23^{o}SINAPE - Simpósio Nacional de Probabilidade e Estatística

Concurso Melhor Tese de Doutorado

Estimação e diagnóstico em modelos multivariados para dados censurados

Larissa Avila Matos

Prof. Dr. Victor Hugo Lachos Dávila - Orientador
Prof. Dr. Luis Mauricio Castro Cepero - Coorientador

Sumário

Introdução

Preliminares

Trabalho 1

Trabalho 2

Trabalho 3

Trabalho 4

Trabalho 5

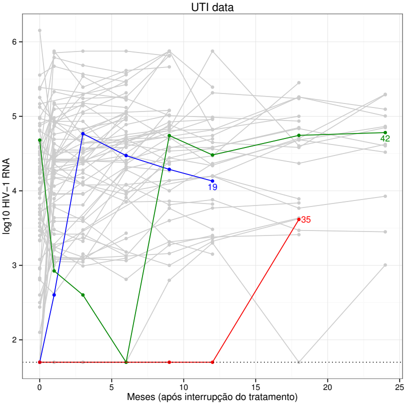
Referências

Motivação

- Em muitos estudos clínicos, o uso de modelos longitudinais tem mostrado um crescimento significativo nos últimos anos e tornou-se uma ferramenta poderosa para modelar tais resultados.
- Em pesquisa sobre a AIDS, o estudo da dinâmica do vírus da imunodeficiência humana (HIV) tem recebido atenção significativa na literatura biomédica, permitindo-nos compreender a patogênese do HIV, e avaliar a eficácia da terapia anti-retroviral (ARV).
- A maioria dos ensaios clínicos na terapia ARV avaliam as taxas/mudanças nas cargas virais HIV-1 RNA, que são coletados longitudinalmente.
- O monitoramento dessas cargas virais é considerado fundamental.
- Entretanto, dependendo dos ensaios de diagnóstico utilizados, as medidas das cargas virais podem estar sujeitas a limites de detecção superiores ou inferiores, valores acima ou abaixo desses limites não são quantificados.

Estudos Motivacionais de dados do HIV

- Estudo 1: **UTI data**
  - → 72 crianças e adolescentes infectados pelo HIV no período perinatal;
  - → 7% dos dados são censurados a esquerda.

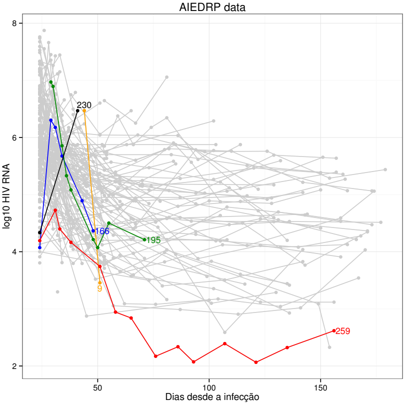


Estudos Motivacionais de dados do HIV

- Estudo 2: **AIEDRP study**

  $\rightarrow$ 320 indivíduos não tratados com infecção aguda pelo HIV do Programa AIEDRP, um grande estudo observacional multicêntrico de indivíduos com infecção aguda e precoce do HIV;

  $\rightarrow$ 22% dos dados são censurados a direita.

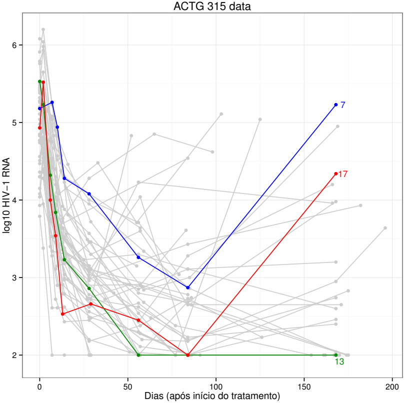


Estudos Motivacionais de dados do HIV

- Estudo 3: **ACTG 315 data**

  → 46 pacientes infectados pelo HIV tratados com um potente coquetel de drogas anti-retroviral;

  → 11% dos dados são censurados a esquerda.



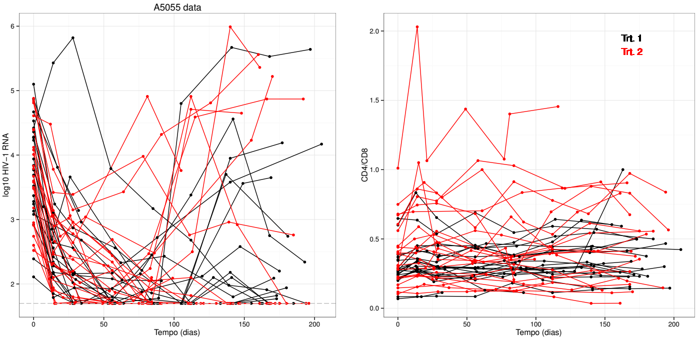
Estudos Motivacionais de dados do HIV

- Estudo 4: **A5055 study**

  → 44 pacientes infectados pelo HIV-1 tratados com uma de duas terapias anti-retrovirais;

  → 2 variáveis respsostas: $\log_{10}$(RNA) e CD4/CD8, onde CD4 e CD8 são dois marcadores imunológicos freqüentemente utilizados para o monitoramento da progressão do HIV;

  → 33% dos dados são censurados a esquerda.



Trabalhos recentes

Modelos Longitudinais

Estudos Longitudinais para dados censurados utilizando distribuição normal

- Samson *et al.* (2006) [*Computational Statistical & Data Analysis*]
- Vaida *et al.* (2007) [*Computational Statistical & Data Analysis*]
- Vaida & Liu (2009) [*Journal of Computational and Graphical Statistics*]
- Matos *et al.* (2013a) [*Computational Statistical & Data Analysis*]

Estudos Longitudinais para dados censurados utilizando distribuições com caudas pesadas

- Lachos *et al.* (2011) [*Biometrics*]
- **Garay et al. (2014)** [*Statistical Methods in Medical Research*]
- **Matos et al. (2013b)** [*Statistica Sinica*]
- **Wang et al. (2015)** [*Statistical Methods in Medical Research*]

Proposta

- **Objetivo:** modelar variáveis com respostas multivariadas censuradas utilizando distribuições com caudas pesadas.

