

# MODELOS DE REGRESSÃO NÃO-LINEARES CONSIDERANDO UMA MISTURA FINITA DE DISTRIBUIÇÕES NORMAIS PARA OS ERROS

Vicente Garibay CANCHO<sup>1</sup>  
Jorge Alberto ACHCAR<sup>2</sup>  
Edwin Moises Marcos ORTEGA<sup>3</sup>

- RESUMO: Neste artigo exploramos o uso de técnicas de Monte Carlo via Cadeias de Markov (MCMC) para desenvolver uma análise bayesiana do modelo de regressão não-linear (MRNL) considerando que o termo aleatório do modelo tem uma mistura finita de distribuições normais. Utilizamos o algoritmo de Metropolis-Hasting para gerar amostras da distribuição conjunta *a posteriori* dos parâmetros do modelo. Ilustramos a metodologia com dois exemplos considerando dados reais e dados simulados.
- PALAVRAS-CHAVE: Regressão não-linear; seleção de modelos; MCMC; mistura de normais.

## 1 Introdução

Seja a variável resposta  $Y_i$  e as variáveis independentes  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^\top$  relacionadas pelo seguinte modelo aditivo:

$$Y_i = \eta(\mathbf{x}_i; \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

em que  $\eta(\mathbf{x}_i; \boldsymbol{\theta})$  é uma função não-linear contínua e diferenciável em  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$  e  $\varepsilon_i$  é o erro aleatório associado ao modelo. Tradicionalmente são desenvolvidos procedimentos inferenciais para o modelo dado em (1) supondo

---

<sup>1</sup>Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação – ICMC, Universidade de São Paulo – USP, CEP:13560-970, São Carlos, SP, Brasil. E-mail: garibay@icmc.usp.br.

<sup>2</sup>Departamento de Estatística, Universidade Federal de São Carlos - UFSCar, Caixa Postal 676, CEP 13565-905, São Carlos, SP, Brasil. E-mail: jachcar@power.ufscar.com.br.

<sup>3</sup>Departamento Ciências Exatas, Universidade de São Paulo, campus de Piracicaba – ESALQ/USP, Caixa Postal 9, CEP 13418-900, Piracicaba, SP, Brasil. E-mail: edwin@ciagri.usp.br.

que os termos aleatórios são variáveis aleatórias independentes e identicamente distribuídas com distribuição normal com média zero e variância  $\sigma^2$  (desconhecida). Se a suposição de normalidade dos erros no modelo de regressão não-linear não é satisfeita, então teremos o problema de falta de robustez da suposição distribucional, que é comumente encontrado nos modelos estatísticos baseados na distribuição normal. Na prática, a aproximação mais comumente adotada para alcançar a normalidade envolve transformação de variáveis. Embora tais métodos possam fornecer resultados empíricos razoáveis deveriam ser evitados considerando uma família paramétrica mais flexível. Como proposta de flexibilização, consideramos que  $\varepsilon_i$  segue uma mistura finita de  $J$  distribuições normais com média zero, isto é,

$$f(\varepsilon_i) = \sum_{j=1}^J p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2) \quad (2)$$

onde  $p_j > 0$ ,  $\sum_{j=1}^J p_j = 1$ ,  $E(\varepsilon_i) = \sum_{j=1}^J p_j \mu_j = 0$  (restrição do modelo) e  $\phi(\varepsilon_i | \mu_j, \sigma_j^2)$  é a função de densidade de uma distribuição normal com média  $-\infty < \mu_j < \infty$  e variância  $\sigma_j^2 > 0$ ,  $j = 1, 2, \dots, J$ . Assim, das expressões (1) e (2) podemos observar que a distribuição de  $Y_i | \boldsymbol{\theta}$  é uma mistura finita de distribuições normais.

Procedimentos inferenciais clássicos para os modelos de mistura baseado no estimador de máxima verossimilhança pode ser dificultoso, mesmo para os casos mais simples como a mistura de duas distribuições normais (ver por exemplo, Titterton et al, 1985). Neste trabalho desenvolvemos um procedimento bayesiano baseado em métodos de Monte Carlo via Cadeias de Markov (MCMC) para os MRNL considerando que a distribuição da componente aleatória segue uma mistura de distribuições normais (ver por exemplo, Roeder e Wasserman, 1997, ou Diebolt e Robert, 1994).

A apresentação do artigo segue os seguintes passos: na seção 2, introduzimos uma análise Bayesiana assumindo variáveis latentes para a simplificação das distribuições *a posteriori* condicionais necessárias para o amostrador de Gibbs; na subseção 2.2, introduzimos algumas técnicas Bayesiana de seleção de modelos; na seção 3, introduzimos dois exemplos de aplicação assumindo que os erros têm misturas de duas distribuições normais; finalmente, na seção 4, apresentamos algumas conclusões.

## 2 Análise Bayesiana

Sejam  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$   $n$  pares de observações em que,  $y_i$  é a variável resposta que se relaciona com as variáveis independentes  $\mathbf{x}_i$  pelo modelo dado em (1). Considerando que os termos aleatórios  $\varepsilon_i$  seguem uma distribuição com f.d.p dada em (2), a função de verossimilhança do modelo proposto é dada por

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^n \left( \sum_{j=1}^J p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2) \right) \quad (3)$$

em que o vetor de parâmetros é  $\gamma = (\boldsymbol{\theta}, \mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2, p_1, \dots, p_j)$  e  $\varepsilon_i = y_i - \eta(\mathbf{x}_i; \boldsymbol{\theta})$ .

Por outro lado, assumimos que a densidade *a priori* conjunta dos parâmetros do modelo proposto seja  $\pi(\gamma)$ . Assim, a densidade *a posteriori* conjunta de  $\gamma$  é dada por

$$\pi(\gamma|\mathbf{y}, \mathbf{x}) \propto \prod_{i=1}^n \left( \sum_{j=1}^J p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2) \right) \pi(\gamma). \quad (4)$$

Observe em (4) que obter as distribuições marginais *a posteriori* de  $\gamma$  é bastante complexo para qualquer que seja  $\pi(\gamma)$ . Uma alternativa para situações desse tipo é o uso do algoritmo de Gibbs. Com a finalidade de simplificar as distribuições condicionais necessárias para o algoritmo de Gibbs, recorreremos a uma variável latente (Tanner e Wong, 1987) que permite escrever a função de verossimilhança como produto das componentes do modelo de observações independentes.

