

## Modelos Não Lineares de Efeitos Mistos sob a distribuição slash: Aplicação do Algoritmo SAEM em dados de Cefamandol

João Lucas Gomes Aguiar<sup>1</sup>, José Clelto Barros Gomes<sup>1</sup>

<sup>1</sup>Universidade Federal do Amazonas, Departamento de Estatística, Av. Rodrigo Otávio Jordão Ramos, 6200, Coroado I, 69080-900, Manaus AM, Brasil

**Palavras-Chave:** NLME, Efeitos mistos, Distribuição Slash, SAEM.

### Introdução

Modelos Não Lineares de Efeitos Mistos (NLME) são ferramentas essenciais para a análise de dados longitudinais e de medidas repetidas, com aplicações em áreas como farmacologia e biologia, conforme apontado por Davidian e Giltinan<sup>2</sup> em dados de cefamandol. A suposição de normalidade nos erros e efeitos aleatórios, comum nesses modelos, é frequentemente inadequada devido a presença de observações atípicas. Para contornar essa limitação, a literatura propõe o uso de distribuições de caudas pesadas. Destaca-se a família de distribuições de Mistura de Escala de Normais Multivariadas (SMN), que inclui as distribuições t, normal contaminada e slash como Meza et. al<sup>4</sup> com caudas pesadas. A estimação dos parâmetros nesses modelos pode ser feita por diversos métodos, uma abordagem popular é o algoritmo EM (Expectation-Maximization), mas sua aplicação é limitada quando o passo de esperança não possui solução analítica. Para superar isso, variantes como o MCEM (Monte Carlo EM), que utiliza simulações de Monte Carlo são propostas. No entanto, o MCEM tem um alto custo computacional e dificuldades de convergência em modelos muito complexos. Nesse cenário, o algoritmo SAEM (Stochastic Approximation EM), proposto por Delyon et. al<sup>3</sup>, emerge como uma alternativa mais eficiente e com melhores garantias de convergência.

### Material e Métodos

A família de distribuições SMN, conforme descrita em Andrews e Mallows<sup>1</sup>, que generaliza a distribuição normal ao introduzir uma variável aleatória de escala, escrevendo a variável  $Y$  na seguinte forma estocástica.

$$Y = \mu + J^{-1/2}Z, \quad J \sim H(j; g) \quad (1)$$

em que  $Z$  é um vetor aleatório com distribuição normal multivariada, com de média zero e matriz covariância  $\Sigma$ , e  $J$  é uma variável aleatória positiva, independente de  $Z$ , que representa a escala da mistura. A variável  $J$  tem função de distribuição acumulada  $H(j; g)$ , onde  $g$  denota os parâmetros da mistura (que podem ser escalares ou vetoriais). A estrutura do modelo NLME é definida em dois estágios. Para um conjunto de  $n$  indivíduos, com um vetor de observações  $Y_i$  de dimensão  $n_i \times 1$  para cada

indivíduo  $i$ , o modelo é:

$$\begin{aligned} Y_i &= f_i(\phi_i; \mathbf{X}_i) + \epsilon_i, \\ \phi_i &= \mathcal{A}_i\beta + \mathbf{b}_i, \end{aligned} \quad (2)$$

onde o primeiro estágio descreve a variabilidade intraindividual e o segundo, a populacional. Aqui,  $f_i$  é uma função não linear;  $\mathbf{X}_i$  é o vetor de covariáveis;  $\phi_i$  é o vetor de parâmetros para o  $i$ -ésimo indivíduo;  $\mathcal{A}_i$  é a matriz de delineamento;  $\beta$  é o vetor de efeitos fixos e  $\mathbf{b}_i$  é o vetor de efeitos aleatórios. Para incorporar a robustez, assumimos que tanto os erros aleatórios ( $\epsilon_i$ ) quanto os efeitos aleatórios ( $\mathbf{b}_i$ ) seguem distribuições da família SMN:

$$\begin{aligned} Y_i | \mathbf{b}_i &\stackrel{ind}{\sim} SMN_{n_i}(f_i(\phi_i, \mathbf{X}_i), \sigma^2 \mathbf{I}_{n_i}; H_1), \\ \mathbf{b}_i &\stackrel{ind}{\sim} SMN_q(\mathbf{0}, \Psi; H_2), \quad i = 1, \dots, n, \end{aligned} \quad (3)$$

