

Inferência Bayesiana em Modelos Assimétricos

Márcia D'Elia Branco

Universidade de São Paulo
Instituto de Matemática e Estatística

Coloquio Inter-Institucional - UFRJ - 22/10/2008

Introdução

Objetivos

Distribuições sob seleção

Casos especiais: normal e t assimetrizadas

Inferência Bayesiana/MCCM

Modelos Lineares Mistos

Modelos binários

Roteiro

1 Introdução

Introdução

Objetivos

Distribuições sob seleção

Casos especiais: normal e t assimetrizadas

Inferência Bayesiana/MCCM

Modelos Lineares Mistos

Modelos binários

Roteiro

- 1 Introdução
- 2 Objetivos

Roteiro

- 1 Introdução
- 2 Objetivos
- 3 Distribuições sob seleção

Roteiro

- 1 Introdução
- 2 Objetivos
- 3 Distribuições sob seleção
- 4 Casos especiais: normal e t assimetrizadas

Roteiro

- 1 Introdução
- 2 Objetivos
- 3 Distribuições sob seleção
- 4 Casos especiais: normal e t assimetrizadas
- 5 Inferência Bayesiana/MCCM

Roteiro

- 1 Introdução
- 2 Objetivos
- 3 Distribuições sob seleção
- 4 Casos especiais: normal e t assimetrizadas
- 5 Inferência Bayesiana/MCCM
- 6 Modelos Lineares Mistos

Roteiro

- 1 Introdução
- 2 Objetivos
- 3 Distribuições sob seleção
- 4 Casos especiais: normal e t assimetrizadas
- 5 Inferência Bayesiana/MCCM
- 6 Modelos Lineares Mistos
- 7 Modelos binários

Preliminares

$X \in \mathbb{R}^d$ vetor aleatório, abs. contínuo.

Qual distribuição de probabilidade podemos utilizar para X ?

Preliminares

$X \in \mathbb{R}^d$ vetor aleatório, abs. contínuo.

Qual distribuição de probabilidade podemos utilizar para X ?

1) Normal multivariada: $X \sim N_d(\mu, \Sigma)$.

$$f(x \mid \mu, \Sigma) = |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right\}$$

$\mu \in \mathbb{R}^d$ e Σ matrix $d \times d$ definida positiva,
com $E[X] = \mu$ e $Cov[X] = \Sigma$.

Propriedades:

- Fechada por marginalização, condicionamento e c.linear
- $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2) \Leftrightarrow X_1, \dots, X_d$ são independentes.
- $(x - \mu)^t \Sigma^{-1} (x - \mu) \sim \chi_d^2$

2) T-multivariada: $X \sim T_d(\mu, \Sigma; \nu)$.

$$f(x) = \frac{\Gamma[(\nu + d)/2]}{\Gamma[\nu/2](\pi\nu)^{d/2}} |\Sigma|^{-1/2} [\nu + (x - \mu)^t \Sigma^{-1} (x - \mu)]^{-\frac{\nu+d}{2}}.$$

Propriedades:

- Fechada por marginalização e combinação linear.
- $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2) \Leftrightarrow \text{Cov}(X_i, X_j) = 0 \nRightarrow$ independentes.
- $(x - \mu)^t \Sigma^{-1} (x - \mu) \sim F - \text{Snedecor}$

2) T-multivariada: $X \sim T_d(\mu, \Sigma; \nu)$.

$$f(x) = \frac{\Gamma[(\nu + d)/2]}{\Gamma[\nu/2](\pi\nu)^{d/2}} |\Sigma|^{-1/2} [\nu + (x - \mu)^t \Sigma^{-1} (x - \mu)]^{-\frac{\nu+d}{2}}.$$

Propriedades:

- Fechada por marginalização e combinação linear.
- $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2) \Leftrightarrow \text{Cov}(X_i, X_j) = 0 \nRightarrow$ independentes.
- $(x - \mu)^t \Sigma^{-1} (x - \mu) \sim F - \text{Snedecor}$

3) Mistura de normais no parâmetro de escala.