- **Solução clássica:** No contexto frequentista, a principal hipótese assumida é a de que os termos aleatórios seguem uma distribuição normal ou t de Student multivariada e o algoritmo EM é utilizado para a estimação dos parâmetros.

- **Problema:** alguns conjuntos de dados não são compatíveis com a suposição de normalidade, seja pela cauda pesada ou pela presença de valores atípicos. E dependendo da distribuição escolhida para os termos aleatórios o algoritmo EM não pode ser implementado.

- **Proposta:** utilizar distribuições mais flexíveis para os termos aleatórios. Neste caso, trabalharemos com a chamada classe de distribuições de mistura de escala normal (SMN) e para o procedimento de estimação nós vamos adotar o algoritmo SAEM.

A Tese

- A tese consistiu em 5 artigos (atualmente aceitos) para modelar dados longitudinais com resposta censurada:

  Trabalho 1: **MATOS, LARISSA A.**; CASTRO, LUIS M.; LACHOS, VICTOR H.. *Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads.* TEST, v. 25, p. 627-653, 2016.

  Trabalho 2: **MATOS, LARISSA A.**; BANDYOPADHYAY, DIPANKAR; CASTRO, LUIS M.; LACHOS, VICTOR H.. *Influence assessment in censored mixed-effects models using the multivariate Students-t.* JOURNAL OF MULTIVARIATE ANALYSIS, v. 141, p. 104-117, 2015.

  Trabalho 3: **MATOS, LARISSA A.**; LACHOS, VICTOR H.; LIN, T.I.; CASTRO, LUIS M. (2018). *Heavy-tailed longitudinal regression models for censored data: A robust parametric approach.* (TEST - doi: [10.1007/s11749-018-0603-5])

  Trabalho 4: LACHOS, VICTOR H.; **MATOS, LARISSA A.**; CASTRO, LUIS M.; CHEN, MING-HUI (2018). *Flexible Longitudinal Linear Mixed Models for Multiple Censored Responses Data.* (Statistics in Medicine - Conditionally accepted)

  Trabalho 5: **MATOS, LARISSA A.**; CASTRO, LUIS M.; CABRAL, CELSO RÔMULO BARBOSA; LACHOS, VICTOR H. (2018). *Multivariate Measurement Error Models Based on Student-t Distribution under Censored Responses.* (Statistics).

Distribuições da classe SMN

Andrews & Mallows (1974); Lange & Sinsheimer (1993)

Representação estocástica

$$\mathbf{y} = \boldsymbol{\mu} + \kappa(\mathrm{U})^{1/2}\mathbf{Z}, \tag{1}$$

- $\mu \in \mathbb{R}$ um parâmetro de locação;
- $Z \sim N(0, \boldsymbol{\Sigma})$;
- U uma v.a. positiva com f.d.a. $\mathrm{H}(u|\boldsymbol{\nu})$ e f.d.p. $h(u|\boldsymbol{\nu})$;
- $\boldsymbol{\nu}$ um escalar ou um vetor de parâmetro indexando a distribuição de U;
- $\kappa(\mathrm{U})$ uma função de pesos;
- Z e U são independentes;
- Notação: $\mathbf{y} \sim \mathrm{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathrm{H})$.

Distribuições da classe SMN

- Distribuições condicionais:

$$\begin{aligned} \mathbf{y}|\mathrm{U}=u &\sim N(\boldsymbol{\mu},\kappa(u)\boldsymbol{\Sigma}), \\ \mathrm{U}=u &\sim h(u|\boldsymbol{\nu}). \end{aligned} \tag{2}$$

- Densidade:

$$f(\mathbf{y}) = \int_0^\infty \phi_p(\mathbf{y};\boldsymbol{\mu},\kappa(u)\boldsymbol{\Sigma})d\mathrm{H}(u|\boldsymbol{\nu}). \tag{3}$$

Distribuições da classe SMN

Casos particulares: $\mathbf{y} \in \mathbb{R}^p$

- **Normal multivariada**
  - $P(U = 1) = 1$;
  - Função de distribuição: $\mathrm{N}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- **t de Student multivariada**
  - $\mathrm{U} = Gama(\nu/2, \nu/2)$;
  - $\kappa(u) = 1/u$;
  - Função de distribuição:

$$\mathrm{T}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{p+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2},$$

onde $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$.

- **Slash multivariada**
  - $\mathrm{U} = Beta(\nu, 1)$;
  - $\kappa(u) = 1/u$;
  - Função de distribuição:

$$\mathrm{SL}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) du, \quad u \in (0,1), \quad \nu > 0.$$

- **Normal Contaminada multivariada**
  - U é uma v.a. discreta com f.p. $h(u|\boldsymbol{\nu}) = \nu_1 \mathbb{I}_{\{\nu_2\}}(u) + (1-\nu_1)\mathbb{I}_{\{1\}}(u)$ e $\boldsymbol{\nu} = (\nu_1, \nu_2)$;
  - $\kappa(u) = 1/u$;
  - Função de distribuição:

$$\mathrm{CN}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \nu_1 \phi_p(\mathbf{y}; \boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Sigma}) + (1-\nu_1)\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

  O parâmetro ν_1 pode ser interpretado como a proporção de outliers e ν_2 como um fator de escala.

Algoritmo SAEM

Algoritmo EM - Dempster *et al.* (1977)

Seja $\boldsymbol{\theta}$ o vetor de parâmetros e $\boldsymbol{y}_c = (\mathbf{y}^\top, \boldsymbol{q}^\top)$ o vetor de dados completos, i.e., os dados observados $\mathbf{y}^\top$ e os dados faltantes (ou as variáveis latentes) $\boldsymbol{q}^\top$. O algoritmo EM consiste basicamente em dois passos: A esperança (Passo E) e a maximização (Passo M).