Consideramos inicialmente o caso particular quando  $J = 2$ , que corresponde ao modelo de mistura de duas distribuições normais. Seja  $\mathbf{v}_i = (v_{i1}, v_{i2})$ ,  $i = 1, \dots, n$ , a variável latente, onde  $v_{i1} + v_{i2} = 1$  e  $v_{i1} | \gamma, y_i, \mathbf{x}_i \sim b(1, h_{i1})$  (uma distribuição de Bernoulli) com probabilidade de sucesso  $h_{i1}$  dada por

$$h_{i1} = \frac{p_1 \phi(\varepsilon_i | \mu_1, \sigma_1^2)}{\sum_{j=1}^2 p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2)}. \quad (5)$$

Assim, a distribuição conjunta de  $\mathbf{v}_1, \dots, \mathbf{v}_n$  é dada por

$$f(\mathbf{v}_1, \dots, \mathbf{v}_n) \propto \frac{\prod_{i=1}^n \prod_{j=1}^2 [p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2)]^{v_{ij}}}{\prod_{i=1}^n \left\{ \sum_{j=1}^2 p_j \phi(\varepsilon_i | \mu_j, \sigma_j^2) \right\}}. \quad (6)$$

Combinando (6) e (4), temos a função de densidade conjunta *a posteriori* de  $\gamma$  e  $\mathbf{v}$  dado  $\mathbf{y}, \mathbf{x}$ :

$$\pi(\gamma, \mathbf{v} | \mathbf{y}, \mathbf{x}) \propto \prod_{i=1}^n [p_1 \phi(\varepsilon_i | \mu_1, \sigma_1^2)]^{v_{i1}} [p_2 \phi(\varepsilon_i | \mu_2, \sigma_2^2)]^{v_{i2}} \pi(\gamma). \quad (7)$$

Similar resultado pode ser obtido quando  $J > 2$ .

O amostrador de Gibbs em cada ciclo é composto por duas etapas. Dados valores iniciais

$$\boldsymbol{\gamma}^{(0)} = (\gamma_1^{(0)}, \dots, \gamma_q^{(0)}),$$

- (i) Gerar amostras  $\mathbf{v}^{(1)} = (\mathbf{v}_1^{(1)}, \dots, \mathbf{v}_n^{(1)})$ , onde  $\mathbf{v}_i^{(1)} = (v_{i1}^{(1)}, v_{i2}^{(1)})$  da distribuição Bernoulli  $b(1, h_{i1})$ , onde  $h_{i1}$  é dado por (5);

(ii) Gerar uma amostra de  $\gamma$  das distribuições condicionais

$$\pi(\gamma_1 | \gamma_{-\gamma_1}^{(0)}, \mathbf{v}^{(1)}, \mathbf{y}, \mathbf{x}), \dots, \pi(\gamma_J | \gamma_{-\gamma_J}^{(1)}, \mathbf{v}^{(1)}, \mathbf{y}, \mathbf{x})$$

$$\text{onde } \gamma_{-\gamma_i} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_J).$$

## 2.1 Distribuições condicionais

Para o modelo regressão não-linear (1) com mistura de duas distribuições normais para o termo aleatório e as restrições da distribuição do erro, isto é, com  $p_1 + p_2 = 1$  e  $p_1\mu_1 + p_2\mu_2 = 0$ , se  $p_1 = p$ ,  $p_2 = 1 - p$  e  $\mu_2 = -\frac{p\mu_1}{1-p}$ , pode-se demonstrar que a densidade conjunta *a posteriori* (7), de  $\gamma = (\boldsymbol{\theta}, p, \mu_1, \sigma_1, \sigma_2)$  e  $\mathbf{v}$  dados  $\mathbf{y}, \mathbf{x}$  é dada por,

$$\pi(\gamma, \mathbf{v} | \mathbf{y}, \mathbf{x}) \propto p^r (1-p)^{n-r} \prod_{i=1}^n [\phi(\varepsilon_i | \mu_1, \sigma_1^2)]^{v_{i1}} \quad (8)$$

$$[\phi(\varepsilon_i | -p\mu_1/(1-p), \sigma_2^2)]^{1-v_{i1}} \pi(\gamma)$$

$$\text{onde } r = \sum_{i=1}^4 v_{i1} \text{ e } \varepsilon_i = y_i - \eta(\mathbf{x}_i; \boldsymbol{\theta}).$$

Assumimos as seguintes densidades *a priori* independentes para  $\gamma = (\boldsymbol{\theta}, p, \mu, \sigma_1, \sigma_2)$ :

$$\begin{aligned} p &\sim \text{Beta}(a, b); \quad a \text{ e } b \text{ conhecidos,} \\ \mu_1 &\sim N(c, d); \quad c \text{ e } d \text{ conhecidos,} \\ \sigma_j^2 &\sim GI(e_j, f_j); \quad e_j \text{ e } f_j \text{ conhecidos } j = 1, 2, \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), \end{aligned} \quad (9)$$

onde  $\text{Beta}(a, b)$  denota a distribuição beta com parâmetros  $a$  e  $b$ ,  $N(c, d)$  é a distribuição normal com média  $c$  e variância  $d$ ,  $GI(e, f)$  denota a distribuição gama inversa com parâmetros  $e$  e  $f$  e  $\pi(\boldsymbol{\theta})$  é a distribuição *a priori* conjunta de  $\boldsymbol{\theta}$ .