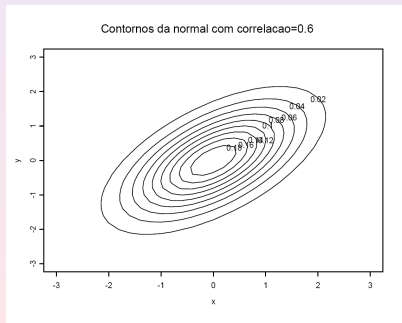
$$f(x) = |\Sigma|^{-1/2} \int_0^\infty \exp \left\{ -\frac{1}{2} (x - \mu)^t \frac{\Sigma^{-1}}{w} (x - \mu) \right\} dG(w)$$

3) Elíptica: $X \sim El_d(\mu, \Sigma; g^{(d)})$.

$$f(x | \mu, \Sigma) = |\Sigma|^{-1/2} g \left(-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right)$$

- A função geradora $g^{(d)}$ satisfaz algumas condições para f ser fdp .
- Σ não é a matrix de covariâncias.
- Simetria.
- Curtoses diferentes da normal (caudas pesadas e caudas leves).

Gráfico de contornos para elíptica bivarida



Exemplo 1: $(X, Y) \sim N_2(\mu, \Sigma)$

X peso dos indivíduos adultos do sexo masculino.

Y altura dos indivíduos adultos do sexo masculino.

Amostra: Foi medido o peso e altura de n jogadores de basquete.

Não é uma amostra aleatória dessa população. É uma amostra especial (seleta).

É razoável supor que $Y > \mu_y$. O que observamos é um valor do seguinte v.a.

$$(X^*, Y^*) = (X, Y \mid Y > \mu_y)$$

Essa tem distribuição assimétrica.

Objetivos

- Propor distribuições de probabilidades multivariadas que acomodem assimetria, caudas pesadas e multimodalidade.

Objetivos

- Propor distribuições de probabilidades multivariadas que acomodem assimetria, caudas pesadas e multimodalidade.
- Flexibilizar a suposição de normalidade usualmente considerada em modelos lineares e não lineares.

Objetivos

- Propor distribuições de probabilidades multivariadas que acomodem assimetria, caudas pesadas e multimodalidade.
- Flexibilizar a suposição de normalidade usualmente considerada em modelos lineares e não lineares.
- Manter a estrutura paramétrica do modelo com parâmetros interpretáveis.

Objetivos

- Propor distribuições de probabilidades multivariadas que acomodem assimetria, caudas pesadas e multimodalidade.
- Flexibilizar a suposição de normalidade usualmente considerada em modelos lineares e não lineares.
- Manter a estrutura paramétrica do modelo com parâmetros interpretáveis.
- Evitar transformações de dados.

Objetivos

- Propor distribuições de probabilidades multivariadas que acomodem assimetria, caudas pesadas e multimodalidade.
- Flexibilizar a suposição de normalidade usualmente considerada em modelos lineares e não lineares.
- Manter a estrutura paramétrica do modelo com parâmetros interpretáveis.
- Evitar transformações de dados.
- Relativa facilidade no processo inferencial.

Distribuições de probabilidade induzidas por processos de seleção

Definição: Sejam $\mathbf{U} \in \mathbb{R}^q$ e $\mathbf{V} \in \mathbb{R}^p$ dois vetores aleatórios, e denotamos por C um conjunto mensurável em \mathbb{R}^q . Nós definimos como distribuição sob seleção a distribuição condicional $(\mathbf{V}|\mathbf{U} \in C)$. Então, um vetor aleatório $\mathbf{X} \in \mathbb{R}^p$ tem distribuição sob seleção se $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$. Usamos a notação $\mathbf{X} \sim SLCT_{p,q}$ cujos parâmetros dependem das caracterizações de \mathbf{U} , \mathbf{V} , e C . Se \mathbf{V} tem fdp $f_{\mathbf{V}}$ então \mathbf{X} tem fdp

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x}) \frac{P(\mathbf{U} \in C | \mathbf{V} = \mathbf{x})}{P(\mathbf{U} \in C)},$$

- Essa definição unifica as diversas distribuições assimetrizadas apresentadas na literatura nos últimos anos (Arellano-Valle, Branco and Genton, 2006).