- **Passo E:** Calcula a esperança da log-verossimilhança completa condicionada ao vetor de dados observados, denotada por

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E\left[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right],$$

  onde $\widehat{\boldsymbol{\theta}}^{(k)}$ é a estimativa de $\boldsymbol{\theta}$ na k–ésima iteração.

- **Passo M:** Maximiza a log-verossimilhança completa em relação aos parâmetros do modelo, substituindo os dados latentes por seus valores esperados condicionais obtidos no passo E.

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\operatorname{argmax}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

Algoritmo SAEM

Algoritmo MCEM - Wei & Tanner (1990)

- **Passo E MC:**

  1. Passo de simulação: Gera $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$ da distribuição condicional $f(\mathbf{q}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k-1)})$;

  2. Passo de aproximação: Usando $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$, calcula a esperança da log-verossimilhança completa condicionada usando a aproximação,

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \frac{1}{m}\sum_{l=1}^{m} \ell_c(\boldsymbol{\theta}|\mathbf{q}^{(k,l)}, \mathbf{y}).$$

- **Passo M:** Atualiza $\boldsymbol{\theta}^{(k)}$ de acordo com

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\operatorname{argmax}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

Algoritmo SAEM

Algoritmo SAEM - Delyon *et al.* (1999)

- **Passo E MC:**

  1. **Passo de simulação:** Gera $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$ da distribuição condicional $f(\mathbf{q}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k-1)})$;

  2. **Passo de aproximação estocática:**: Atualiza $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ de acordo com

  $$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \ell_c(\boldsymbol{\theta}|\mathbf{q}^{(k,l)}, \mathbf{y}) - Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)}) \right],$$

  onde δ_k é um parâmetro de suavização, i.e., uma sequência decrescente de números positivos tal que $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$.

- **Passo M:** Atualiza $\boldsymbol{\theta}^{(k)}$ de acordo com

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\operatorname{argmax}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

Algoritmo SAEM

Algoritmo SAEM - Delyon *et al.* (1999)

- Como proposto por Galarza *et al.* (2015), nós vamos considerar o seguinte parâmetro de suavização

$$\delta_k = \begin{cases} 1, & \text{if} \quad 1 \leq k \leq cW; \\ \frac{1}{k - cW}, & \text{if} \quad cW + 1 \leq k \leq W, \end{cases} \tag{4}$$

onde,

  - W é o número máximo de iterações; e
  - c é um ponte de corte ($0 \leq c \leq 1$), que determina a percentagem das iterações iniciais.

  - Outras propostas para o parâmetro de suavização δ_k podem ser encontradas em Kuhn & Lavielle (2005), Jank (2006), entre outros.

Estrutura de correlação damped exponential (DEC)

Estrutura de correlação - Muñoz *et al.* (1992)

Estrutura de correlação damped exponential (DEC) para E_i:

$$\mathrm{E}_i = \mathrm{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij} - t_{ik}|^{\phi_2}}\right], \ i = 1, \ldots, n, \ \ j, k = 1, \ldots, n_i, \qquad (5)$$

- ϕ_1: descreve a autocorrelação entre as observações separadas pela distância absoluta entre os tempos;
- ϕ_2: avalia a aceleração do decaimento da função de autocorrelação.

Para a estrutura DEC, nós temos que:

(a) Se $\phi_2 = 0 \Rightarrow \mathrm{E}_i$ é a estrutura de correlação simétrica;

(b) Se $0 < \phi_2 < 1 \Rightarrow \mathrm{E}_i$ é a estrutura de correlação com taxa de decaimento entre a estrutura simétrica e o modelo AR de primeira ordem (AR(1));

(c) Se $\phi_2 = 1 \Rightarrow \mathrm{E}_i$ é uma estrutura de correlação AR(1);

(d) Se $\phi_2 > 1 \Rightarrow$ é uma estrutura de correlação E_i com taxa de decaimento mais rápida que a do modelo AR(1); e

(e) Se $\phi_2 \rightarrow \infty \Rightarrow \mathrm{E}_i$ é a estrutura de correlação do modelo média móvel de ordem 1 (MA(1)).

Trabalho 1

Censored mixed-effects models for irregularly observed repeated measures with applications to hiv viral loads

Autores: **Larissa A. Matos**, Luis M. Castro e Victor H. Lachos

Revista: *TEST*

Trabalho 2

Influence assessment in censored mixed-effects models using the multivariate student's-t distribution

Autores: **Larissa A. Matos**, Dipankar Bandyopadhyay, Luis M. Castro e Victor H. Lachos

Revista: *Journal of multivariate analysis*

Trabalho 3

Heavy-tailed longitudinal regression models for censored data: A likelihood based perspective

Autores: **Larissa A. Matos**, Victor H. Lachos, T-I Lin e Luis M. Castro

Revista: *TEST*

Trabalho 4

Heavy-Tailed Longitudinal Linear Mixed Models for Multiple Censored Responses Data

Autores: Victor H. Lachos, **Larissa A. Matos**, Luis M. Castro e Ming-Hui Chen

Revista: *Statistics in Medicine*

Modelo linear multivariado mistos censurado

Seja $\mathbf{Y}_i = [\mathbf{y}_{i1} : \ldots : \mathbf{y}_{ir}]$, então

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n; \tag{6}$$

em que:

- $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i) = (\mathbf{y}_{i1}^\top, \ldots, \mathbf{y}_{ir}^\top)^\top$, onde $\mathbf{y}_{ij} = (y_{ij1}, \ldots, y_{ijn_i})^\top$ é o vetor $n_i \times 1$ da j-ésima variável resposta para o indivíduo i;
- $\mathbf{X}_i = \text{Bdiag}\{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{ir}\}$, onde $\mathbf{X}_{ij}$ é a matriz de planejamento $n_i \times p_j$ dos efeitos fixos da j-ésima variável resposta;
- $\mathbf{Z}_i = \text{Bdiag}\{\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{ir}\}$, onde $\mathbf{Z}_{ij}$ é a matriz de planejamento $n_i \times q_j$ dos efeitos aleatórios da j-ésima variável resposta, geralmente $\mathbf{Z}_{ij}$ é um sub conjunto de $\mathbf{X}_{ij}$;
- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_r^\top)^\top$ é o vetor $p \times 1$ de efeitos fixos associado com a matriz de planejamento $\mathbf{X}_i$, $p = \sum_{j=1}^r p_j$;
- $\mathbf{b}_i = (\mathbf{b}_{i1}^\top, \ldots, \mathbf{b}_{ir}^\top)^\top$ é o vetor $q \times 1$ de efeitos aleatórios associado com a matriz de planejamento $\mathbf{Z}_i$, $q = \sum_{j=1}^r q_j$;
- $\boldsymbol{\epsilon}_i = \text{vec}(\mathrm{E}_i) = (\boldsymbol{\epsilon}_{i1}^\top, \ldots, \boldsymbol{\epsilon}_{ir}^\top)^\top$ é o vetor $s_i \times 1$ de erros aleatórios ($s_i = n_i \times r$), onde $\mathrm{E}_i = [\boldsymbol{\epsilon}_{i1} : \ldots : \boldsymbol{\epsilon}_{ir}]$ e $\boldsymbol{\epsilon}_{ij}$ corresponde o erro da j-ésima resposta do i-ésimo indivíduo.

Modelo linear multivariado mistos censurado

- Em vez da suposição habitual de normalidade, nós utilizamos as distribuições SMN, então, o modelo pode ser expresso como

$$\begin{aligned} \mathbf{y}_i \mid \mathbf{b}_i &\overset{\text{ind.}}{\sim} \mathrm{SMN}_{s_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i; \mathrm{H}_1), \\ \mathbf{b}_i &\overset{\text{ind.}}{\sim} \mathrm{SMN}_q(\mathbf{0}, \mathbf{D}; \mathrm{H}_2), \quad i = 1, \ldots, n. \end{aligned} \tag{7}$$

- Usando a representação estocástica (1), a representação hierárquica para o modelo definido em (6) é dada por

$$\begin{aligned} \mathbf{y}_i \mid \mathbf{b}_i, \kappa_i &\overset{\text{ind.}}{\sim} \mathrm{N}_{s_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \kappa_i^{-1}\mathbf{R}_i), \\ \mathbf{b}_i \mid \tau_i &\overset{\text{ind.}}{\sim} \mathrm{N}_q(\mathbf{0}, \tau_i^{-1}\mathbf{D}), \\ \kappa_i &\overset{\text{ind.}}{\sim} \mathrm{H}_1(\nu), \\ \tau_i &\overset{\text{ind.}}{\sim} \mathrm{H}_2(\eta); \end{aligned} \tag{8}$$

onde $\mathbf{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}_i$.

Modelo linear multivariado mistos censurado

Censuras

Para o modelo descrito, os dados observados para o $i-$ésimo indivíduo são representados por $(\mathbf{V}_i, \mathbf{C}_i)$, onde

- $\mathbf{V}_i$ é o vetor de medida não censurada e/ou nível de censura; e
- $\mathbf{C}_i$ é o vetor de indicador de censura,

de tal modo que, considerando o caso de censura a esquerda, temos que

$$\begin{aligned} y_{ijk} &\leq V_{ijk} \quad \text{se} \quad C_{ijk} = 1, \\ y_{ijk} &= V_{ijk} \quad \text{se} \quad C_{ijk} = 0, \end{aligned}$$

com $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ e $k = 1, \ldots, r$; onde $\mathbf{V}_i = [V_{i1} : \ldots : V_{ir}]$ é uma matriz $n_i \times r$ e $\mathbf{C}_i = [C_{i1} : \ldots : C_{ir}]$ é uma matriz $n_i \times r$;

- Para simplificar, os dados com censuradas à esquerda foram apresentados, mas as extensões para censura à direita ou intervalar são imediatas.

Modelo linear multivariado mistos censurado

- O procedimento de estimação do modelo proposto é derivado através da função da log-verossimilhança completa, dada por:

$$
\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^{n}\left[\log f(\mathbf{y}_i|\mathbf{b}_i,\kappa_i) + \log f(\mathbf{b}_i|\tau_i) + \log h_1(\kappa_i|\nu) + \log h_2(\tau_i|\eta)\right] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{R}_i| - \frac{1}{2}\sum_{i=1}^{n}\kappa_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\
&- \frac{1}{2}\sum_{i=1}^{n}\log|\mathbf{D}| - \frac{1}{2}\sum_{i=1}^{n}\tau_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i + \sum_{i=1}^{n}\log h_1(\kappa_i|\nu) + \sum_{i=1}^{n}\log h_2(\tau_i|\eta) + K,
\end{aligned}
$$

onde K é uma constante que não depende do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \nu, \eta)$, e $\mathbf{y}_c = (\mathbf{V}^\top, \mathbf{C}^\top, \mathbf{y}^\top, \boldsymbol{b}^\top, \boldsymbol{\kappa}^\top, \boldsymbol{\tau}^\top)^\top$ (dados aumentados).

Modelo linear multivariado mistos censurado

Estimação paramétrica via algoritmo SAEM

- Função Q: Para o *i*-ésimo indivíduo,

$$
\begin{aligned}
Q_i\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) \quad = \quad & \widehat{\ell h_{1i}}^{(k)} + \widehat{\ell h_{2i}}^{(k)} - \frac{1}{2}\log|\widehat{\mathbf{D}}^{(k)}| - \frac{1}{2}\mathrm{tr}\left(\widehat{\tau \mathbf{b}_i^2}^{(k)}\widehat{\mathbf{D}}_i^{-1(k)}\right) - \frac{1}{2}\sum_{i=1}^{n}\log|\widehat{\mathbf{R}}_i^{(k)}| \\
- \quad & \frac{1}{2}\Bigg[\mathrm{tr}\left(\widehat{\kappa\mathbf{y}_i^2}^{(k)}\widehat{\mathbf{R}}_i^{-1(k)}\right) - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\widehat{\kappa\mathbf{y}_i}^{(k)} + 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{Z}_i\widehat{\kappa\mathbf{b}_i}^{(k)} \\
- \quad & 2\mathrm{tr}\left(\mathbf{Z}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k)}\right) + \mathrm{tr}\left(\mathbf{Z}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{Z}_i\widehat{\kappa\mathbf{b}_i^2}^{(k)}\right) + \widehat{\kappa_i}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\Bigg],
\end{aligned}
$$