De (8), pode-se mostrar que as distribuições condicionais completas requeridas pelo amostrador de Gibbs são dadas por

- (i)  $v_{i1}|\boldsymbol{\theta}, \mu_1, p, \sigma_1^2, \sigma_2^2, \mathbf{y}, \mathbf{x} \sim b(1, h_{i1}), i = 1, \dots, n,$  onde
- $$h_{i1} = \frac{p\phi(\varepsilon_i|\mu_1, \sigma_1^2)}{p\phi(\varepsilon_i|\mu_1, \sigma_1^2) + (1-p)\phi(\varepsilon_i - \frac{p\mu_1}{1-p}, \sigma_2^2)};$$
- (ii)  $\boldsymbol{\theta}|\mu_1, p, \sigma_1^2, \sigma_2^2, \mathbf{y}, \mathbf{x}, \mathbf{v} \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^2\frac{v_{ij}(\varepsilon_i - \mu_j)^2}{\sigma_j^2}\right\}\pi(\boldsymbol{\theta}),$
- (iii)  $\mu_1|\boldsymbol{\theta}, p, \sigma_1^2, \sigma_2^2, \mathbf{y}, \mathbf{x}, \mathbf{v} \sim N(m, \kappa),$  onde
- $$m = \frac{d(k_1(1-p)^2\sigma_2^2 + k_2\sigma_1^2p(1-p)) + c(1-p)\sigma_1^2\sigma_2^2}{d(k_1(1-p)^2\sigma_2^2 + (n-r)p^2\sigma_1^2) + (1-p)^2\sigma_1^2\sigma_2^2}, k_1 = \sum_{i=1}^n v_{i1}\varepsilon_i,$$
- $$k_2 = \sum_{i=1}^n \varepsilon_i(1-v_{i1}) \text{ e } \kappa = \frac{d(1-p)^2\sigma_1^2\sigma_2^2}{d(k_1(1-p)^2\sigma_2^2 + (n-r)p^2\sigma_1^2) + (1-p)^2\sigma_1^2\sigma_2^2}$$
- (iv)  $\sigma_1^2|\boldsymbol{\theta}, \mu_1, p, \sigma_2^2, \mathbf{y}, \mathbf{x}, \mathbf{v} \sim GI\left(\frac{r}{2} + e_1, f_1 + \frac{1}{2}\sum_{i=1}^n v_{i1}(\varepsilon_i - \mu_1)^2\right),$
- (v)  $\sigma_2^2|\boldsymbol{\theta}, \mu_1, p, \sigma_1^2, \mathbf{y}, \mathbf{x}, \mathbf{v} \sim GI\left(\frac{n-r}{2} + e_2, g\right),$  onde
- $$g = f_2 + \frac{1}{2}\sum_{i=1}^n (1-v_{i1})\left(\varepsilon_i + \frac{p\mu_1}{1-p}\right)^2,$$
- (vi)  $p|\boldsymbol{\theta}, \mu_1, \sigma_1^2, \sigma_2^2, \mathbf{y}, \mathbf{x}, \mathbf{v} \propto p^{r+a-1}(1-p)^{n-r+b-1} \times$
- $$\exp\left\{-\frac{1}{2\sigma_2^2}\sum_{i=1}^n v_{i2}\left(\varepsilon_i + \frac{p\mu_1}{1-p}\right)^2\right\}.$$

Assim, uma vez atualizados  $\mu_1$  e  $p$ , podemos determinar  $\mu_2$  por  $\mu_2 = -p\mu_1/(1-p)$ .

## 2.2 Seleção de modelos

A determinação do modelo é um problema fundamental da estatística. Sob o enfoque bayesiano, a literatura sobre a adequação e seleção de modelos apresenta muitas abordagens, sendo a mais popular o fator de Bayes. Várias modificações do fator de Bayes são apresentadas na literatura (veja, por exemplo, Aitikin, 1991; Berger e Perichi, 1996). Geisser e Eddy (1979), Gelfand e Gosh (1998) sugerem uma abordagem preditiva para a seleção de modelos.

Considere uma escolha entre dois modelos paramétricos, cada um sendo denotado por sua densidade conjunta  $f(\mathbf{y}; \boldsymbol{\theta}_i, M_i)$  ou por sua função de verossimilhança  $L(\boldsymbol{\theta}_i; \mathbf{y}, M_i)$ ,  $i = 1, 2$ . Suponha que  $w_i$  é a probabilidade de selecionar o modelo  $M_i$ ,  $i = 1, 2$ , e  $f(\mathbf{y}|M_i)$  é a distribuição preditiva para o modelo  $M_i$ , isto é,

$$f(\mathbf{y}|M_i) = \int f(\mathbf{y}|\boldsymbol{\theta}_i, M_i)\pi(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i \quad (10)$$

onde  $\pi(\boldsymbol{\theta}_i|M_i)$  é a função de densidade *a priori* sob o modelo  $M_i$ . Se  $\mathbf{y}_0$  denota o conjunto de dados observados, então escolhemos o modelo que tem maior valor  $w_i f(\mathbf{y}_0|M_i)$ . Frequentemente considera-se  $w_i = 0,5$ ,  $i = 1, 2$  e o fator de Bayes de  $M_1$  com relação  $M_2$  é definido como

$$B_{12} = \frac{f(\mathbf{y}_0|M_1)}{f(\mathbf{y}_0|M_2)}. \quad (11)$$

A distribuição preditiva (10) pode ser aproximada por Monte Carlo usando  $S$  amostras geradas da densidade *a priori*  $\pi(\boldsymbol{\theta}_i|M_i)$ , isto é,

$$\hat{f}(\mathbf{y}|M_i) = \frac{1}{S} \sum_{s=1}^S f(\mathbf{y}|\boldsymbol{\theta}_i^{(s)}, M_i) \quad (12)$$

onde  $\boldsymbol{\theta}_i^{(s)}$  é a  $s$ -ésima amostra do vetor  $\boldsymbol{\theta}_i$ . Algumas modificações da estimativa (12) para a densidade preditiva são propostas na literatura (veja, por exemplo, Newton e Raftery, 1994; Gelfand e Dey, 1994). Para a seleção de modelos podemos utilizar a técnica da densidade preditiva condicional ordenada (CPO) (ver, Gelfand e Dey, 1994), dada por

$$CPO_r = f(y_r|\mathbf{y}, M_i) = \int f(y_r|\boldsymbol{\theta}_i, \mathbf{y}_{(r)}, M_i)\pi(\boldsymbol{\theta}_i|\mathbf{y}_{(r)}, M_i)d\boldsymbol{\theta}_i \quad (13)$$

onde  $\mathbf{y}_{(r)} = (y_1, \dots, y_{r-1}, y_{r+1}, \dots, y_n)$  e  $\pi(\boldsymbol{\theta}_i|\mathbf{y}_{(r)}, M_i)$  é a correspondente densidade *a posteriori* de  $\boldsymbol{\theta}_i$ , baseada em  $\mathbf{t}_{(r)}$ .