em que  $\Psi$  é a matriz de covariância dos efeitos aleatórios. Utilizando a representação hierárquica das distribuições SMN, o modelo pode ser expresso de forma condicionalmente normal:

$$\begin{aligned} Y_i | \mathbf{b}_i, J_i &\sim \mathcal{N}_{n_i}(f_i(\phi_i, \mathbf{X}_i), J_i^{-1} \sigma^2 \mathbf{I}_{n_i}), \\ \mathbf{b}_i | G_i &\stackrel{ind}{\sim} \mathcal{N}_q(\mathbf{0}, G_i^{-1} \Psi), \\ J_i &\stackrel{ind}{\sim} H_1(j), \quad G_i \stackrel{ind}{\sim} H_2(g), \quad i = 1, \dots, n, \end{aligned} \quad (4)$$

em que  $J_i$  e  $G_i$  são variáveis latentes de escala. A consequência dessa estrutura é que a função de verossimilhança marginal  $L(\theta)$  não possui forma analítica fechada, pois exige a integração sobre os dados latentes ( $\mathbf{b}_i, J_i, G_i$ ). A complexidade da função de verossimilhança marginal inviabiliza a maximização direta. Para estimar os parâmetros do modelo, utilizamos algoritmos iterativos. Embora o algoritmo EM seja uma base para problemas com dados latentes, sua aplicação direta aqui é intratável. Por isso, adotamos o algoritmo SAEM, que substitui o Passo-E por uma aproximação estocástica, garantindo a convergência para as estimativas de máxima verossimilhança sob condições gerais sendo mais computacionalmente eficiente. O passo de simulação do SAEM, é implementado com métodos MCMC (Monte

Carlo por Cadeias de Markov). Especificamente, o algoritmo Metropolis-Hastings é utilizado para gerar amostras das distribuições condicionais dos efeitos aleatórios, por fim a etapa de maximização que atualiza os parâmetros do modelo.

## Resultados e Discussão

O algoritmo SAEM foi aplicado ao banco de dados de Cefamandol<sup>2</sup>, que contém as concentrações plasmáticas do antibiótico em seis indivíduos ao longo do tempo. O dados e os ajustes do modelo podem ser visualizados na Figura 1, baseado no modelo com erro Slash<sub>5</sub> e efeito aleatório Normal.

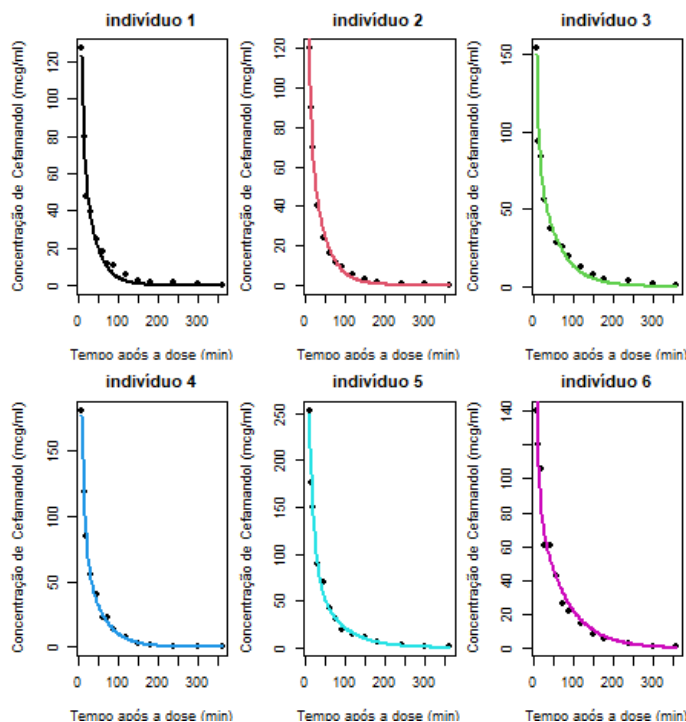


Figura 1: Concentração de Cefamandol (onde os pontos são dados reais e as curvas são os ajustes individuais).

