto  $\alpha_1$ .

$$vi) \alpha_2 | v_{3(\alpha_2)}, D \sim N \left\{ \frac{c_2 \sigma_2^2 \sigma_1 (1 - \rho^2) + \sigma_1 d_2^2 \sum_{i=1}^n \xi_i^{(2)} - \sigma_2 d_2^2 \rho \sum_{i=1}^n C_i^{(2)}}{\sigma_2 [\sigma_1^2 (1 - \rho^2) + n d_1^2]}, \frac{d_2^2 \sigma_2^2 (1 - \rho^2)}{\sigma_2^2 (1 - \rho^2) + n d_2^2} \right\},$$

onde  $\xi_i^{(2)} = y_{i2}^{(\lambda_2)} - \theta_2 Z_i$ ,  $C_i^{(2)} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i$ ;  $i = 1, \dots, n$  e  $v_{3(\alpha_2)}$  é o vetor de todos os parâmetros exceto  $\alpha_2$ .

$$vii) \pi \left( \sigma_1^2 | v_{3(\sigma_1^2)}, D \right) \propto (\sigma_1^2)^{-(e_1+1)} e^{-f_1/\sigma_1^2} \Psi_1 \left( v_3 \right),$$

$$\text{onde } \Psi_1 \left( v_3 \right) = \exp \left\{ -\frac{n}{2} \ln(\sigma_1^2) - \frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1 \sigma_2} \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} \right] \right\},$$

$\varepsilon_{i1} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i$ ,  $i = 1, \dots, n$  e  $v_{3(\sigma_1^2)}$  é o vetor de todos os parâmetros exceto  $\sigma_1^2$ .

$$viii) \pi \left( \sigma_2^2 | v_{3(\sigma_2^2)}, D \right) \propto (\sigma_2^2)^{-(e_2+1)} e^{-f_2/\sigma_2^2} \Psi_2 \left( v_3 \right),$$

$$\text{onde } \Psi_2 \left( v_3 \right) = \exp \left\{ -\frac{n}{2} \ln(\sigma_2^2) - \frac{1}{2(1-\rho^2)} \left[ \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 - \frac{2\rho}{\sigma_1 \sigma_2} \sum_{i=1}^n \varepsilon_{i1} \varepsilon_{i2} \right] \right\},$$

$\varepsilon_{i1} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i$ ,  $i = 1, \dots, n$  e  $v_{3(\sigma_2^2)}$  é o vetor de todos os parâmetros exceto  $\sigma_2^2$ .

$$\text{ix) } \pi \left( \rho | v_{3(\rho)}, D \right) \propto (1 - \rho^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n \varepsilon_{i1}\varepsilon_{i2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right] \right\},$$

onde  $\varepsilon_{i1} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i$ ,  $i = 1, \dots, n$  e  $v_{3(\rho)}$  é o vetor de todos os parâmetros exceto  $\rho$ .

Para gerar  $\alpha_1, \alpha_2, \theta_1, \theta_2$  utilizamos o algoritmo Gibbs sampling e para gerar  $\lambda_1, \lambda_2, \sigma_1^2, \sigma_2^2$  e  $\rho$  precisamos usar o algoritmo de Metropolis- Hastings.

## 5.2 Uma análise Bayesiana considerando erros não correlacionados

### 5.2.1 Função de Verossimilhança

Considerando a transformação Box-Cox para os dados da Tabela 2, e o modelo (5.1), agora assumimos erros não correlacionados, ou seja  $\rho = 0$ , onde as variáveis aleatórias  $\varepsilon_{i1}$  e  $\varepsilon_{i2}$  são independentes com distribuições normais com médias iguais a zero e variâncias  $\sigma_1^2$  e  $\sigma_2^2$ . Observar que temos um caso particular do Modelo M3, com  $\rho = 0$  (erros não correlacionados), que denotamos por Modelo M4.

De (5.1) temos,

$$\begin{aligned} \varepsilon_{i1} &= y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i, \\ \varepsilon_{i2} &= y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i. \end{aligned} \tag{5.11}$$

Considerando as suposições de normalidade e independência observamos que  $y_{i1}^{(\lambda_1)} \sim N \{ \alpha_1 + \theta_1 Z_i; \sigma_1^2 \}$  e  $y_{i2}^{(\lambda_2)} \sim N \{ \alpha_2 + \theta_2 Z_i; \sigma_2^2 \}$ .

A função densidade de probabilidade conjunta de  $y_{i1}^{(\lambda_1)}$  e  $y_{i2}^{(\lambda_2)}$  é dada por

$$f\left(y_{i1}^{(\lambda_1)}, y_{i2}^{(\lambda_2)}\right) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2} \left(y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i\right)^2\right\} \times \quad (5.12)$$

$$\times \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2\sigma_2^2} \left(y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i\right)^2\right\}.$$

A função densidade de probabilidade para  $y_{i1} = x_{i1}/x_{i3}$  e  $y_{i2} = x_{i2}/x_{i3}$  não transformados (variáveis originais) é dada por

$$f(y_{i1}, y_{i2}) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2} \left(y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i\right)^2\right\} \left(\frac{dy_{i1}^{(\lambda_1)}}{dy_{i1}}\right) \times$$

$$\times \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2\sigma_2^2} \left(y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i\right)^2\right\} \left(\frac{dy_{i2}^{(\lambda_2)}}{dy_{i2}}\right),$$

onde  $\frac{dy_{i1}^{(\lambda_1)}}{dy_{i1}} = \frac{\lambda_1 y_{i1}^{\lambda_1-1}}{\lambda_1} = y_{i1}^{\lambda_1-1}$  e  $\frac{dy_{i2}^{(\lambda_2)}}{dy_{i2}} = \frac{\lambda_2 y_{i2}^{\lambda_2-1}}{\lambda_2} = y_{i2}^{\lambda_2-1}$ .

Portanto com os dados originais a distribuição conjunta de  $y_i = (y_{i1}, y_{i2})$  para  $i = 1, \dots, n$ , ou seja,  $\underline{y}_1 = (y_{11}, y_{12})$ ,  $\underline{y}_2 = (y_{21}, y_{22})$ ,  $\underline{y}_3 = (y_{31}, y_{32})$ , ...,  $\underline{y}_n = (y_{n1}, y_{n2})$  é dada por:

$$f\left(\underline{y}_1, \dots, \underline{y}_n\right) \propto (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} \left(\prod_{i=1}^n y_{i1}^{\lambda_1}\right) \left(\prod_{i=1}^n y_{i2}^{\lambda_2}\right) \times$$

$$\times \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2\right\},$$

onde  $\left(\prod_{i=1}^n y_{i1}^{\lambda_1}\right)$  e  $\left(\prod_{i=1}^n y_{i2}^{\lambda_2}\right)$  são os produtos do jacobiano,  $\varepsilon_{i1} = y_{i1}^{(\lambda_1)} - \alpha_1 - \theta_1 Z_i$ ,  $\varepsilon_{i2} = y_{i2}^{(\lambda_2)} - \alpha_2 - \theta_2 Z_i$ ,  $y_{i1}^{(\lambda_1)} = \left[(x_{i1}/x_{i3})^{\lambda_1} - 1\right] / \lambda_1$  se  $\lambda_1 \neq 0$  e  $y_{i2}^{(\lambda_2)} = \left[(x_{i2}/x_{i3})^{\lambda_2} - 1\right] / \lambda_2$  se  $\lambda_2 \neq 0$ .

Neste caso a função de verossimilhança para  $v_4 = (\lambda_1, \lambda_2, \alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$  é dada por

$$L\left(v_4\right) \propto (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} \left(\prod_{i=1}^n y_{i1}^{\lambda_1}\right) \left(\prod_{i=1}^n y_{i2}^{\lambda_2}\right) \times \quad (5.13)$$

$$\times \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2\right\}.$$



### 5.2.2 Informação *a Priori*

Supondo independência entre os parâmetros  $\lambda_1, \lambda_2, \alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2$  e  $\sigma_2^2$ , consideramos as seguintes distribuições *a priori*,

$$\begin{aligned} \text{i) } \lambda_k &\sim N(g_k, h_k^2); k = 1, 2; g_k, h_k^2 \text{ conhecidos,} \\ \text{ii) } \theta_k &\sim N\{a_k, b_k^2\}; k = 1, 2; a_k, b_k^2 \text{ conhecidos,} \\ \text{iii) } \alpha_k &\sim N\{c_k, d_k^2\}; k = 1, 2; c_k, d_k^2 \text{ conhecidos,} \\ \text{iv) } \sigma_k^2 &\sim IG[e_k, f_k]; k = 1, 2; e_k, f_k \text{ conhecidos.} \end{aligned} \quad (5.14)$$