- Essa definição unifica as diversas distribuições assimetrizadas apresentadas na literatura nos últimos anos (Arellano-Valle, Branco and Genton, 2006).
- O conjunto de seleção comumente empregado é

$$C(\beta) = \{\mathbf{u} \in \mathbb{R}^q | \mathbf{u} > \beta\}.$$

- Essa definição unifica as diversas distribuições assimetrizadas apresentadas na literatura nos últimos anos (Arellano-Valle, Branco and Genton, 2006).
- O conjunto de seleção comumente empregado é

$$C(\beta) = \{\mathbf{u} \in \mathbb{R}^q | \mathbf{u} > \beta\}.$$

- A classe proposta por Sahu, Dey e Branco(2003) considera $C(0)$ e distribuição conjunta elíptica para U e V .

Elíptica assimetrizada de Sahu, Dey and Branco (2003)

Seja

$$\mathbf{V} = \mathbf{D}\mathbf{U} + \mathbf{Z} \quad (1)$$

onde $\mathbf{D} = \text{diag}(\delta_1, \delta_2, \dots, \delta_d)$ e

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{Z} \end{pmatrix} \sim El_{2d} \left(\begin{pmatrix} \mathbf{0} \\ \mu \end{pmatrix}, \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \Sigma \end{pmatrix} \right). \quad (2)$$

$\mathbf{X} = [\mathbf{V} \mid \mathbf{U} > 0]$ tem distribuição elíptica-assimétrica e fdp

$$2^d f_{g^{(d)}}(x) F_{g_{q(x)}^{(d)}} \left([\mathbf{I}_d - \mathbf{D}(\Sigma + \mathbf{D}^2)^{-1}\mathbf{D}]^{-1/2} \mathbf{D}(\Sigma + \mathbf{D})^{-1}\mathbf{x} \right) \quad (3)$$

Normal assimetrizada de SDB

A distribuição normal-assimétrica é obtida fazendo $g(u) = e^{u/2}$ e sua fdp dada por

$$2^d |\Sigma + \mathbf{D}^2|^{-1/2} \phi_d((\mathbf{y} - \mu)(\Sigma + \mathbf{D}^2)^{-1/2}) \\ \times \Phi_d \left([\mathbf{I}_d - \mathbf{D}(\Sigma + \mathbf{D}^2)^{-1}\mathbf{D}]^{-1/2} \mathbf{D}(\Sigma + \mathbf{D})^{-1} \mathbf{y}_* \right)$$

Normal assimetrizada de SDB

A distribuição normal-assimétrica é obtida fazendo $g(u) = e^{u/2}$ e sua fdp dada por

$$2^d |\Sigma + \mathbf{D}^2|^{-1/2} \phi_d((\mathbf{y} - \mu)(\Sigma + \mathbf{D}^2)^{-1/2}) \\ \times \Phi_d \left([\mathbf{I}_d - \mathbf{D}(\Sigma + \mathbf{D}^2)^{-1}\mathbf{D}]^{-1/2} \mathbf{D}(\Sigma + \mathbf{D})^{-1} \mathbf{y}_* \right)$$

Propriedades

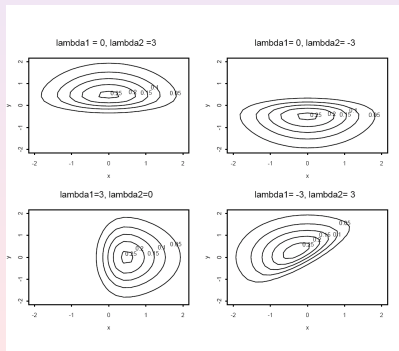
- Fechada por marginalização e combinação linear.
- $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2) \Leftrightarrow X_1, \dots, X_d$ independentes.
- $E[X] = \mu + \left(\frac{2}{\pi}\right)^{1/2} \delta$.
- $Cov[X] = \Sigma + \left(1 - \frac{2}{\pi}\right) D^2$

- Essa densidade coincide com a normal-assimétrica de Azzalini, somente quando $d = 1$.