com

$$
\begin{aligned}
\widehat{\ell h_{1i}}^{(k)} &= E\left[\log h_1(\kappa_i|\nu)|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], & \widehat{\ell h_{2i}}^{(k)} &= E\left[\log h_2(\tau_i|\eta)|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] \\
\widehat{\kappa\mathbf{y}_i^2}^{(k)} &= E\left[\kappa_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], & \widehat{\kappa\mathbf{y}_i}^{(k)} &= E\left[\kappa_i\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], \\
\widehat{\kappa\mathbf{b}_i^2}^{(k)} &= E\left[\kappa_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], & \widehat{\kappa\mathbf{b}_i}^{(k)} &= E\left[\kappa_i\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], \\
\widehat{\tau\mathbf{b}_i^2}^{(k)} &= E\left[\tau_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], & \widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k)} &= E\left[\kappa_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], \\
\widehat{\kappa_i}^{(k)} &= E\left[\kappa_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right].
\end{aligned}
$$

Modelo linear multivariado mistos censurado

SAEM - Passo E

- **Passo de Simulação:** Amostrador de Gibbs

Passo 1. Gere $\mathbf{y}_i^c$ de $f(\mathbf{y}_i^c|\mathbf{V}_i^c, \mathbf{y}_i^o, \mathbf{b}_i^{(k,l-1)}, \kappa_i^{(k,l-1)}, \tau_i^{(k,l-1)}, \widehat{\boldsymbol{\theta}}^{(k)})$, onde

$$\mathbf{y}_i^c|\mathbf{V}_i^c, \mathbf{y}_i^o, \mathbf{b}_i, \kappa_i, \tau_i, \boldsymbol{\theta} \sim \mathrm{TN}_{s_i^c}(\boldsymbol{\mu}_i, \kappa_i^{-1}\mathbf{S}_i; \mathbb{A}_i),$$

com $\mathbb{A}_i = \{\mathbf{y}_i^c = (y_{i1}^c, \ldots, y_{is_i^c}^c)^\top | y_{i1}^c \leq V_{i1}^c, \ldots, y_{is_i^c}^c \leq V_{is_i^c}^c\}$,

$$\begin{aligned} \boldsymbol{\mu}_i &= (\mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{Z}_i^c\mathbf{b}_i) + \mathbf{R}_i^{co}(\mathbf{R}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta} - \mathbf{Z}_i^o\mathbf{b}_i) \quad \text{e} \\ \mathbf{S}_i &= \mathbf{R}_i^{cc} - \mathbf{R}_i^{co}(\mathbf{R}_i^{oo})^{-1}\mathbf{R}_i^{oc}. \end{aligned}$$

Então $\mathbf{y}_i^{(k,l)} = (y_{i1}, \ldots, y_{is_i^o}, y_{is_i^o+1}^{c(k,l)}, \ldots, y_{is_i}^{c(k,l)})$ é amostra gerada.

Passo 2. Gere $\mathbf{b}_i^{(k,l)}$ from $f(\mathbf{b}_i|\mathbf{y}_i^{(k,l)}, \kappa_i^{(k,l-1)}, \tau_i^{(k,l-1)}, \widehat{\boldsymbol{\theta}}^{(k)})$, onde

$$\mathbf{b}_i|\mathbf{y}_i, \kappa_i, \tau_i \sim \mathrm{N}_q(\boldsymbol{\Psi}_i\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\kappa_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \boldsymbol{\Psi}_i),$$

com $\boldsymbol{\Psi} = (\kappa_i\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i + \tau_i\mathbf{D}^{-1})^{-1}$ (Arellano-Valle *et al.*, 2005, Lemma 2).

Modelo de regressão não linear censurado

SAEM - Passo E

Passo 3. Gere $\kappa_i^{(k,l)}$ de $f(\kappa_i|\mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \tau_i^{(k,l-1)}, \widehat{\boldsymbol{\theta}}^{(k)})$.

Passo 4. Gere $\tau_i^{(k,l)}$ de $f(\tau_i|\mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \kappa_i^{(k,l)}, \widehat{\boldsymbol{\theta}}^{(k)})$.

Observação: Note que, dado $\mathbf{y}_i \mid \mathbf{b}_i$ independente de τ_i; $\mathbf{b}_i$ independente de κ_i; e κ_i e τ_i mutuamente independente, então temos que

$$f(\kappa_i|\mathbf{y}_i, \mathbf{b}_i, \tau_i) \propto f(\mathbf{y}_i|\mathbf{b}_i, \kappa_i) f(\kappa_i)$$

e

$$f(\tau_i|\mathbf{y}_i, \mathbf{b}_i, \kappa_i) \propto f(\mathbf{b}_i|\tau_i) f(\tau_i).$$

Modelo de regressão não linear censurado

SAEM - Passo E

Distribuição de $\boldsymbol{\epsilon}_i$	Distribuição de κ_i	Distribuição de $\kappa_i \mid \mathbf{y}_i, \mathbf{b}_i, \tau_i$
$\mathrm{T}_{si}(\mathbf{0}, \mathbf{R}_i, \nu)$	$\mathrm{Gamma}(\nu/2, \nu/2)$	$\mathrm{Gamma}\left((\nu + s_i)/2, (D^2_{e_i} + \nu)/2\right)$
$\mathrm{SL}_{si}(\mathbf{0}, \mathbf{R}_i, \nu)$	$\mathrm{Beta}(\nu, 1)$	$\mathrm{TGamma}\left(\nu + s_i/2, D^2_{e_i}/2, 1\right)$
$\mathrm{CN}_{si}(\mathbf{0}, \mathbf{R}_i, \nu_1, \nu_2)$	$\nu_1 \mathbb{I}_{\{\nu_2\}}(\kappa_i) + (1 - \nu_1)\mathbb{I}_{\{1\}}(\kappa_i)$	$P(\kappa_i = \nu_2) = 1 - P(\kappa_i = 1) = p_1/p_1 + p_2$ $p_1 = \nu_1 \nu_2^{s_i/2} exp\{-\frac{1}{2} D^2_{e_i} \nu_2\}$ $p_2 = (1 - \nu_1) exp\{-\frac{1}{2} D^2_{e_i}\}$