Uma análise diagnóstica inicial, utilizando a distância de Mahalanobis (Figura 2) indicou a presença de indivíduos atípicos. Essas observações justificaram a aplicação de modelos com caudas pesadas.

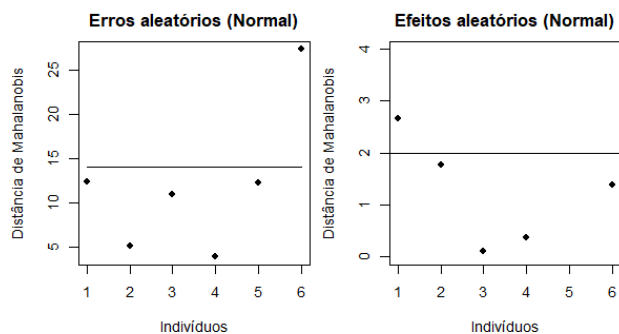


Figura 2: Distância de Mahalanobis para Normal-Normal

Quatro modelos diferentes foram ajustados utilizando uma malha de graus de liberdade e seus resultados são

apresentados na Tabela 1. A comparação dos modelos, baseada nos critérios de informação, demonstrou a superioridade do modelo Slash<sub>5</sub>-Normal, onde apresentou o menor AIC (545,68) e BIC (544,22), sendo considerado o mais adequado e parcimonioso para descrever os dados.

Tabela 1: Estimativa dos parâmetros (Est.) com erros padrões (E.P.) para os modelos ajustados aos dados do cefamandol.

Par	NN	S <sub>5</sub> N	NS <sub>2,25</sub>	S <sub>5,25</sub> S <sub>5</sub>
	Est. (E.P.)	Est. (E.P.)	Est. (E.P.)	Est. (E.P.)
$\hat{\phi}_0$	385,56 (33,25)	384,83 (37,64)	388,90 (16,84)	385,08 (118,69)
$\hat{\eta}_1$	-1,74 (0,10)	-1,78 (0,08)	-1,75 (0,02)	-2,16 (0,10)
$\hat{\eta}_2$	115,60 (8,15)	104,08 (9,42)	116,37 (2,89)	102,21 (6,42)
$\hat{\tau}_0$	-3,76 (0,09)	-3,82 (0,09)	-3,72 (0,03)	-3,85 (0,19)
$\hat{d}_1$	0,10 (0,04)	0,10 (0,04)	0,06 (0,03)	0,83 (1,37)
$\hat{d}_2$	0,08 (0,04)	0,08 (0,03)	0,05 (0,03)	0,07 (0,02)
$\hat{\sigma}^2$	25,23 (4,86)	15,86 (4,74)	25,22 (2,82)	16,04 (5,79)
$\hat{L}_o(\hat{\theta})$	-277,10	-265,84	-277,11	-270,02
AIC	568,20	545,68	568,23	554,04
BIC	566,74	544,22	566,77	282,56

## Conclusões

Os resultados deste estudo demonstraram a eficácia da abordagem de NLME com distribuições de cauda pesada, especificamente a distribuição slash, para a análise de dados longitudinais. Especificamente, o modelo Slash<sub>5</sub>-Normal demonstrou ser o mais adequado e parcimonioso para descrever os dados por apresentar o menor AIC e BIC, além do maior valor de log-verossimilhança. Adicionalmente, a metodologia proposta, que emprega o algoritmo SAEM, mostrou-se uma alternativa viável e computacionalmente eficiente para a estimação dos parâmetros.

## Agradecimentos

Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) pela bolsa concedida por meio do Programa de Apoio à Iniciação Científica do Amazonas (PAIC) e pela Universidade Federal do Amazonas (UFAM).

## Referências

- [1] Andrews, D. F.; Mallows, C. L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 36:99–102. 1974.
- [2] Davidian, M.; Giltinan, D. M. *Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall. 1995.
- [3] Delyon, B.; Lavielle, M.; Moulines, E. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128. 1999.
- [4] Meza, C.; Osorio, F.; De La Cruz, R. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*, 22:121–139. 2012.
- [5] Rogers, W. H.; Tukey, J. Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica*, 26:211–226. 1972.