Usando o amostrador de Gibbs, as densidades em (13) pode ser aproximada por

$$\hat{f}(y_r|\mathbf{t}_{(r)}M_i) = \frac{1}{S} \sum_{s=1}^S f(y_r|\mathbf{t}_{(r)}, \boldsymbol{\theta}_i^{(s)}, M_i) \quad (14)$$

onde  $\boldsymbol{\theta}_i^{(s)}$  denota  $s$ -ésima amostra de Gibbs do vetor  $\boldsymbol{\theta}_i$ . Podemos utilizar as estimativas  $C\hat{P}O_r = \hat{f}(t_r|\mathbf{t}_{(r)}M_i)$  na seleção dos modelos, considerando um gráfico dos valores  $C\hat{P}O_r$  versus o número da observação  $r$  ( $r = 1, \dots, n$ ) para diferentes modelos; valores maiores (em média) indicam o melhor modelo. Também pode-se considerar para escolha do melhor modelo, aquele que maximiza  $MCPO = \sum_{r=1}^n \log(C\hat{P}O_r)/n$  (veja, Ibrahim, Chen e Sinha, 2001; Achcar, 2001).

Geisser e Eddy (1979) sugerem que o produto das densidades preditivas  $\prod_{r=1}^n f(y_r|\mathbf{y}_{(r)}, M_i)$  pode ser usado como um indicador na seleção de modelos. Se temos dois modelos  $M_1$  e  $M_2$ , a razão

$$PSFB_{12} = \frac{\prod_{r=1}^n f(y_r|\mathbf{y}_{(r)}, M_1)}{\prod_{r=1}^n f(y_r|\mathbf{y}_{(r)}, M_2)}. \quad (15)$$

pode ser usada como uma aproximação do fator de Bayes, denominada de pseudo-fator de Bayes. Assim, por exemplo, escolheremos o modelo  $M_1$  se  $PSFB_{12} > 1$ .

Ibrahim, Chen e Sinha (1998) propõem um critério bayesiano chamado de medida  $L$  para a avaliação e comparação de modelos. O critério da medida  $L$  é construído a partir da distribuição preditiva *a posteriori* dos dados, e pode ser escrito como soma de dois componentes, um envolvendo as médias da distribuição preditiva *a posteriori* e outro envolvendo as variâncias.

Seja  $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$  o vetor de respostas futuras com a mesma função de densidade de  $\mathbf{Y}|\boldsymbol{\theta}$ . A medida  $L$  é definida como

$$L = \sum_{i=1}^n Var(Z_i|\mathbf{Y}) + \tau \sum_{i=1}^n (\mu_i - y_i)^2. \quad (16)$$

onde  $\mu_i = E(Z_i|\mathbf{Y})$  e  $0 < \tau < 1$ . Valores pequenos da medida  $L$  indicam um melhor modelo. Ibrahim, Chen e Sinha (1998) obtiveram resultados teóricos para o caso de modelos lineares, mostrando que certos valores de  $\tau$  têm propriedades altamente desejáveis. Baseada em sua exploração teórica, Ibrahim, Chen e Sinha (1998) consideram que  $\tau = 1/2$  é uma escolha desejável e justificável para a seleção de modelos.

### 3 Exemplos

Com a finalidade de ilustrar a metodologia apresentada nas seções anteriores consideramos dois conjuntos de dados: um simulado a partir de um modelo logístico (Ratkowsky, 1983) com erro normal e dados reais apresentados por Sousa (1998) referentes a dados de crescimento de um cavalo pantaneiro.

#### 3.1 Exemplo 1

Os dados da Tabela 1 foram simulados a partir de um modelo logístico normal (Ratkowsky, 1983), expresso por

$$Y_i = \frac{\theta_1}{1 + \theta_2 \exp\{-\theta_3 x_i\}} + \varepsilon_i, \quad i = 1, \dots, n, \quad (17)$$

onde  $\varepsilon_i \sim N(0, \sigma^2)$ , independentes com parâmetros  $\theta_1 = 139$ ,  $\theta_2 = 8$ ,  $\theta_3 = 0,5$  e  $\sigma^2 = 10$ . Geramos um total de 32 observações considerando a variável  $x$  com valores entre 0,25 e 25. As estimativas dos parâmetros (erro padrão assintótico) de  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  e  $\sigma^2$  são dados por  $\hat{\theta}_1 = 139,86$  (1,665),  $\hat{\theta}_2 = 8,7425$  (0,838),  $\hat{\theta}_3 = 0,527$  (0,026) e  $\hat{\sigma}^2 = 25,012$  (6,253). As estimativas de máxima verossimilhança dos parâmetros do modelo logístico foram obtidas usando a sub-rotina MAXBFGS do aplicativo OX (ver Doornik, 1996).