$D^2_{e_i} = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)$

Distribuição de $\mathbf{b}_i$	Distribuição de τ_i	Distribuição de $\tau_i \mid \mathbf{y}_i, \mathbf{b}_i, \kappa_i$
$\mathrm{T}_q(\mathbf{0}, \mathbf{D}, \eta)$	$\mathrm{Gamma}(\eta/2, \eta/2)$	$\mathrm{Gamma}\left((\eta + q)/2, (D^2_{\mathbf{b}_i} + \eta)/2\right)$
$\mathrm{SL}_q(\mathbf{0}, \mathbf{D}, \eta)$	$\mathrm{Beta}(\eta, 1)$	$\mathrm{TGamma}\left(\eta + q/2, D^2_{\mathbf{b}_i}/2, 1\right)$
$\mathrm{CN}_q(\mathbf{0}, \mathbf{D}, \eta_1, \eta_2)$	$\eta_1 \mathbb{I}_{\{\eta_2\}}(\tau_i) + (1 - \eta_1)\mathbb{I}_{\{1\}}(\tau_i)$	$P(\tau_i = \eta_2) = 1 - P(\tau_i = 1) = q_1/q_1 + q_2$ $q_1 = \eta_1 \eta_2^{q/2} exp\{-\frac{1}{2} D^2_{\mathbf{b}_i} \eta_2\}$ $q_2 = (1 - \eta_1) exp\{-\frac{1}{2} D^2_{\mathbf{b}_i}\}$

$D^2_{\mathbf{b}_i} = \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i$

Modelo linear multivariado misto censurado

SAEM - Passo E

- **Passo de Aproximação:** $(\mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \kappa_i^{(k,l)}, \tau_i^{(k,l)})$, $l = 1, \ldots, m$:

$$\begin{aligned}
\widehat{\kappa\mathbf{y}_i^2}^{(k)} &= \widehat{\kappa\mathbf{y}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} \mathbf{y}_i^{(k,l)\top} - \widehat{\kappa\mathbf{y}_i^2}^{(k-1)} \right), \\
\widehat{\kappa\mathbf{y}_i}^{(k)} &= \widehat{\kappa\mathbf{y}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} - \widehat{\kappa\mathbf{y}_i}^{(k-1)} \right), \\
\widehat{\kappa\mathbf{b}_i^2}^{(k)} &= \widehat{\kappa\mathbf{b}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{b}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\kappa\mathbf{b}_i^2}^{(k-1)} \right), \\
\widehat{\kappa\mathbf{b}_i}^{(k)} &= \widehat{\kappa\mathbf{b}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{b}_i^{(k,l)} - \widehat{\kappa\mathbf{b}_i}^{(k-1)} \right), \\
\widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k)} &= \widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k-1)} \right), \\
\widehat{\tau\mathbf{b}_i^2}^{(k)} &= \widehat{\tau\mathbf{b}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \tau_i^{(k,l)} \mathbf{b}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\tau\mathbf{y}_i^2}^{(k-1)} \right), \\
\widehat{\kappa_i}^{(k)} &= \widehat{\kappa_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} - \widehat{\kappa_i}^{(k-1)} \right), \\
\widehat{\ell h_{1i}}^{(k)} &= \widehat{\ell h_{1i}}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \log h_1(\kappa_i^{(k,l)} | \widehat{\nu}^{(k-1)}) - \widehat{\ell h_{1i}}^{(k-1)} \right), \\
\widehat{\ell h_{2i}}^{(k)} &= \widehat{\ell h_{2i}}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \log h_2(\tau_i^{(k,l)} | \widehat{\eta}^{(k-1)}) - \widehat{\ell h_{2i}}^{(k-1)} \right).
\end{aligned}$$

Modelo linear multivariado misto censurado

SAEM - Passo CM

Atualizamos $\widehat{\boldsymbol{\theta}}^{(k)}$ pela maximização de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, o que leva as seguintes expressões:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \widehat{\kappa}_i^{(k)} \mathbf{X}_i^{\top} \widehat{\mathbf{R}}_i^{-1(k)} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^{\top} \widehat{\mathbf{R}}_i^{-1(k)} \left(\widehat{\kappa \mathbf{y}_i}^{(k)} - Z_i \widehat{\kappa \mathbf{b}_i}^{(k)}\right), \\
\widehat{\sigma_{jl}^2}^{(k+1)} &= \begin{cases} \left(\sum_{i=1}^{n} n_i\right)^{-1} \sum_{i=1}^{n} \operatorname{tr}\left(\widehat{\boldsymbol{\Omega}}_i^{-1(k)} \widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)}\right) & \text{for } j = l, \\ \left(2\sum_{i=1}^{n} n_i\right)^{-1} \sum_{i=1}^{n} \operatorname{tr}\left[\widehat{\boldsymbol{\Omega}}_i^{-1(k)} \left(\widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)} + \widehat{\kappa \boldsymbol{\epsilon}}_{ilj}^{(k)}\right)\right] & \text{for } j \neq l, \end{cases} \\
\widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi} \in (0,1) \times \mathbb{R}^+}{\operatorname{argmax}} \left\{ -\frac{r}{2} \sum_{i=1}^{n} \log |\boldsymbol{\Omega}_i(\boldsymbol{\phi}, \mathbf{t}_i)| - \frac{1}{2} \sum_{i=1}^{n} \operatorname{tr}\left[\left(\widehat{\boldsymbol{\Sigma}}^{(k)} \otimes \boldsymbol{\Omega}_i(\boldsymbol{\phi}, \mathbf{t}_i)\right)^{-1} \widehat{\kappa \mathrm{E}_i}\right] \right\}, \\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau \mathbf{b}_i^2}^{(k)}, \\
\widehat{\nu}^{(k+1)} &= \underset{\nu}{\operatorname{argmax}} \sum_{i=1}^{n} \widehat{\ell h_{1i}}^{(k)}(\nu), \\
\widehat{\eta}^{(k+1)} &= \underset{\eta}{\operatorname{argmax}} \sum_{i=1}^{n} \widehat{\ell h_{2i}}^{(k)}(\eta).
\end{aligned}
$$