### 5.2.3 *Posteriori Conjunta*

Considerando a função de verossimelhança (5.13) e as *prioris* (5.14) a distribuição *a posteriori* conjunta de  $v_4 = (\lambda_1, \lambda_2, \alpha_1, \alpha_2, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$  é dada por

$$\begin{aligned} \Pi(v_4 | \text{dados}) &\propto \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2h_k^2}(\lambda_k - g_k)^2\right] \right\} \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2d_k^2}(\alpha_k - c_k)^2\right] \right\} \times \\ &\times \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2b_k^2}(\theta_k - a_k)^2\right] \right\} \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(e_k+1)} e^{-f_k/\sigma_k^2} \right\} \times \\ &\times (\sigma_1^2)^{-n/2} (\sigma_2^2)^{-n/2} \left( \prod_{i=1}^n y_{i1}^{\lambda_1} \right) \left( \prod_{i=1}^n y_{i2}^{\lambda_2} \right) \times \\ &\times \exp\left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n \varepsilon_{i1}^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n \varepsilon_{i2}^2 \right\}. \end{aligned} \quad (5.15)$$

### 5.2.4 Distribuições Condicionais para o amostrador de Gibbs

Dada a distribuição à *posteriori* (5.15) as distribuições condicionais dos parâmetros necessárias para o algoritmo Gibbs sampling e Metropolis-Hastings são dadas por

$$\text{i) } \pi\left(\lambda_k | v_{4(\lambda_k)}, D\right) \propto \exp\left\{-\frac{1}{2h_k^2}(\lambda_k - g_k)^2\right\} \Psi_k\left(v_4\right),$$

onde  $\Psi_k \left( \underset{\sim}{v}_4 \right) = \exp \left\{ -\frac{1}{\sigma_k^2} \left( \sum_{i=1}^n y_{ik}^{(\lambda_k)^2} - 2\alpha_k \sum_{i=1}^n y_{ik}^{(\lambda_k)} - 2\theta_k \sum_{i=1}^n Z_i y_{ik}^{(\lambda_k)} \right) + \lambda_k \sum_{i=1}^n \ln y_{ik} \right\}$  e  $k = 1, 2$ .

$$\text{ii) } \theta_k | v_{4(\theta_k)}, D \sim N \left\{ \frac{a_k \sigma_k^2 + b_k^2 \sum_{i=1}^n Z_i u_i^{(k)}}{\sigma_k^2 + b_k^2 \sum_{i=1}^n Z_i^2}; \frac{b_k^2 \sigma_k^2}{\sigma_k^2 + b_k^2 \sum_{i=1}^n Z_i^2} \right\},$$

onde  $k = 1, 2$  e  $u_i^{(k)} = y_{ik}^{(\lambda_k)} - \alpha_k; i = 1, \dots, n$ .

$$\text{iii) } \alpha_k | v_{4(\alpha_k)}, D \sim N \left\{ \frac{c_k \sigma_k^2 + d_k^2 \sum_{i=1}^n \xi_i^{(k)}}{\sigma_k^2 + n d_k^2}; \frac{d_k^2 \sigma_k^2}{\sigma_k^2 + n d_k^2} \right\}, \quad (5.16)$$

onde  $k = 1, 2$  e  $\xi_i^{(k)} = y_{ik}^{(\lambda_k)} - \theta_k Z_i; i = 1, \dots, n$ .

$$\text{iv) } \sigma_k^2 | v_{4(\sigma_k^2)}, D \sim IG \left[ e_k + \frac{n}{2}; f_k + \frac{1}{2} \sum_{i=1}^n \varepsilon_{ik}^2 \right],$$

onde  $k = 1, 2$  e  $\varepsilon_{ik} = y_{ik}^{(\lambda_k)} - \alpha_k - \theta_k Z_i; i = 1, \dots, n$ .

Como as distribuições condicionais para  $\alpha_k, \theta_k, \sigma_k^2$  para  $k = 1, 2$  são conhecidas, usamos o algoritmo Gibbs sampling; para  $\lambda_k, k = 1, 2$ , usamos o algoritmo de Metropolis-Hastings.

## 5.3 Aplicação da Metodologia

### 5.3.1 Aplicação da metodologia considerando erros correlacionados

A metodologia é ilustrada considerando os dados da Tabela 2.

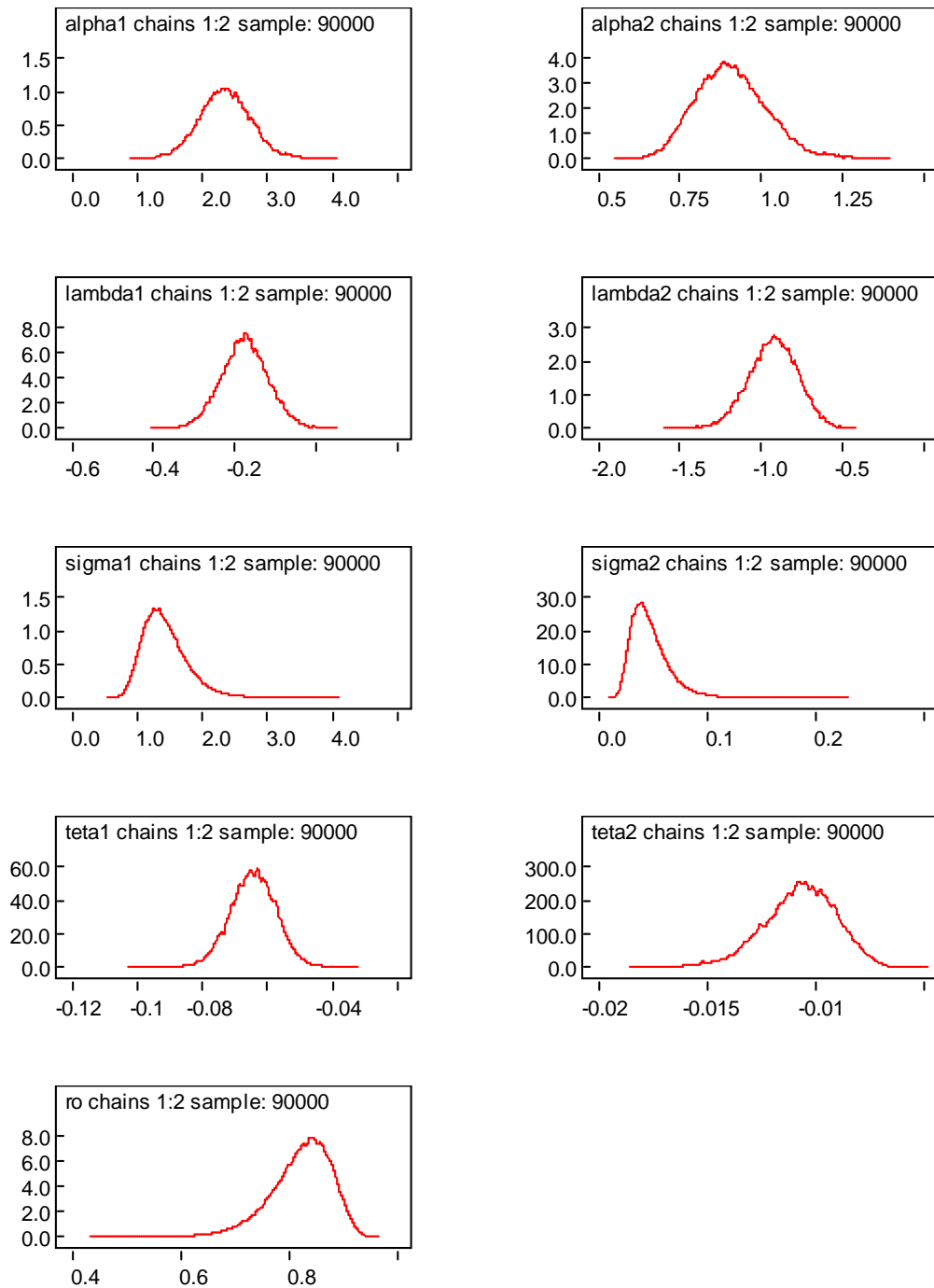
Assumindo o modelo M3, ou seja, transformação Box-Cox para os dados com erros normais correlacionados para o modelo (5.1), e distribuições *a priori* não informativas para  $\alpha_k, \theta_k, \sigma_k^2, \lambda_k$  e com hiperparâmetros  $a_1 = a_2 = c_1 = c_2 = g_1 = g_2 = 0$ ,  $b_1^2 = b_2^2 = d_1^2 = d_2^2 = h_1^2 = h_2^2 = 100000$ ,  $e_1 = e_2 = f_1 = f_2 = 1000$ , geramos duas cadeias paralelas com 50000 iterações cada uma onde foram descartadas as 5000 primeiras iter-

ações, considerando um salto de tamanho 20 para eliminar a correlação entre as amostras. Na geração das amostras, utilizamos o *software* Winbugs. A convergência do algoritmo foi verificada usando o método de Gelman e Rubin (1992). Na Tabela 6, temos os sumários *a posteriori* de interesse. Observamos a convergência do algoritmo Gibbs sampling ( $\sqrt{\hat{R}} < 1.1$  para todos os casos).