Com a finalidade de mostrar que o modelo logístico, em que o erro é uma mistura de duas distribuições normais, é robusto a observações discrepantes, decidimos modificar a variável resposta, fazendo  $y_{32} = 180$ . Assim, as estimativas de máxima verossimilhança (EMV) dos parâmetros do modelo logístico dada em

Tabela 1 - Dados gerados do modelo logístico

$i$	$x_i$	$y_i$	$i$	$x_i$	$y_i$
1	0,25	11,599	17	4,75	73,479
2	0,50	18,991	18	5,00	79,582
3	0,75	18,172	19	6,00	98,411
4	1,25	25,693	20	7,00	108,620
5	1,50	30,664	21	8,00	122,140
6	1,75	31,541	22	9,00	132,300
7	2,00	35,050	23	10,00	135,270
8	2,25	38,247	24	11,00	135,300
9	2,50	45,343	25	12,00	132,770
10	2,75	45,016	26	13,00	136,310
11	3,25	55,827	27	14,00	131,070
12	3,50	63,793	28	15,00	139,290
13	3,75	63,211	29	16,00	139,120
14	4,00	66,331	30	18,00	140,310
15	4,25	78,545	31	20,00	141,420
16	4,50	76,458	32	25,00	144,040

(17) resultaram  $\hat{\theta}_1 = 144,03$  (2,930),  $\hat{\theta}_2 = 8,405$  (1,309),  $\hat{\theta}_3 = 0,499$  (0,043),  $\hat{\sigma}^2 = 70,209$  (17,554). Observe que os EMV foram afetados pela observação perturbada; também a estimativa do erro do modelo foi muito afetada.

A fim de analisarmos os dados da Tabela 1 considerando a observação perturbada, ( $y_{32} = 180$ ) em uma perspectiva bayesiana, inicialmente consideramos o modelo não-linear de crescimento logístico com termo aleatório modelado como uma mistura de duas distribuições normais. Para esse modelo consideramos a função de densidade *a priori* dada em (9) com  $a = b = 0,001$ ,  $c = 0$ ,  $d = 10^6$ ,  $e_1 = f_1 = 0,01$ ,  $e_2 = f_2 = 0,01$  e  $\theta_j \sim N(0, 10^3)$ ,  $j = 1, 2, 3$ . Observar que com essa escolha para os hiperparâmetros das distribuições *a priori* temos prioris não informativas para os parâmetros. Observar também que com as restrições  $\sum_{j=1}^2 p_j \mu_j = 0$  e  $p_1 = 1 - p_2$  contornamos o problema de não identificabilidade do modelo ( $p_1 = p$ ,  $p_2 = 1 - p$ ). Com essa escolha dos hiperparâmetros da distribuição *a priori*, geramos duas cadeias separadas com 50.000 iterações; monitorou-se a convergência da amostras Gibbs usando o método de Gelman e Rubin (1992), que utiliza a técnica de análise de variância para determinar se mais iterações são necessárias. Para cada parâmetro em cada cadeia as 10.000 primeiras iterações foram descartadas para eliminar o efeito do valores iniciais e para evitar o problema de autocorrelação foram tomadas amostras de 10 em 10 o que totaliza uma amostra final de tamanho 8.000. Na Tabela 2 reportamos o resumo *a posteriori* das distribuições dos parâmetros do modelo juntamente com os resultados da estimativa dos fatores de redução de escala



potencial  $\hat{R}$  (Gelman e Rubin, 1992). Observamos valores bastante próximos de 1, indicando que as cadeias convergiram.

Tabela 2 - Sumário *a posteriori* para o modelo logístico com erro mistura de normais

Parâmetro	Média	DP	Interv. cred. (95%)	$\hat{R}$
$\theta_1$	142,200	2,068	(138,800 ; 146,700)	1,002
$\theta_2$	7,183	0,542	(6,096 ; 8,218)	1,000
$\theta_3$	0,489	0,016	(6,096; 8,218)	1,001
$p$	0,919	0,0418	(0,821 ; 0,980)	1,000
$\mu_1$	-3,311	1,706	(0,081 ; 0,769)	1,001
$\mu_2$	37,83	2,279	(33,050 ; 41,900)	1,004
$\sigma_1^2$	7,13	1,513	(4,761 ; 10,710)	0,997
$\sigma_2^2$	1,118	0,403	(0,587 ; 2,082)	1,001

Uma estimativa de Monte Carlo da variância do erro do modelo logístico misto,  $Var(\varepsilon_i) = p\sigma_1^2 + (1-p)\sigma_2^2$ , baseada na amostra de tamanho 8.000 resultou 6,650 (1,432), que é uma quantidade mais próxima de  $\sigma^2 = 10$  comparada com o EMV dado por  $\hat{\sigma}^2 = 70,209$  (17,554).

Consideramos agora o modelo não-linear de crescimento logístico (17) e os dados da Tabela 1 com a observação  $y_{32}$  perturbada e assumindo uma distribuição normal para o erro. Para esse modelo adotamos as seguintes distribuições *a priori* independentes:  $\theta_j \sim N(0, 10^3), j = 1, 2, 3$ .  $\sigma^2 \sim GI(0, 01, 0, 01)$ , o que leva *a priori* não informativas. Com essa escolha, geramos duas cadeias separadas de Gibbs cada uma com 50.000 iterações e utilizamos o método de Gelman e Rubin (1992) para verificar a convergência das cadeias. Para cada parâmetro em cada cadeia as 10.000 primeiras iterações foram descartadas visando eliminar o efeito dos valores iniciais e daí foram tomadas amostras de 10 em 10 totalizando uma amostra final de tamanho 8.000. As quantidades *a posteriori* de interesse obtidas a partir das amostras selecionadas são dadas na Tabela 3, onde observamos que os fatores de redução potencial são menores de 1,1 ( $\hat{R} < 1,01$ ), indicando a convergência das amostras geradas.

Tabela 3 - Sumário *a posteriori* para o modelo logístico com distribuição normal para o erro

Parâmetro	Média	DP	Interv. cred. (95%)	$\hat{R}$
$\theta_1$	142,500	2,970	(136,500 ; 148,400)	1,001
$\theta_2$	8,178	1,306	(6,072 ; 11,17)	0,999
$\theta_3$	0,488	0,043	(0,410 ; 0,576)	1,002
$\sigma^2$	68,230	19,910	(39,850 ; 115,700)	1,002

Podemos observar na Tabela 3 que as estimativas dos parâmetros do modelo são próximas às estimativas de máxima verossimilhança e a estimativa da variância dos erros do modelo é influenciada fortemente pela observação 32, que não acontece no modelo logístico misto, o que mostra que este último modelo é robusto à observações discrepantes. Na Tabela 4 temos os valores preditos ( $z_i$ ) pelo modelo logístico (17) com erro mistura de duas distribuições normais (MRNL-LM) e modelo logístico com erro normal (MRNL-LN), notando que o valor predito da observação  $y_{32}$  é igual ao respectivo valor simulado.