Modelo linear multivariado misto censurado

Função de verossimilhança

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) = \prod_{i=1}^{n} \int \left[\int_0^{\infty} f(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) h_1(\kappa_i) d\kappa_i \right] f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i.$$

Particionando $\mathbf{y}_i$, temos que

$$\begin{aligned} L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) &= \prod_{i=1}^{n} \int \left[\int_0^{\infty} \phi_{s_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^c \boldsymbol{\beta} - \mathbf{Z}_i^c \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i^{oo}) \Phi_{s_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa_i^{-1} \mathbf{S}_i) h_1(\kappa_i) d\kappa_i \right] \\ &\times \quad f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i = \prod_{i=1}^{n} \int g(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i \end{aligned} \tag{9}$$

onde $g(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) = \int_0^{\infty} \phi_{s_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^c \boldsymbol{\beta} - \mathbf{Z}_i^c \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i^{oo}) \Phi_{s_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa_i^{-1} \mathbf{S}_i) h_1(\kappa_i|\nu) d\kappa_i$.
Calculamos a integral em (9) usando o método de amostragem por importância. Então, temos que

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) = \prod_{i=1}^{n} \int g(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) \frac{f(\mathbf{b}_i|\boldsymbol{\theta})}{f^{\star}(\mathbf{b}_i|\boldsymbol{\theta})} d\mathbf{b}_i,$$

onde $f^{\star}$ é a função de importância. Consequentemente, $L_o(\boldsymbol{\theta}; \mathbf{y}_i^{obs})$ é estimada através da seguinte aproximação

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) = \prod_{i=1}^{n} \left[\frac{1}{M} \sum_{m=1}^{M} g(\mathbf{y}_i|\mathbf{b}_{im}, \kappa_i; \boldsymbol{\theta}) \frac{f(\mathbf{b}_{im}|\boldsymbol{\theta})}{f^{\star}(\mathbf{b}_{im}|\boldsymbol{\theta})} \right],$$

com $\mathbf{b}_{i1}, \ldots, \mathbf{b}_{im}$ gerado de $f^{\star}(\mathbf{b}_i|\boldsymbol{\theta})$.

Modelo linear multivariado misto censurado

Critérios para seleção do modelo

- **AIC e BIC**

$$\mathrm{AIC} = 2\,m - 2\,\ell_{max} \text{ e } \mathrm{BIC} = m\,log\,N - 2\,\ell_{max}.$$

- **Decomposição AIC e BIC (Zhang et al., 2014)**

Seja $\mathbf{y}_{i1}^{\star} = (\mathbf{y}_{i1}^{\top}, \ldots, \mathbf{y}_{ir^{\star}}^{\top})^{\top}$ e $\mathbf{y}_{i2}^{\star} = (\mathbf{y}_{ir^{\star}+1}^{\top}, \ldots, \mathbf{y}_{ir}^{\top})^{\top}$, onde $\mathbf{y}_i = (\mathbf{y}_{i1}^{\star\top}, \mathbf{y}_{i2}^{\star\top})^{\top}$ e $r^{\star} \in \{1, \ldots, r\}$, então podemos decompor o AIC e BIC da seguinte forma

$$\mathrm{AIC} = \mathrm{AIC}_{\mathbf{y}_1^{\star}} + \mathrm{AIC}_{\mathbf{y}_2^{\star}|\mathbf{y}_1^{\star}} \text{ e } \mathrm{BIC} = \mathrm{BIC}_{\mathbf{y}_1^{\star}} + \mathrm{BIC}_{\mathbf{y}_2^{\star}|\mathbf{y}_1^{\star}}.$$

- **Medida para medir o ganho da modelagem conjunta**

$$\Delta\mathrm{AIC} = \mathrm{AIC}_{\mathbf{y}_{2,0}^{\star}} - \mathrm{AIC}_{\mathbf{y}_2^{\star}|\mathbf{y}_1^{\star}} \text{ and } \Delta\mathrm{BIC} = \mathrm{BIC}_{\mathbf{y}_{2,0}^{\star}} - \mathrm{BIC}_{\mathbf{y}_2^{\star}|\mathbf{y}_1^{\star}}.$$

Modelo linear multivariado misto censurado

Estudo - A5055

$$
\begin{aligned}
y_{i1k} &= \beta_{10} + \beta_{11} t_{ik} + \beta_{12}\text{treat}_i + \beta_{13} t_{ik}^{0.5} + \beta_{14}\text{treat}_i \times t_{ik} + b_{i10} + b_{i11} t_{ik} + e_{i1k},\\
y_{i2k} &= \beta_{20} + \beta_{21} t_{ik} + \beta_{22}\text{treat}_i + \beta_{23}\text{treat}_i \times t_{ik} + b_{i20} + b_{i21} t_{ik} + e_{i2k},
\end{aligned}
$$

i=1,...,44,

- y_{i1k} é a resposta $\log_{10}$ (RNA) para o individuo i no tempo t_k;
- y_{i2k} é a resposta log(CD4/CD8) para o individuo i no tempo t_k;
- 316 observações;
- 33% das observações censuradas à esquerda;
- $t_{ik} = \text{dia}_{ik}/7$ (semana), para $k = 1, \ldots, s_i$, onde dia: 0, 7, 14, 28, 56, 84, 112, 140 e 168;
- treat_i é o indicador de tratamento ($=0$ para tratamento 1; $=1$ para tratamento 2);
- b_{ij0} e b_{ij1} são os interceptos aleatórios e as inclinações aleatórias, respectivamente, para y_{ijk}, $j = 1, 2$.