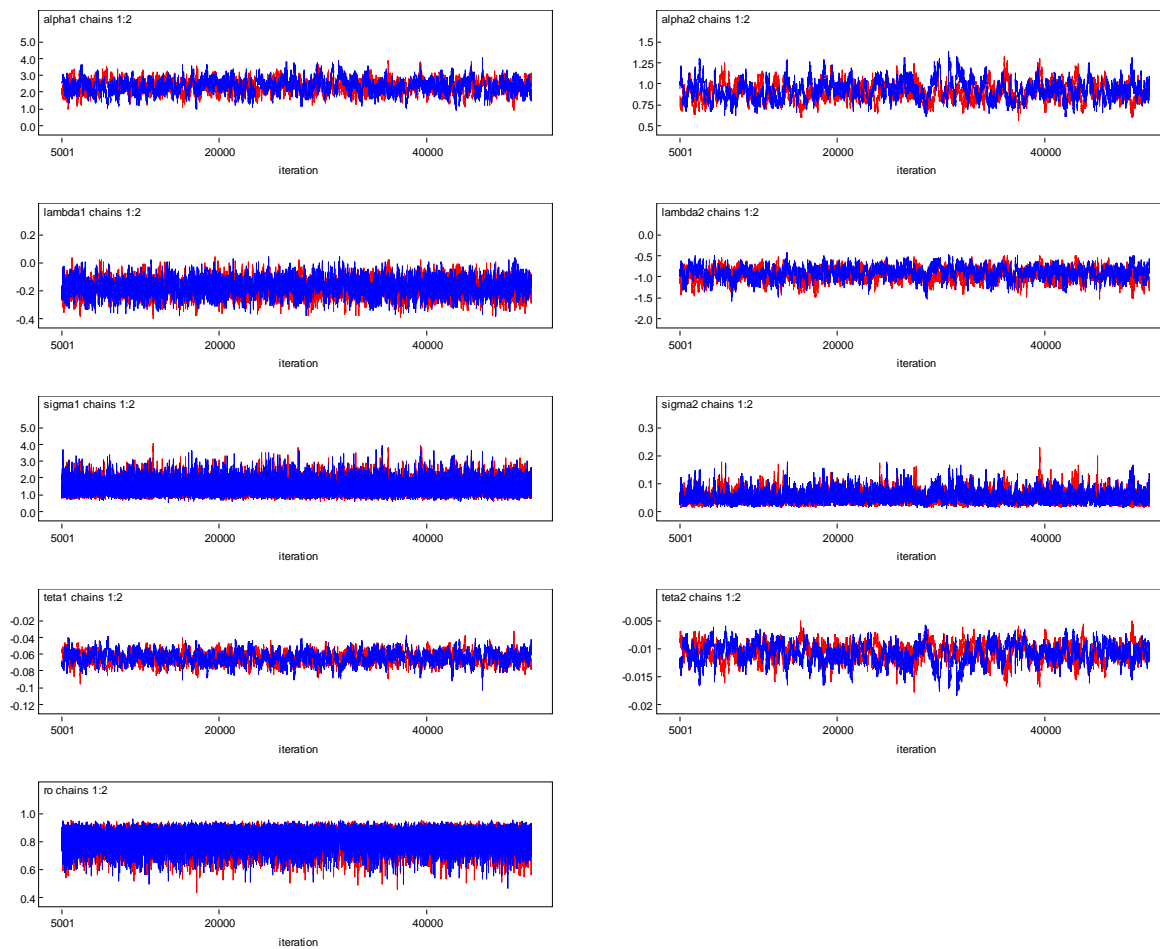
**Tabela 6:** Sumários *a posteriori* para o modelo M3

Parâmetros	Média	Int.Cred.95%	$\hat{R}$
$\alpha_1$	2.315	(1.544; 3.198)	0.999
$\alpha_2$	0.8948	(0.6998; 1.146)	0.994
$\theta_1$	-0.06382	(-0.07841; -0.04912)	1.001
$\theta_2$	-0.01049	(-0.01422; -0.007444)	1.001
$\sigma_1^2$	1.436	(0.902; 2.252)	1.001
$\sigma_2^2$	0.0466	(0.02205; 0.09293)	0.996
$\lambda_1$	-0.1782	(-0.2956; -0.05238)	1.000
$\lambda_2$	-0.9405	(-1.251; -0.6427)	1.002
$\rho$	0.8202	(0.6916; 0.908)	1.001

Dos resultados da Tabela 6, observamos que com a transformação Box-Cox, também temos efeitos de profundidade significativos.

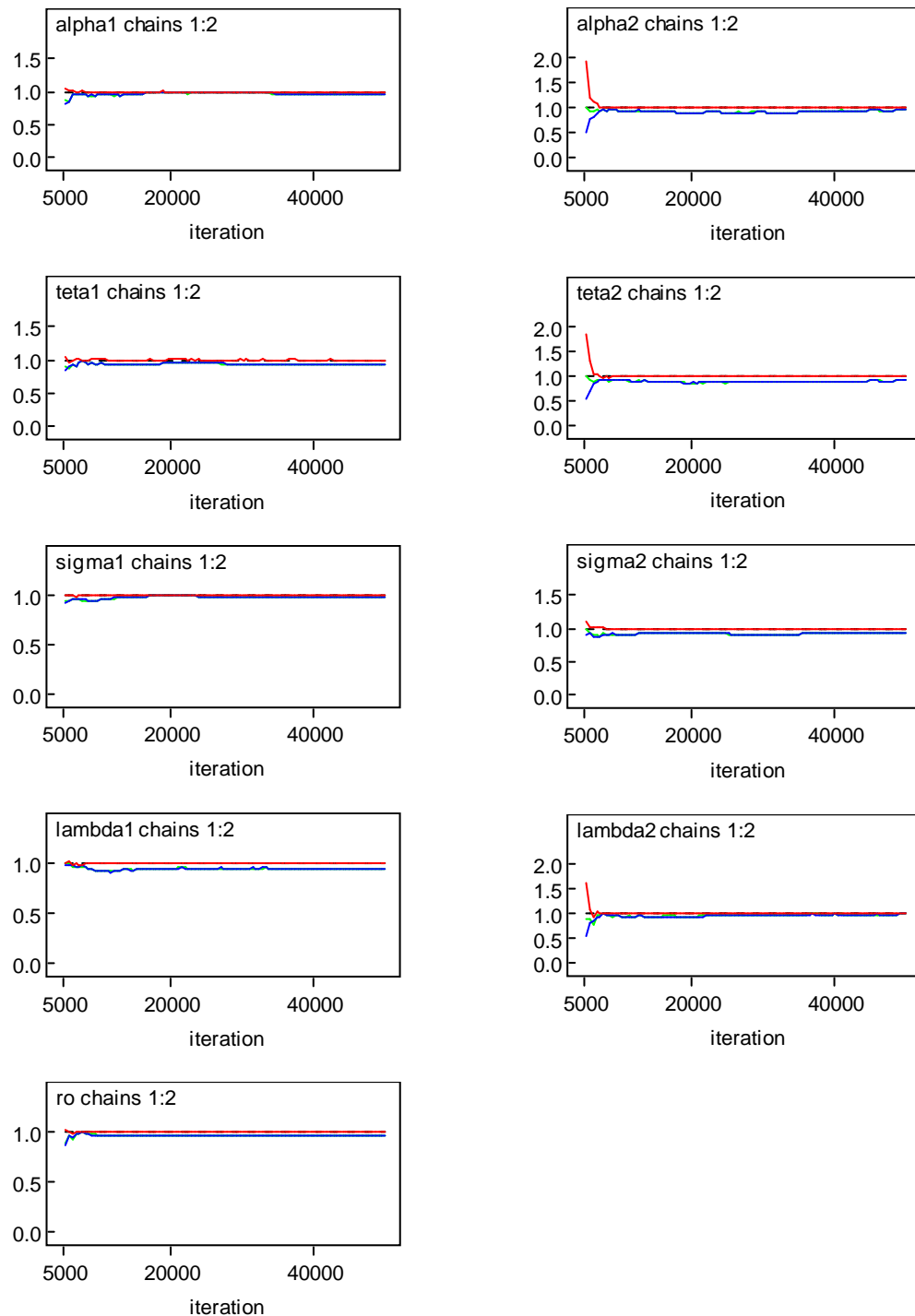


**Figura 8:** Gráficos das densidades geradas pelo amostrador de Gibbs



**Figura 9:** Gráficos dos traços *a posteriori*

Observando a Figura 9, verificamos uniformidade dos gráficos das amostras geradas, indicando convergência para as cadeias.



**Figura 10:** Teste gráfico de Gelman e Rubin

Nas Figuras 8, 9 e 10 temos as representações gráficas dos sumários *a posteriori*, gráficos das distribuições *a posteriori* condicionais aproximadas pelas amostras geradas via amostrador de Gibbs, gráficos de convergência das amostras geradas e o teste gráfico de Gelman e Rubin.