Tabela 4 - Valores preditos de acordo com os modelos logístico com erro-mistura de duas distribuições (MRNL-LM) e com erro normal (MRNL-LN)

MRNL				MRNL			
		LM	LN			LM	LN
i	$y_i$	$z_i$	$z_i$	i	$y_i$	$z_i$	$z_i$
1	11,599	16,12	17,52	17	73,479	80,22	79,30
2	18,991	18,21	19,48	18	79,582	84,39	83,60
3	18,172	20,62	21,80	19	98,411	99,58	99,42
4	25,693	25,83	26,50	20	108,620	111,90	112,20
5	30,664	28,75	29,21	21	122,140	120,90	122,10
6	31,541	31,90	32,22	22	132,300	127,30	129,20
7	35,050	35,25	35,22	23	135,270	131,60	133,80
8	38,247	38,81	38,70	24	135,300	134,20	136,90
9	45,343	42,43	42,02	25	132,770	135,90	139,10
10	45,016	46,30	45,84	26	136,310	137,10	140,30
11	55,827	54,55	53,72	27	131,070	137,80	141,00
12	63,793	58,66	57,86	28	139,290	138,20	141,50
13	63,211	63,03	62,04	29	139,120	138,50	141,90
14	66,331	67,34	66,37	30	140,310	138,80	142,10
15	78,545	71,66	70,75	31	141,420	138,80	142,50
16	76,458	76,03	75,04	32	<b>180,000</b>	180,00	142,40

Comparamos os ajustes do modelo logístico normal e do modelo logístico com mistura aos dados da Tabela 1, considerando a observação 32 perturbada. Para isso, utilizamos as técnicas descritas na seção 2.1. Com isso, foram calculadas as medidas  $L$  (com  $\tau = 1/2$ ), MCPO e o critério de informação bayesiana (BIC) usada por Sahu e Dey (2000), para o modelo logístico com erro normal e logístico com erro uma mistura de duas distribuições normais. Com as amostras de Gibbs geradas dos modelos considerados na Tabela 5 mostramos as estimativas de Monte Carlo dos critérios bayesianos baseados nas amostras simuladas de Gibbs.

Podemos observar da Tabela 5 que o melhor modelo é o logístico com mistura de duas distribuições normais para a componente aleatória. Na Figura 1, apresentamos o gráfico de  $CPO_r$  para os modelos logístico misto e logístico normal

Tabela 5 - Critérios de seleção de modelos-Exemplo 1

Critério	Modelo	
	Logístico misto	Logístico normal
<i>MCPO</i>	-2,926	-3,776
<i>L</i>	456,521	3305,230
<i>BIC</i>	201,280	238,663

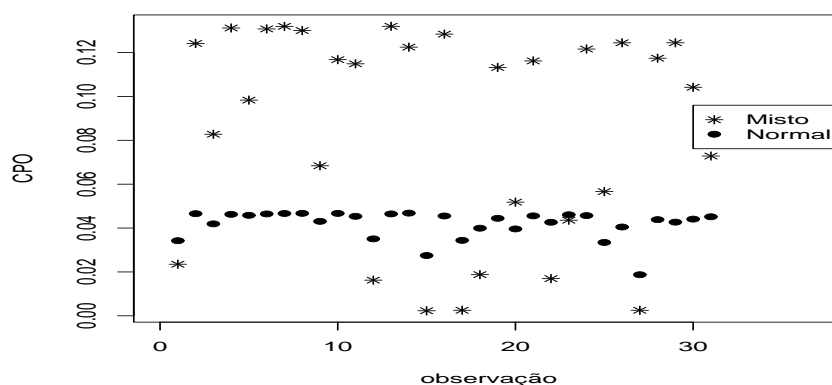


Figura 1 - Gráfico de CPO contra o número de observações para o modelo logístico normal e misto.

*versus* o número das observações. Observamos que aproximadamente 78% das observações estão melhor ajustadas pelo modelo logístico misto. O pseudofator de Bayes (15), comparando os modelos misto e normal é igual a  $1,761347 \times 10^{12}$ , indicando uma forte evidência favorável ao modelo misto. Esses resultados indicam que o modelo não-linear misto é robusto à presença de observações discrepantes.

### 3.2 Exemplo 2

Consideramos o conjunto de dados apresentados por Sousa (1998) (ver Tabela 6), os quais se referem ao crescimento de um cavalo pantaneiro. A variável é a altura na cercelha do cavalo (em cm). Esses dados foram analisados por Oliveira e Sousa (1996) utilizando o modelo não-linear de crescimento tipo-Weibull com componente aleatória normal.

Neste trabalho analisamos os dados da Tabela 6 de uma perspectiva bayesiana, considerando os modelos logístico dado em (17) e o modelo tipo-Weibull,

$$Y_i = \theta_1 - \theta_2 \exp\{-\theta_3 x_i^{\theta_4}\} + \varepsilon_i, \quad i = 1, \dots, 15. \quad (18)$$

Tabela 6 - Dados de crescimento de um cavalo pantaneiro

Idade (dias)	5	54	131	195	286	412	476	586
Altura (cm)	86	96	107	114	121	124	124	129
Idade (dias)	682	785	817	874	1494	1530	1711	
Altura(cm)	131	135	135	135	138	138	138	

onde as componentes aleatórias dos modelos seguem uma mistura de duas distribuições normais dada em (2). Como em Sousa (1998), dividimos a idade por 1.800 para restringir os valores de  $x_i$  no intervalo  $(0,1)$ . Com o propósito de comparar o ajuste dos modelos, adotamos as mesmas distribuições *a priori* para os parâmetros envolvidos nos dois modelos.