- Esse conjunto de dados foi analisado anteriormente por Wang *et al.* (2015).

Modelo linear multivariado misto censurado

Estudo - A5055

Critérios para seleção do modelo para os modelos *SMN-MLMEC* sob estrutura DEC:

	Distribuição ϵ / Distribuição $\mathbf{b}$								
	N/N	SL/N	T/N	N/SL	N/T	SL/SL	SL/T	T/SL	T/T
AIC	789.85	742.18	739.59	791.98	792.29	744.47	744.54	741.85	741.51
BIC	896.62	853.41	850.81	903.20	903.51	860.14	860.21	857.52	857.19

Modelo linear multivariado misto censurado

Estudo - A5055

Estimativas dos parâmetros usando distribuições T/N para a estrutura DEC e UNC:

Estrutura	Parâmetros	Estimativa (SE)	Parâmetros	Estimativa (SE)	Parâmetros	Estimativa (SE)
DEC	β_{10}	3.743 (0.134)	d_{11}	0.1446 (0.0829)	σ_{11}	0.409 (0.076)
	β_{11}	0.130 (0.026)	d_{21}	0.0011 (0.0133)	σ_{21}	-0.039 (0.020)
	β_{12}	-0.005 (0.067)	d_{22}	-0.0884 (0.1182)	σ_{22}	0.050 (0.011)
	β_{13}	-0.957 (0.098)	d_{31}	-0.0011 (0.0033)	ϕ_1	0.704 (0.065)
	β_{14}	-0.007 (0.025)	d_{32}	0.0034 (0.0027)	ϕ_2	0.632 (0.131)
	β_{20}	-1.284 (0.077)	d_{33}	-0.0122 (0.0116)	ν	4.737 (0.003)
	β_{21}	0.005 (0.005)	d_{41}	-0.0004 (0.0004)		
	β_{22}	0.252 (0.084)	d_{42}	0.2727 (0.0861)		
	β_{23}	-0.003 (0.007)	d_{43}	0.0008 (0.0015)		
			d_{44}	0.0001 (0.0001)		
	loglik	-344.79	AIC	739.59	BIC	850.81
UNC	β_{10}	3.718 (0.135)	d_{11}	0.4089 (0.1463)	σ_{11}	0.263 (0.053)
	β_{11}	0.129 (0.026)	d_{21}	-0.0112 (0.0153)	σ_{21}	-0.024 (0.012)
	β_{12}	0.003 (0.091)	d_{22}	-0.0964 (0.1251)	σ_{22}	0.028 (0.005)
	β_{13}	-0.955 (0.075)	d_{31}	0.0002 (0.0030)	ν	4.340 (0.004)
	β_{14}	-0.008 (0.027)	d_{32}	0.0054 (0.0029)		
	β_{20}	-1.278 (0.076)	d_{33}	-0.0132 (0.0116)		
	β_{21}	0.005 (0.004)	d_{41}	-0.0006 (0.0004)		
	β_{22}	0.286 (0.081)	d_{42}	0.2953 (0.0785)		
	β_{23}	-0.006 (0.006)	d_{43}	0.0002 (0.0015)		
			d_{44}	0.0001 (0.0001)		
	loglik	-357.97	AIC	761.94	BIC	864.26

Modelo linear multivariado misto censurado

Estudo - A5055

Critérios para seleção do modelo. Decomposição do AIC e BIC para o melhor modelo:

AIC	739.59	BIC	850.81
$\text{AIC}_{\mathbf{y}_2^\star \mid \mathbf{y}_1^\star}$	92.65	$\text{BIC}_{\mathbf{y}_2^\star \mid \mathbf{y}_1^\star}$	158.80
$\text{AIC}_{\mathbf{y}_{2,0}^\star}$	125.26	$\text{BIC}_{\mathbf{y}_{2,0}^\star}$	166.58
ΔAIC	32.61	ΔBIC	7.77

Trabalho 5

Multivariate Measurement Error Models Based on Student-t Distribution under Censored Responses

Autores: **Larissa A. Matos**, Luis M. Castro, Celso R. B. Cabral e Victor H. Lachos

Revista: *Statistics*

Referências I

Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B*, pages 99–102.

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. (2005). Skew-normal linear mixed models. *Journal of Data Science*, **3**, 415–438.

Delyon, B., Lavielle, M. & Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of Statistics*, pages 94–128.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,*, **39**, 1–38.

Garay, A. M., Castro, L. M., Leskow, J. & Lachos, V. H. (2014). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, page DOI: 10.1177/0962280214551191.

Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics*, **15**(4), 803–829.

Kuhn, E. & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, **49**(4), 1020–1038.

Lachos, V. H., Bandyopadhyay, D. & Dey, D. K. (2011). Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. *Biometrics*, **67**, 1594–1604.

Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.

Matos, L., Lachos, V., Balakrishnan, N. & Labra, F. (2013a). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistical & Data Analysis*, **57**(1), 450–464.

Matos, L., Prates, M., Chen, M.-H. & Lachos, V. (2013b). Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**, 1323–1345.

Muñoz, A., Carey, V., Schouten, J. P., Segal, M. & Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, **48**, 733–742.

Samson, A., Lavielle, M. & Mentré, F. (2006). Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Computational Statistics & Data Analysis*, **51**(3), 1562–1574.

Referências II

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**(4), 797–817.

Vaida, F., Fitzgerald, A. & DeGruttola, V. (2007). Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational Statistics & Data Analysis*, **51**(12), 5718–5730.

Wang, W.-L., Lin, T.-I. & Lachos, V. H. (2015). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research*, page doi: 0962280215620229.

Wei, G. C. & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**(411), 699–704.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., Wang, P. & Shen, W. (2014). Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Statistics in Medicine*, **33**(27), 4715–4733.