Através da Figura 10, ou seja, o teste gráfico do critério de Gelman e Rubin, verificamos

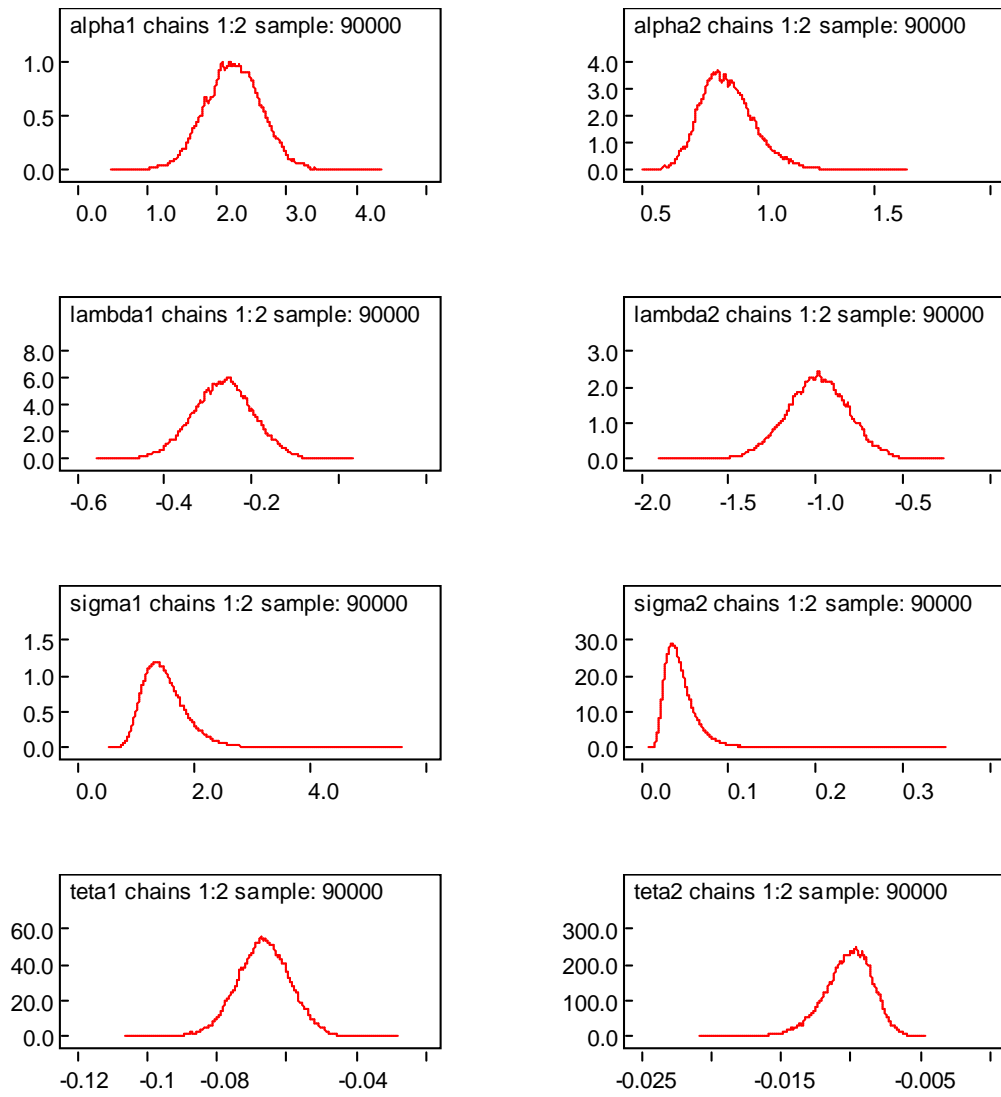
a convergência para todas as cadeias.

### 5.3.2 Aplicação da metodologia considerando erros não correlacionados

Assumindo os dados com transformação de Box-Cox e erros normais não correlacionados ( $\rho = 0$ ), ou seja, o modelo M4, temos na Tabela 7 os sumários *a posteriori* de interesse (mesmo esquema de simulação prévio e mesmos hiperparâmetros para as distribuições *a priori*). Também geramos duas cadeias paralelas com 50000 iterações cada uma onde foram descartadas as 5000 primeiras iterações, considerando um salto de tamanho 10. Observamos a convergência do algoritmo Gibbs sampling ( $\sqrt{\hat{R}} < 1.1$  para todos os casos).

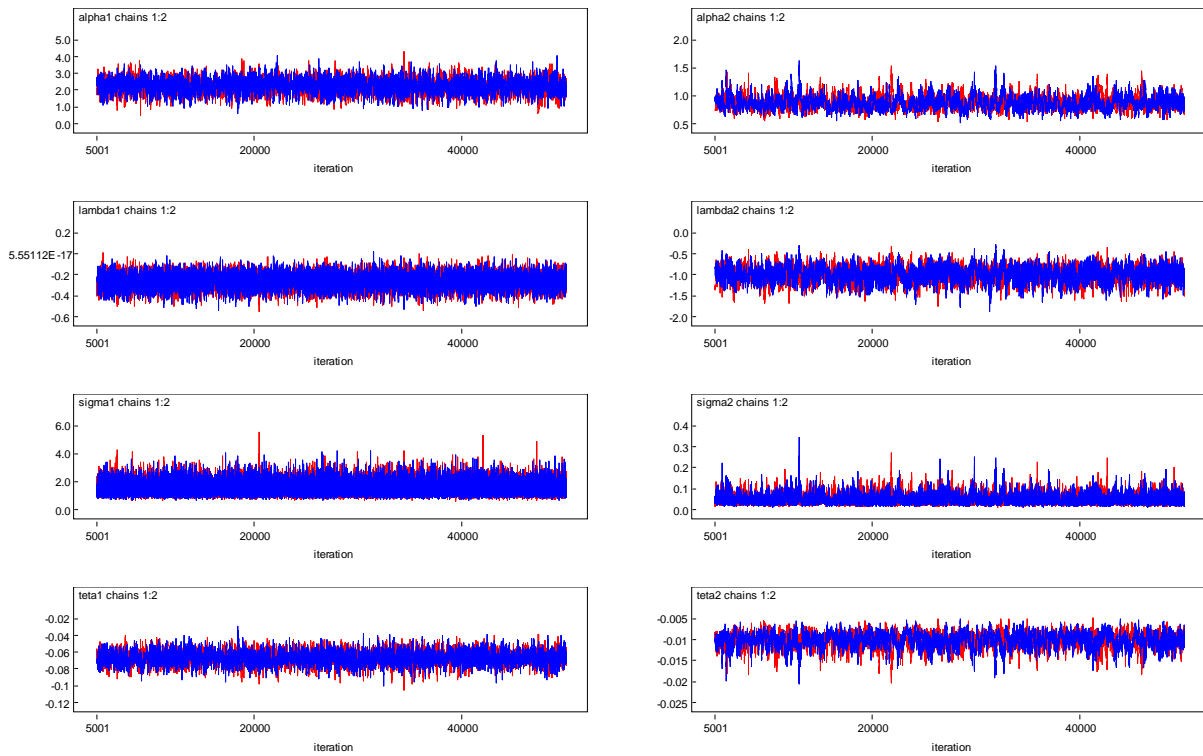
**Tabela 7:** Sumários *a posteriori*.

Parâmetros	Média	Int.Cred.95%	$\hat{R}$
$\alpha_1$	2.261	(1.481;3.005)	0.998
$\alpha_2$	0.8728	(0.6771; 1.158)	1.001
$\theta_1$	-0.067	(-0.08188; -0.05308)	0.983
$\theta_2$	-0.0102	(-0.01412; -0.007284)	1.020
$\sigma_1^2$	1.519	(0.9393; 2.436)	1.010
$\sigma_2^2$	0.04523	(0.02171; 0.09281)	0.996
$\lambda_1$	-0.2644	(-0.4053; -0.1309)	0.999
$\lambda_2$	-0.987	(-1.33; -0.6408)	1.000