Considerando a densidade *a priori* dada em (9), com  $\theta_j \sim N(0, 10^6)$ ,  $j = 1, 2, 3$  ou  $4$ ,  $p \sim \text{Beta}(1,1)$ ,  $\mu_1 \sim N(0, 10^6)$ ,  $\sigma_j^2 \sim \text{GI}(0,01; 0,01)$ ,  $j = 1, 2$ , geramos duas cadeias paralelas cada uma com 110.000 iterações. Como no exemplo 1 monitoramos a convergência das amostras Gibbs usando o método de Gelman e Rubin (1992). Para cada parâmetro as 10.000 primeiras iterações foram descartadas e para evitar o problema de autocorrelação das amostras geradas, foram tomadas amostras de 20 em 20, o que totaliza uma amostra final de tamanho 10.000. Na Tabela 7, reportamos o resumo *a posteriori* dos parâmetros do modelo juntamente com os resultados da estimativa dos fatores de redução de escala potencial  $\hat{R}$  (veja, Gelman e Rubin, 1992) para todos os parâmetros. Observamos valores bastante próximos de 1, o que indica que as cadeias convergiram.

Consideramos agora os modelos (17)-(18) com componente aleatório que segue uma distribuição normal, com média zero e variância  $\sigma^2$ . Ajustamos esses modelos aos dados da Tabela 6, considerando que os parâmetros dos modelos têm as seguintes densidades *a priori* independentes,  $\theta_j \sim N(0, 10^6)$ ,  $j = 1, 2, 3$  ou  $4$ ,  $\sigma^2 \sim \text{GI}(0,01; 0,01)$ . Com essa escolha da distribuição *a priori* geramos duas cadeias separadas de Gibbs, cada uma com 110.000 iterações, e utilizamos o método de Gelman e Rubin (1992) para verificar a convergência das cadeias. Para cada parâmetro, as 10.000 primeiras iterações foram descartadas para eliminar o efeito dos valores iniciais, e então foram tomadas amostras de 20 em 20, o que totaliza uma amostra final de tamanho 10.000. As quantidades *a posteriori* de interesse obtidas a partir das amostras selecionadas são dadas na Tabela 8, onde observamos que os fatores de redução potencial são próximos de *um* indicando a convergência das amostras geradas.

Comparamos os ajustes dos modelos logístico e tipo-Weibull aos dados da Tabela 6. Na Tabela 9 temos as estimativas de Monte Carlo dos critérios bayesianos descritos na seção 2.1. Observamos que o modelo tipo-Weibull com erro mistura de duas distribuições normais se ajusta melhor aos dados.

Com as amostras de Gibbs geradas, temos na Tabela 10 as estimativas do pseudofator de Bayes (11) para seleção do melhor modelo para os dados da Tabela 6.

Tabela 7 - Sumário *a posteriori* para os parâmetros dos modelos logístico e tipo-Weibull com mistura de normais como erro, para os dados da Tabela 6

Modelo	Parâmetro	Média	DP	Int. cred. (95%)	$\hat{R}$
Logístico	$\theta_1$	136,1	0,9021	(134,5 ; 138,0)	0,996
	$\theta_2$	0,5394	0,02898	( 0,4772 ; 0,585)	0,996
	$\theta_3$	8,332	0,6589	(6,672; 9,377)	1,003
	$\sigma_1^2$	2,438	2,446	(0,5479 ; 8,374)	0,997
	$\sigma_2^2$	1,269	1,093	(0,3846 ; 4,072)	0,999
	$\mu_1$	-2,549	1,883	(-4,808 ; -0,4236)	0,998
	$\mu_2$	1,286	0,9362	(0,2826; 2,6512)	0,999
	$p$	0,3523	0,1455	(0,1372 ; 0,6579)	1,000
	$Var(\varepsilon_i)$	1,81	1,265	( 0,6302 ; 5,089)	1,007
Tipo-Weibull	$\theta_1$	138,7	1,69	(135,8 ; 142,6)	1,000
	$\theta_2$	54,37	4,287	(47,49 ; 63,86)	0,998
	$\theta_3$	4,856	1,135	( 3,233 ; 7,771)	1,001
	$\theta_4$	0,8353	0,1117	(0,6302 ; 1,088)	1,001
	$\sigma_1^2$	1,35	0,7133	(0,5198 ; 3,238)	1,007
	$\sigma_2^2$	1,229	0,5939	(0,4821 ; 2,737)	1,009
	$\mu_1$	-0,4722	1,795	(-3,364 ;3,009)	0,999
	$\mu_2$	0,2187	0,9078	(-1,592 ; 1,837)	1,000
	$p$	0,375	0,182	(0,07047 ; 0,7676)	1,007
$Var(\varepsilon_i)$	0,8353	0,1117	(0,6302 ; 1,088)	1,007	

Tabela 8 - Sumário *a posteriori* para os parâmetros dos modelos logístico e tipo-Weibull com erro normal para os dados da Tabela 6

Modelo	Parâmetro	Média	DP	Int. cred. (95%)	$\hat{R}$
Logístico	$\theta_1$	136,5	1,2881	(133,9 ;139,0)	1,000
	$\theta_2$	0,546	0,03386	(0,4814 ;0,6164 )	1,001
	$\theta_3$	8,155	0,9831	(6,465; 9,377)	0,999
	$\sigma^2$	6.637	3,397	(2.826 ; 15.17)	1,000
Tipo-Weibull	$\theta_1$	139,12	1,414	(136,6 ; 142,3)	1,000
	$\theta_2$	55,47	2,92	(50,09 ; 61,75)	0,998
	$\theta_3$	4,464	0,7097	(3,248 ; 6,025)	1,001
	$\theta_4$	0,7994	0,07768	(0,6537 ; 0,9592)	1,001
	$\sigma^2$	2,439	1,32	(1,005 ; 5,75)	1,007

Segundo esse critério, o modelo não-linear tipo-Weibull é o melhor modelo para esses dados.