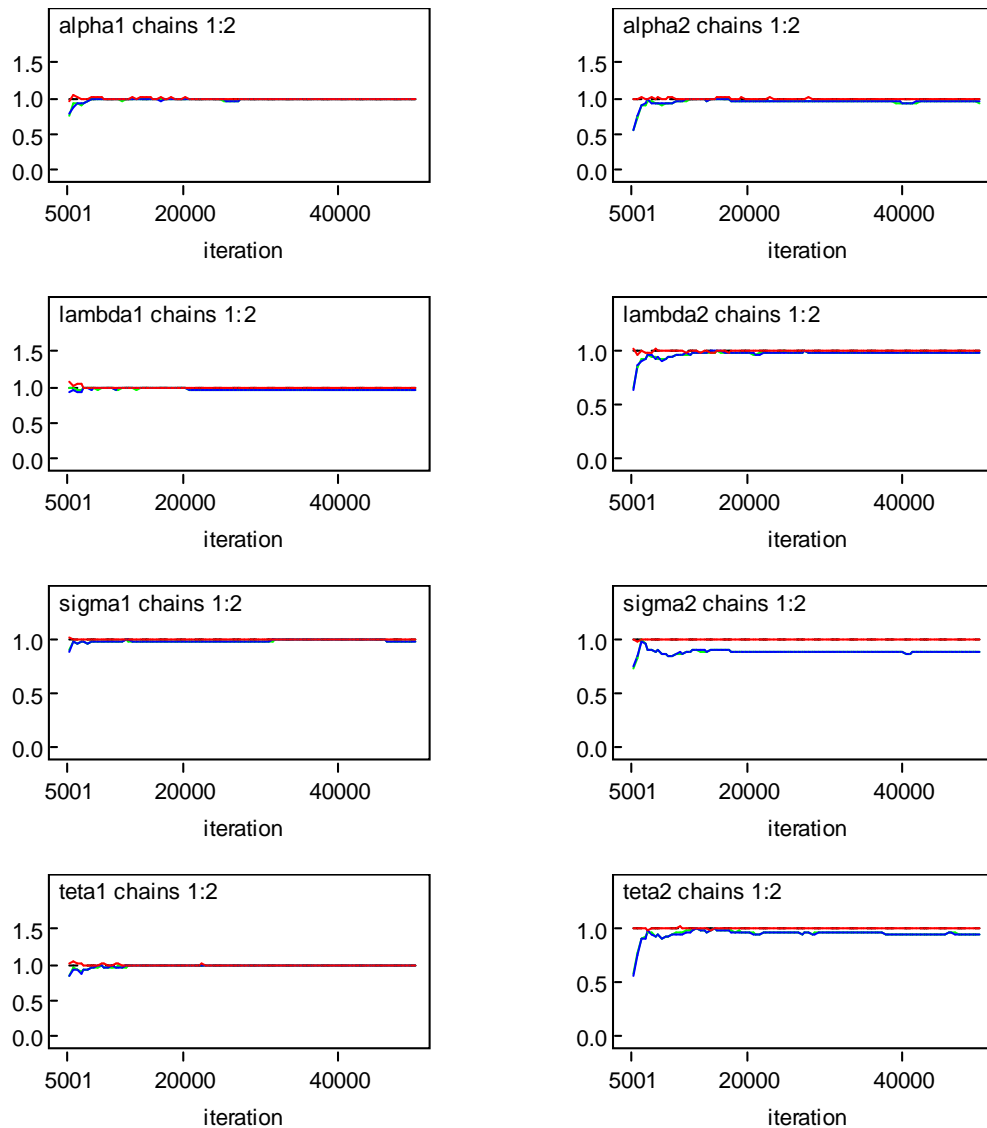
**Figura 11:** Gráficos das densidades geradas pelo amostrador de Gibbs





**Figura 12:** Traços *a posteriori*

Observando a Figura 12, verificamos uniformidade dos gráficos das amostras geradas, indicando convergência para as cadeias.



**Figura 13:** Teste gráfico de Gelman e Rubin

Nas Figuras 11, 12 e 13 temos as representações gráficas dos sumários *a posteriori*, gráficos das distribuições *a posteriori* condicionais aproximadas pelas amostras geradas via amostrador de Gibbs, gráficos de convergência das amostras geradas e o teste gráfico de Gelman e Rubin, onde verificamos a convergência para todas as cadeias.

### 5.3.3 Seleção de Modelos

Para a seleção dos modelos M3 (transformação Box-Cox com erros correlacionados) e M4 (transformação Box-Cox com erros não correlacionados), usamos os critérios DIC e BIC implementados no *software* Winbugs.

**Tabela 8:** Medidas de adequabilidade para os modelos

<b>Modelos</b>	<b>DIC</b>	<b>BIC</b>
M3	127.492	30.16624
M4	173.929	27.25856

Novamente notamos que os dados da Tabela 2 se adequam melhor ao modelo considerando erros correlacionados.

# Capítulo 6

## Uma Análise Bayesiana para Dados Composicionais Assumindo a Distribuição Potência Exponencial para os Erros

Consideramos nesta análise a transformação ALR para os dados composicionais da Tabela 2, isto é  $y_{i1} = \log(x_{i1}/x_{i3})$  e  $y_{i2} = \log(x_{i2}/x_{i3})$  para  $i = 1, \dots, n$ .

Assumir o modelo (2.1) e considerar os erros  $\varepsilon_{ik}$  e  $\varepsilon_{i2}$ ,  $i = 1, \dots, n$  independentes com distribuição potência exponencial (ver Box e Tiao, 1992, pg.157) dada por,

$$p(\varepsilon_{ik}/\sigma_k, \beta_k^*) = \frac{\omega(\beta_k^*)}{\sigma_k} \exp\left\{-c(\beta_k^*) \left|\frac{\varepsilon_{ik}}{\sigma_k}\right|^{\frac{2}{1+\beta_k^*}}\right\}, \quad (6.1)$$

onde  $-\infty < \varepsilon_{ik} < \infty$ ,  $c(\beta_k^*) = \left\{\frac{\Gamma[\frac{3}{2}(1+\beta_k^*)]}{\Gamma[\frac{1}{2}(1+\beta_k^*)]}\right\}^{\frac{1}{1+\beta_k^*}}$ ,  $\omega(\beta_k^*) = \frac{\{\Gamma[\frac{3}{2}(1+\beta_k^*)]\}^{\frac{1}{2}}}{(1+\beta_k^*)\{\Gamma[\frac{1}{2}(1+\beta_k^*)]\}^{\frac{3}{2}}}$ ,  $\sigma_k > 0$ ,  $-1 < \beta_k^* \leq 1$  e  $k = 1, 2$ .

Em (6.1) temos que  $E(\varepsilon_{ik}) = 0$  e que o parâmetro  $\beta_k^*$  é uma medida de curtose, indicando a existência da não-normalidade da distribuição potência exponencial. Denotamos este modelo como Modelo M5.

Observamos que a distribuição potência exponencial (6.1) é um modelo que engloba algumas distribuições comumente usadas para o erro do modelo (2.1).

## 6.1 Alguns casos especiais da distribuição potência exponencial

### 6.1.1 (1) Distribuição Normal

Se  $\beta_k^* = 0$ , temos a distribuição normal para  $\varepsilon_{ik}$  com média zero e variância  $\sigma_k^2$ ,  $k = 1, 2$ , isto é,

$$p(\varepsilon_{ik}/\sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{1}{2\sigma_k^2}\varepsilon_{ik}^2\right\}, \quad -\infty < \varepsilon_{ik} < \infty,$$

onde  $c(\beta_k^*) = 1/2$  e  $\omega(\beta_k^*) = \frac{1}{\sqrt{2\pi}}$ , que chamaremos de modelo M6.

### 6.1.2 (2) Distribuição Exponencial Dupla

Se  $\beta_k^* = 1$ , temos a distribuição exponencial dupla para  $\varepsilon_{ik}$ ,  $k = 1, 2$ , que chamaremos de modelo M7, dado por

$$p(\varepsilon_{ik}/\sigma_k) = \frac{1}{\sqrt{2}\sigma_k} \exp\left\{-\sqrt{2}\left|\frac{\varepsilon_{ik}}{\sigma_k}\right|\right\}, \quad -\infty < \varepsilon_{ik} < \infty,$$

### 6.1.3 (3) Distribuição Retangular

Se  $\beta_k^*$  tende para  $-1$ , temos uma distribuição retangular, para  $\varepsilon_{ik}$ ,  $k = 1, 2$ , denotamos por modelo M8.

$$\lim_{\beta_k^* \rightarrow -1} p(\varepsilon_{ik}/\beta_k^*, \sigma_k) = \frac{1}{2\sqrt{3}\sigma_k}, \quad -\sqrt{3}\sigma_k < \varepsilon_{ik} < \sqrt{3}\sigma_k.$$

Dessa forma temos uma grande flexibilidade de ajuste para os dados composicionais, assumindo os erros com distribuição potência exponencial (6.1).

## 6.2 Análise Bayesiana considerando a Distribuição Potência Exponencial

### 6.2.1 Função de Verossimilhança

Considerando o modelo M5, a função de verossimilhança para

$v_5 = (\alpha_1, \alpha_2, \theta_1, \theta_2, \beta_1^*, \beta_2^*, \sigma_1^2, \sigma_2^2)$  é dada por

$$\begin{aligned} L(v_5) &\propto \prod_{i=1}^n p(\varepsilon_{i1} | v_5) p(\varepsilon_{i2} | v_5) \\ &\propto \prod_{i=1}^n \frac{\omega(\beta_1^*)}{\sigma_1} \exp\left\{-c(\beta_1^*) \left|\frac{\varepsilon_{i1}}{\sigma_1}\right|^{\frac{2}{1+\beta_1^*}}\right\} \frac{\omega(\beta_2^*)}{\sigma_2} \exp\left\{-c(\beta_2^*) \left|\frac{\varepsilon_{i2}}{\sigma_2}\right|^{\frac{2}{1+\beta_2^*}}\right\}, \end{aligned} \quad (6.2)$$

onde  $\varepsilon_{i1} = y_{i1} - \alpha_1 - \theta_1 Z_i$  e  $\varepsilon_{i2} = y_{i2} - \alpha_2 - \theta_2 Z_i$

### 6.2.2 Informação *a Priori*

Assumir as seguintes distribuições *a priori* para os parâmetros:

$$\begin{aligned} \text{ii) } \theta_k &\sim N\{a_k, b_k^2\}; k = 1, 2; a_k, b_k^2 \text{ conhecidos,} \\ \text{iii) } \alpha_k &\sim N\{c_k, d_k^2\}; k = 1, 2; c_k, d_k^2 \text{ conhecidos,} \\ \text{iv) } \sigma_k^2 &\sim IG[e_k, f_k]; k = 1, 2; e_k, f_k \text{ conhecidos,} \\ \text{v) } \beta_k^* &\sim U[-1, 1]; k = 1, 2, \end{aligned} \quad (6.3)$$

onde  $N\{\mu, \sigma^2\}$  denota uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ ,  $IG[a, b]$  denota uma distribuição gama inversa com média  $b/(a-1)$  e variância  $b^2/[(a-1)^2(a-2)]$ ,  $a > 2$  e  $U[c, d]$ , denota uma distribuição uniforme no intervalo  $[c, d]$ .

### 6.2.3 *Posteriori Conjunta*

Assumindo independência *a priori* entre os parâmetros do modelo a distribuição à *posteriori* conjunta é dada por,

$$\begin{aligned} \Pi \left( v_5/\text{dados} \right) &\propto \left\{ \prod_{k=1}^2 \exp \left[ -\frac{1}{2b_k^2} (\theta_k - a_k)^2 \right] \right\} \left\{ \prod_{k=1}^2 \exp \left[ -\frac{1}{2d_k^2} (\alpha_k - d_k)^2 \right] \right\} \times \\ &\times \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(e_k+1)} e^{-f_k/\sigma_k^2} \right\} \left\{ \prod_{i=1}^n \frac{\omega(\beta_1^*)}{\sigma_1} \exp \left[ -c(\beta_1^*) \left| \frac{\varepsilon_{i1}}{\sigma_1} \right|^{\frac{2}{1+\beta_1^*}} \right] \right\} \\ &\times \left\{ \frac{\omega(\beta_2^*)}{\sigma_2} \exp \left[ -c(\beta_2^*) \left| \frac{\varepsilon_{i2}}{\sigma_2} \right|^{\frac{2}{1+\beta_2^*}} \right] \right\}. \end{aligned} \quad (6.4)$$

Para obtermos amostras da distribuição *a posteriori* conjunta (6.4) podemos usar o método MCCM (Monte Carlo em Cadeias de Markov), como o algoritmo de Gibbs sampling. Para isso, precisamos determinar as distribuições condicionais de todos os parâmetros.

Uma simplificação possível é dada quando usamos o *software* Winbugs (Spiegelhalter *et al.*, 1999) onde especificamos apenas a verossimilhança e as distribuições *a priori* dos parâmetros do modelo, não sendo necessário especificar as distribuições condicionais.

Observe que se  $\beta_k^* = 0$  (distribuição normal para os erros) as distribuições condicionais para  $\theta_k$ ,  $\alpha_k$  e  $\sigma_k^2$  necessárias para o algoritmo de Gibbs sampling são dadas em (4.16).

## 6.3 Análise dos Dados do Oceano Ártico

Para análise dos dados introduzidos na Tabela 2, consideramos a transformação ALR com distribuição para os erros potência exponencial, dada em (6.1). Na geração das amostras, utilizamos o *software* Winbugs (ver Spiegelhalter, 1999). O código utilizado do modelo M5 está descrito no apêndice A. A convergência do algoritmo foi verificada usando os critérios de Gelman e Rubin (1992) implementados no *software* Winbugs. Para todos modelos propostos geramos duas cadeias cada uma com 55000 iterações partindo de pontos iniciais diferentes e descartamos as primeiras 5000 iterações (“burn-in”). Além disso, consideramos um salto de tamanho 10, para eliminar a correlação entre as amostras.

### 6.3.1 Modelo M5 (Distribuição Potência Exponencial)

Considerando a verossimilhança (6.2) e assumindo as distribuições *a priori* (6.3) com valores dos hiperparâmetros dadas por  $a_1 = a_2 = c_1 = c_2 = 0$ ,  $b_1^2 = b_2^2 = d_1^2 = d_2^2 = 100000$ ,  $e_1 = e_2 = f_1 = f_2 = 1000$  temos na Tabela 9 os sumários *a posteriori* e os valores do fator de redução de escala  $\hat{R}$  (critério Gelman e Rubin). Para todos os parâmetros observamos a convergência do algoritmo Gibbs *sampling*, pois  $\sqrt{\hat{R}} < 1.1$  para todos os casos.

Na Tabela 9 observamos que a profundidade apresenta um efeito significativo para as proporções de componentes na mistura (zero não esta incluído nos intervalos de credibilidade de  $\theta_1$  e  $\theta_2$ ).

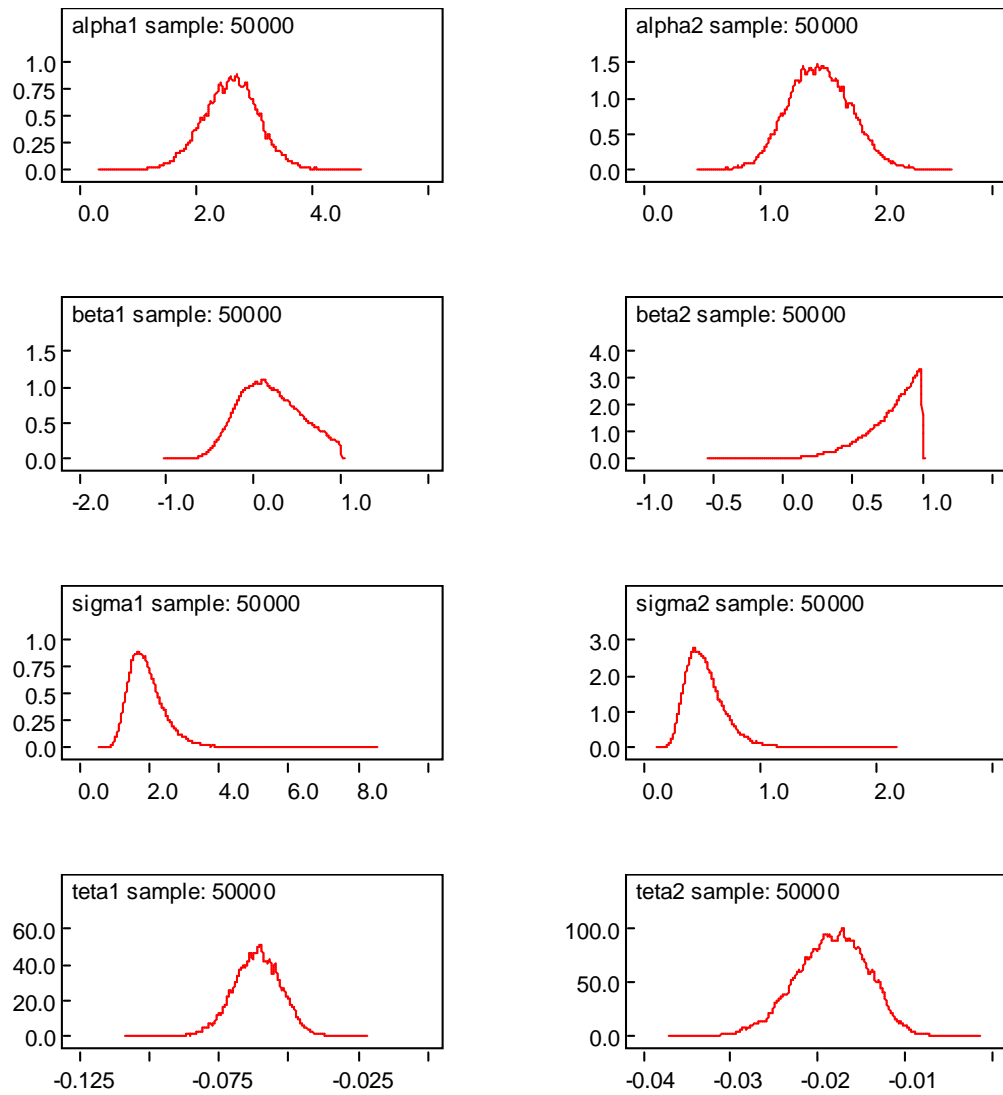
**Tabela 9:** Sumários *a posteriori* para o Modelo M5

Parâmetros	Média	Int.Cred 95%	$\hat{R}$
$\alpha_1$	2.611	(1.661; 3.612)	1.001
$\alpha_2$	1.511	(0.990; 2.077)	1.003
$\theta_1$	-0.061	(-0.078; -0.046)	1.001
$\theta_2$	-0.018	(-0.027; -0.0105)	1.002
$\sigma_1^2$	1.809	(1.706; 3.249)	1.003
$\sigma_2^2$	0.487	(0.284; 0.952)	1.000
$\beta_1^*$	0.192	(-0.418; 0.900)	1.010
$\beta_2^*$	0.769	(0.293; 0.993)	1.002

Também observarmos que  $\beta_1^*$  poderia ser igual a zero ( pois o zero esta incluído em seu intervalo de credibilidade)

Na Figura 14, temos os gráficos das densidades *a posteriori* para todos os parâmetros.