Tabela 9 - Comparação dos modelos não-linear logístico e tipo-Weibull com erros normais e erros-mistura de duas distribuições normais – Exemplo 2

Critério	Modelo logístico		Modelo tipo-Weibull	
	Misto	Normal	Misto	Normal
<i>MCPO</i>	-2,502	-2,756	-2,041	-2,663
<i>L</i>	50,598	153,105	30,933	57,327
<i>BIC</i>	65,441	77,503	62,341	64,292

Tabela 10 - Pseudofator de Bayes

Modelo	PSFB
Tipo-Weibull misto <i>vs</i> tipo-Weibull normal	11420,90
Tipo-Weibull misto <i>vs</i> logístico misto	1012,42
Tipo-Weibull misto <i>vs</i> logístico normal	46083,00
Logístico misto <i>vs</i> logístico normal	45,52
Logístico misto <i>vs</i> tipo-Weibull Normal	11,28
Tipo-Weibull normal <i>vs</i> logístico Normal	4,03

## 4 Conclusões

Neste artigo, consideramos classes de modelos não-lineares (MRNL) com erros distribuídos como uma mistura de distribuições normais. Em muitas situações práticas, a suposição usual de normalidade para o erro em modelos não-lineares não é satisfatória. Isso pode ser observado em gráficos de resíduos, onde muitas vezes há a presença de multimodalidade. Portanto, a suposição de misturas finitas de distribuições normais para o erro em geral leva à melhores inferências, e principalmente melhores previsões para respostas de interesse. O uso de métodos Bayesianos, especialmente considerando técnicas de simulação de Monte Carlo em cadeias de Markov tem sido muito útil para a determinação de sumários *a posteriori* de interesse em modelos assumindo misturas finitas de distribuições. Os resultados obtidos neste artigo, considerando misturas de duas distribuições normais para o erro, pode facilmente ser estendido para misturas com mais de dois componentes.

Além de melhores inferências, o uso de uma mistura finita de distribuições normais para o erro em modelos de regressão não-lineares leva a resultados robustos à presença de observações discordantes (“outliers”).

CANCHO, G. V.; ACHCAR, J. A.; ORTEGA, E. M. M. Non-linear regression models assuming that the error term has a finite mixture of normal distributions. *Rev. Mat. Estat.*, São Paulo, v.24, n.3, p.35-60, 2006.

- **ABSTRACT:** *In this paper, we explore the use of Markov Chains Monte Carlo (MCMC) methods to develop a Bayesian analysis for non-linear regression (NLMR) models assuming that the error term has a finite mixture of normal distributions. We use the Metropolis-Hastings algorithm to simulate samples for the model. We illustrate the the proposed methodology with two examples, considering simulated and real data sets.*
- **KEYWORDS:** *Non-linear regression; selection of models; Markov Chain Monte Carlo; mixture of normal distribution.*

## Referências

- ACHCAR, J. A. Bayesian analysis for software reliability data. In: BALAKRISHNAN, N.; RAO, C. R. (ed.), Handbook of statistics. Holland: Elsevier Science, 2001. v.20, p.733-748. (Advances in Reliability)
- AITKIN, M. Posterior factor. *J. R. Stat. Soc. Ser. B.*, London, v.53, p.111-142, 1991.
- BERGER, J.; PERICCHI, L. R. The intrinsic Bayes factor. *J. Am. Stat. Assoc.*, New York, v. 91, p.109-122, 1996.
- DIEBOLT, J.; ROBERT, C. P. Estimation of finite mixture distributions through Bayesian sampling, *J. R. Stat. Soc. Ser. B*, London, v.56, p.363-375, 1994.
- DOORNIK, J. A. *Ox*: An object-oriented matrix programming language. London: International Thomson Business Press, 1996. 349p.
- GEISSER, S.; EDDY, W. A predictive approach to model selection. *J. Am. Stat. Assoc.*, New York, v.74, p.153-160, 1979.
- GELFAND, A. E.; DEY, D. K. Bayesian model choice: asymptotics and exact calculations. *J. R. Stat. Soc. Ser. B.*, London, v.56, p.501-514, 1994.
- GELFAND, A. E.; GOSH, S. K. Model choice. A minimum posterior predictive loss approach. *Biometrika*, London, v.85, p.398-409, 1998.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.*, Hayward, v.7 p.457-511, 1992.
- IBRAHIM, J. G.; CHEN, M. -H.; SINHA. Criterion based methods for Bayesian model assessment. *Stat. Sinc.*, Taiwan, v.11, p.419-443, 1998.
- IBRAHIM, J. G.; CHEN, M. -H.; SINHA. Bayesian survival analysis. New York: Springer Verlag, 2001. p.479.
- NEWTON, M. A.; RAFTERY, A. E. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J. R. Stat. Soc. Ser. B.*, London, v.56, p.1-48, 1994.
- OLIVEIRA, M. R.; SOUZA, G. S. Medidas de não-linearidade em regressão não-linear: In: SINAPE, 12., 1996, Caxambu. *Resumos...* Caxambu:ABE, 1996. p.295-296.

RATKOWSKY, D. A. *Nonlinear regression models*. New York: Marcel Dekker, 1983. p.276.

ROBERT, C. P. Mixtures of distributions: inference and estimation. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHATER, D. J. (Ed.). *Markov chain Monte Carlo in practice*. London: Chapman and Hall, 1996. p.486.

ROEDER, K.; WASSERMAN, L. Practical Bayesian density estimation using mixtures of normals. *J. Am. Stat. Assoc.*, New York, v.92, p.894-902, 1997.

SAHU, S. K.; DEY, D. K. A comparison of frailty and other models for bivariate survival data. *Lifetime Data Anal.*, Boston, v.6, p.207-228, 2000.

SOUSA, G. S. *Introdução aos modelos de regressão linear e não-linear*. Brasília: Embrapa-SP/Embrapa-SEA, 1998. p.489.

TANNER, M.; WONG, W. The calculation of posterior distribution by data augmentation. *J. Am. Stat. Assoc.*, New York, v.82, p.528-550, 1987.

TETTERINGTON, D. M.; SMITH, A. F. M.; MAKOV, U. E. *Statistical analysis of finite mixture distributions* New York: Wiley, 1995. p.243.

Recebido em 24.05.2006.

Aprovado após revisão em 25.08.2006.