**Figura 14:** gráficos das densidades geradas pelo amostrador de Gibbs

### 6.3.2 Modelo M6 (Distribuição Normal)

Considerando o modelo em (4.1) (erros normais) e assumindo as distribuições *a priori* (6.2) com valores dos hiperparâmetros dados por  $a_1 = a_2 = c_1 = c_2 = 0$ ,  $b_1^2 = b_2^2 = d_1^2 = d_2^2 = 100000$ ,  $e_1 = e_2 = f_1 = f_2 = 1000$ , os valores para os parâmetros de interesse são os mesmos dados na Tabela 4.

### 6.3.3 Modelo M7 (Distribuição Exponencial dupla)

Considerando o modelo em (4.1) e assumindo as distribuições *a priori* (6.2) com valores dos hiperparâmetros  $a_1 = a_2 = c_1 = c_2 = 0$ ,  $b_1^2 = b_2^2 = d_1^2 = d_2^2 = 100000$ ,

$e_1 = e_2 = f_1 = f_2 = 1000$ . A Tabela 10 mostra os sumários *a posteriori* para os parâmetros do Modelo M7, onde observamos o efeito significativo da profundidade.

**Tabela 10:** Sumários *a posteriori* para Modelo M7

Parâmetros	Média	Int.Cred 95%	$\hat{R}$
$\alpha_1$	2.337	(1.2; 3.324)	1.001
$\alpha_2$	1.392	(0.905; 1.924)	1.002
$\theta_1$	-0.056	(-0.073; -0.040)	1.000
$\theta_2$	-0.017	(-0.0245; -0.009)	1.003
$\sigma_1^2$	2.398	(1.222; 4.423)	1.003
$\sigma_2^2$	0.542	(0.278; 0.998)	1.000

### 6.3.4 Modelo M8 (Distribuição Retangular)

Considerando agora o modelo (4.1) e assumindo as distribuições *a priori* (6.2) com valores dos hiperparâmetros  $a_1 = a_2 = c_1 = c_2 = 0$ ,  $b_1^2 = b_2^2 = d_1^2 = d_2^2 = 100000$ ,  $e_1 = e_2 = f_1 = f_2 = 1000$  e  $\beta^* = -0.75$ , temos na Tabela 11 os sumários *a posteriori* para os parâmetros do Modelo M8.

**Tabela 11:** Sumários *a posteriori* para Modelo M8

Parâmetros	Média	Int.Cred 95%	$\hat{R}$
$\alpha_1$	3.503	(2.506; 4.301)	1.001
$\alpha_2$	2.798	(2.255; 3.258)	1.002
$\theta_1$	-0.077	(-0.095; -0.048)	1.000
$\theta_2$	-0.034	(-0.0471; -0.016)	1.003
$\sigma_1^2$	2.712	(2.137; 3.588)	1.003
$\sigma_2^2$	0.973	(0.771; 1.284)	1.000

Da mesma forma que para os outros modelos, observamos um efeito significativo da profundidade.

Na Tabela 12, temos as estimativas Monte Carlo para o DIC e BIC baseados nas amostras de Gibbs para os modelos M5 (distribuição potência exponencial); M6 (distribuição normal); M7 (distribuição exponencial dupla); M8 (distribuição retangular).

**Tabela 12:** Medidas de adequabilidade para os modelos

<b>Modelos</b>	<b>DIC</b>	<b>BIC</b>
M5	221.262	25.9976
M6	228.988	24.6442
M7	222.151	22.4264
M8	250.802	24.19796

Dos resultados da Tabela 12 concluímos que pelo critério DIC (menor valor) e o critério BIC (maior valor) o modelo M5 é melhor ajustado pelos dados da Tabela 2.

# Capítulo 7

## Considerações finais

Dados Composicionais se apresentam na forma de porcentagem e podem aparecer em várias áreas de aplicação. Frequentemente, a restrição de soma igual a um, é ignorada ou incorporada de forma inapropriada na modelagem estatística, obtendo resultados imprecisos ou inadequados.

A partir de uma aplicação com dados reais, utilizamos as transformações ALR e a Box-Cox, considerando a distribuição normal para erros correlacionados e não correlacionados, observando que os modelos com erros correlacionados melhor se ajustam aos dados reais observados na Tabela 2, e que a transformação Box-Cox é a mais adequada aos dados.

Utilizamos uma generalização da distribuição normal; a distribuição potência exponencial, considerando a transformação ALR vimos que considerando erros não correlacionados, esta generalização nos leva a melhores resultados.

O uso de métodos MCMC é uma ótima alternativa para o desenvolvimento da análise Bayesiana para dados composicionais. O uso do *software* Winbugs garante uma grande simplificação das amostras das distribuições *a posteriori* de interesse para dados composicionais.

Estes resultados podem ser estendidos facilmente para outras aplicações com dados composicionais considerando mais que três componentes

# Referências Bibliográficas

- [1] Aitchison, J., (1982). The Statistical Analysis of Compositional Data. *Journal of Royal Statistical Society, B, 2: 139-177.*
- [2] Aitchison, J., Shen, S. M., (1985). Logistic-normal distributions: Some properties and uses. *Biometrika, 47:136-146.*
- [3] Aitchison, J., (1986). The Statistical Analysis of Compositional Data. *Chapman and Hall.*
- [4] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. Information Theory (eds B. N. Petrov and F. Csáki)* 267–281
- [5] Box, G.E.P., Cox, D.R., (1964). An Analysis of Transformations . *Journal of the Royal Statistical Society, B, 26: 211-52.*
- [6] Box, G.E.P., Tiao, G.C., (1992). Bayesian Inference in Statistical Analysis . *Reading: Addison-Wesley.*
- [7] Carlin, B.P., Louis, T., (2000). Bayes and Empirical Bayes methods for data analysis, *2<sup>nd</sup> ed, London: Chapman and Hall.*
- [8] Coakley, J. P., Rust, B. R., (1968). Sedimentation in an Arctic lake. *Journal of Sedimentary Petrology, 38:1290-1300.*
- [9] Gelfand, A.E., Smith, A.F.M., (1990). sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85, 398-409.*

- [10] Gelman, S., Gelman, D., (1984). Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- [11] Gelman, A., Rubin, D.B., (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.
- [12] Iyengar, M., Dey, D.K., (1996a). Bayesian Analysis of Compositional Data. *Department of Statistics, University of Connecticut, Storrs, CT 06269-3120*
- [13] Iyengar, M., Dey, D.K., (1996b). Box-Cox Transformations in Bayesian Analysis of Compositional Data. *Department of Statistics, University of Connecticut*.
- [14] Johnson, R., Wichern, D., (1998). Applied multivariate statistical analysis. *New Jersey : Prentice Hall*.
- [15] McAlister, D., (1879). The law of the geometric mean. *Proc. Roy Soc.*, 29, 367-375.
- [16] Paulino, C. D.; Turkman, M. A. A.; Murteira, B. (2003). *Estatística Bayesian, Fundação Calouste Gulbenkian, Lisboa*.
- [17] Rayens, W.S., Srinivasan, C., (1991a). Box-Cox transformations in the analysis of compositional data. *Journal of Chemometrics*, 5:227-239.
- [18] Rayens, W.S., Srinivasan, C., (1991b). Estimation in compositional data. *Journal of Chemometrics*, 5:361-374.
- [19] Roberts, G.O., Smith, A.F.M., (1993). Bayesian methods via the Gibbs Sampler and related Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society, B, Cambridge, v.55,n.1, 3-23*.
- [20] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-466.
- [21] Spiegelhalter, D.J, Thomas, A, Best, N.G., (1999). Winbugs: Bayesian inference using Gibbs sampling. *Cambridge: MRC Biostatistics Unit*.

- [22] Spiegelhalter, D.J, Best, N.G., Vander Linde, A., (2002). A Bayesian Measures of model complexity and fit (with discussion) *J. Royal Statist Soc, B*, 64 583-639.

Apêndice

# Apêndice A

## Programas

Seguem as linhas de instruções em WinBugs versão 1.4.1 para os procedimentos MCMC usados nas análises Bayesianas apresentadas nesta dissertação.

Código do programa utilizado no *software* Winbugs, a fim de obtermos a distribuição condicional dos parâmetros de interesse, sob o modelo Modelo M1

```
Model
{
  for( i in 1 : N ) {
    Y[i] <- ( ( sqrt(tau1*tau2) ) / ( ( 2 * 3.1415926 ) * sqrt(1-pow(ro,2) ) ) ) * exp( - ( ( 0.5 * tau1 * pow(box1[i],2) *
pow(ro,2) ) - ( ( 0.5 * tau2 * pow(box2[i],2) ) / ( 1-pow(ro,2) ) ) + ( ( sqrt(tau1*tau2) * ro * box1[i] * box2[i] ) / ( 1-
pow(ro,2) ) ) ) ) ) / const
  }
  for( i in 1 : N ) {
    mu1[i] <- alpha1 + teta1 * x[i]
    cox1[i] <- ( log(y1[i]) )
    box1[i] <- cox1[i] - mu1[i]
    mu2[i] <- alpha2 + teta2 * x[i]
    cox2[i] <- ( log(y2[i]) )
    box2[i] <- cox2[i] - mu2[i]
  }
  alpha1 ~ dnorm( 0.0, 1.0E-6)
```





Código do programa utilizado no *software* Winbugs, a fim de obtermos a distribuição condicional dos parâmetros de interesse, sob o modelo Modelo M2

```

Model
{
  for( i in 1 : N ) {
    Y[i] <- ( ( sqrt(tau1/(2 * 3.1415926) ) ) * ( sqrt(tau2/(2 * 3.1415926) ) ) *exp( - 0.5*
tau1*pow(box1[i],2) - 0.5* tau2*pow(box2[i],2))) / const
  }
  for( i in 1 : N ) {
    mu1[i] <- alpha1 + teta1 * x[i]
    cox1[i]<-(log(y1[i]))
    box1[i]<-cox1[i]-mu1[i]
    mu2[i] <- alpha2 + teta2 * x[i]
    cox2[i]<-(log(y2[i]))
    box2[i]<-cox2[i]-mu2[i]
  }
  alpha1~dnorm( 0.0,1.0E-6)
  teta1~dnorm( 0.0,1.0E-6)
  tau1 ~dgamma(0.001,0.001)
  alpha2 ~dnorm( 0.0,1.0E-6)
  teta2 ~dnorm( 0.0,1.0E-6)
  tau2 ~dgamma(0.001,0.001)
  sigma1<-1/tau1
  sigma2<-1/tau2
  for( i in 1 : N ) {
    Uns[i] ~dbern(Y[i])
  }
}

Dados

list(const=1,N=39,y1=c(25.8333, 22.46875, 3.8409, 7.9090909, 20, 51.153846, 26.9375,
5.4489796, 0.5149501, 1.1828358, 10.107692, 117.33333, 0.6, 0.5408163, 2.0427807, 0.2958904,

```

```

0.5954692, 0.0958333, 2.1794872, 1.3869565, 0.2567567, 0.4899713, 0.3079178, 0.1170731,
0.0498084, 0.3175487, 0.1443965, 0.1589861, 0.0782778, 0.1809291, 0.1050328, 0.0957446,
0.1598063, 0.1459695, 0.1577825, 0.1330376, 0.1578947,0.050505,0.0398406),
  y2=c(6.5000,7.7812,2.7348,6.1970,7.5714,24.7692,34.5625,3.7551,1.8073,
1.5485,4.2769,48.3333,1.8483,3.5612,2.3048,1.4438,1.6408,0.9875,1.4824,1.9609,1.4459,
  1.3754,1.6246,1.3341,0.8659,1.4680,1.0108,1.1452,0.8787,1.2616,1.0832,1.0319,1.2615,
  1.0305,0.9723,1.0843,1.3484,0.9697,0.9522),x=c(10.4, 11.7, 12.8, 13.0, 15.7, 16.3, 18.0,
18.7, 20.7, 22.1, 22.4, 24.4, 25.8, 32.5, 33.6, 36.8, 37.8, 36.9, 42.2, 47.0, 47.1, 48.4, 49.4,
49.5, 59.2, 60.1, 61.7, 62.4, 69.3, 73.6, 74.4, 78.5, 82.9, 87.7, 88.1, 90.4, 90.6, 97.7, 103.7),
Uns=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))

```

Valores Iniciais

```
list(alpha1=0.5,teta1=-0.01,tau1=0.5, alpha2=0.5,teta2=-0.01,tau2=0.5)
```

```
list(alpha1=1,teta1=1,tau1=1, alpha2=1,teta2=1,tau2=0.2)
```

Código do programa utilizado no *software* Winbugs, a fim de obtermos a distribuição condicional dos parâmetros de interesse, sob o modelo potência exponencial (Modelo M5)

```

Model
{
for( i in 1 : N ) {
L1[i] <- wbeta1*sqrt(tau1)*exp(-cbeta1*f1[i])
L2[i] <- wbeta2*sqrt(tau2)*exp(-cbeta2*f2[i])
L[i] <- L1[i]*L2[i]/const
}
C1 <- 1+beta1
A1 <-exp(loggam(1.5*C1))
B1 <-exp(loggam(0.5*C1))
cbeta1 <- pow(A1/B1, 1/C1)
wbeta1 <- pow(A1, 0.5)/(C1*pow(B1, 1.5))
C2 <- 1+beta2
A2 <-exp(loggam(1.5*C2))
B2 <-exp(loggam(0.5*C2))
cbeta2 <- pow(A2/B2, 1/C2)
wbeta2 <- pow(A2, 0.5)/(C2*pow(B2, 1.5))
for( i in 1 : N ) {
mu1[i]<- alpha1 + teta1 * x[i]
yt1[i]<- log(y1[i])
e1[i]<- yt1[i] - mu1[i]
d1[i]<- abs(e1[i]*sqrt(tau1))
f1[i]<- pow(d1[i], 2/C1)
mu2[i]<- alpha2 + teta2 * x[i]
yt2[i]<- log(y2[i])
e2[i]<- yt2[i] - mu2[i]
d2[i]<- abs(e2[i]*sqrt(tau2))
f2[i]<- pow(d2[i], 2/C2)
}
}

```

```
alpha1 ~ dnorm(0,0.0001)
teta1 ~ dnorm(0, 0.0001)
tau1 ~ dgamma(0.001,0.001)
beta1 ~ dunif(-0.99,1)
sigma1 <- 1/tau1
alpha2 ~ dnorm(0, 0.0001)
teta2 ~ dnorm(0, 0.0001)
tau2 ~ dgamma(0.001,0.001)
beta2 ~ dunif(-0.99,1)
sigma2 <- 1/tau2
for( i in 1 : N ) {
  Uns[i] ~ dbern(L[i])
}
}
Dados
list(const=1, N=39,y1=c(25.8333, 22.46875, 3.8409, 7.9090909, 20, 51.153846, 26.9375,
5.4489796, 0.5149501, 1.1828358, 10.107692, 117.33333, 0.6, 0.5408163, 2.0427807, 0.2958904,
0.5954692, 0.0958333, 2.1794872, 1.3869565, 0.2567567, 0.4899713, 0.3079178, 0.1170731,
0.0498084, 0.3175487, 0.1443965, 0.1589861, 0.0782778, 0.1809291, 0.1050328, 0.0957446,
0.1598063, 0.1459695, 0.1577825, 0.1330376, 0.1578947, 0.050505, 0.0398406),
  y2=c(6.5000,7.7812,2.7348,6.1970,7.5714,24.7692,34.5625,3.7551,1.8073,1.5485,
4.2769,48.3333,1.8483,3.5612,2.3048,1.4438,1.6408,0.9875,1.4824,1.9609,1.4459,
1.3754,1.6246,1.3341,0.8659,1.4680,1.0108,1.1452,0.8787,1.2616,1.0832,1.0319,
1.2615,1.0305,0.9723,1.0843,1.3484,0.9697,0.9522),
  x=c(10.4, 11.7, 12.8, 13.0, 15.7, 16.3, 18.0, 18.7, 20.7, 22.1, 22.4, 24.4, 25.8, 32.5, 33.6,
36.8, 37.8, 36.9, 42.2, 47.0, 47.1, 48.4, 49.4, 49.5, 59.2, 60.1, 61.7, 62.4, 69.3, 73.6, 74.4,
78.5, 82.9, 87.7, 88.1, 90.4, 90.6, 97.7, 103.7),
  Uns=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))
Valores Iniciais
list(alpha1=0.5, teta1=0,tau1=1, alpha2=0.5, teta2=0, tau2=1, beta1=0, beta2=0)
```