- Essa densidade coincide com a normal-assimétrica de Azzalini, somente quando $d = 1$.
- Extensões desta densidade podem ser vistas em Arellano-Valle and Genton(2005). A SN de SDB é a única onde a independência é caracterizada pela diagonalidade de Σ .

- Essa densidade coincide com a normal-assimétrica de Azzalini, somente quando $d = 1$.
- Extensões desta densidade podem ser vistas em Arellano-Valle and Genton(2005). A SN de SDB é a única onde a independência é caracterizada pela diagonalidade de Σ .
- Os parâmetros δ_i estão associados a assimetria em cada uma das dimensões. $\delta_i = 0$ representa simetria. Se $\delta_i < 0$ ($\delta_i > 0$) representa assimetria negativa (positiva) na i -ésima dimensão. Isso não é verdade para SN de Azzalini.

Gráfico de contornos para normal-assimétrica



t-assimetrizada de SDB

Sua fdp é dada por

$$2^d t_d(x \mid \mu, \Sigma + D^2, \nu) P(U > 0 \mid V = x)$$

A última probabilidade é calculada em relação a distribuição t com $\nu + d$ graus de liberdades.

t-assimetrizada de SDB

Sua fdp é dada por

$$2^d t_d(x \mid \mu, \Sigma + D^2, \nu) P(U > 0 \mid V = x)$$

A última probabilidade é calculada em relação a distribuição t com $\nu + d$ graus de liberdades.

- Flexibiliza a assimetria e curtose
- É fechada por marginalização e combinações lineares
- $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2) \not\Rightarrow \text{Cov}(X_i, X_j) = 0$

Representações estocásticas

- $X \sim SN(\mu, \Sigma, D)$ e $W \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ então

$$Y = \frac{X}{\sqrt{W}}$$

tem distribuição $ST(\mu, \Sigma, D, \nu)$.

Representações estocásticas

- $X \sim SN(\mu, \Sigma, D)$ e $W \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ então

$$Y = \frac{X}{\sqrt{W}}$$

tem distribuição $ST(\mu, \Sigma, D, \nu)$.

- U e V tem distribuição $T(0, \Sigma, \nu)$ então

$$Y = \mu + D|U| + V$$

tem distribuição $ST(\mu, \Sigma, D, \nu)$.

Representação Hierárquica

$$Y | Z, W \sim N \left(\mu + D \frac{|Z|}{\sqrt{W}}, \frac{\Sigma}{W} \right)$$

$$Z \sim N(0, I) \quad \text{e} \quad W \sim \text{Gamma} \left(\frac{\nu}{2}, \frac{\nu}{2} \right)$$

Inferência estatística

- Dificuldades em trabalhar diretamente com a função de verossimilhança.
- Teoria assintótica de máxima verossimilhança. Questões: singularidade da matriz de informação de Fisher, lenta convergência para normalidade. Solução: reparametrização. Uso do algoritmo EM.
- Abordagem Bayesiana. Questões: algoritmos de simulação (MCCM), escolha de prioris, análise de sensibilidade.

Exemplo simples: $X = \psi + \tau Z$, com $Z \sim SN(\alpha)$, onde

$$f(z) = 2\phi(z)\Phi(\delta z)$$

$$\delta = \frac{\alpha}{\sqrt{1+\alpha}} \in [-1, 1].$$

- $E[Z] = \delta\sqrt{\frac{2}{\pi}}$
- $V[Z] = 1 - \frac{2}{\pi}\delta^2$
- $\delta = 0 \rightarrow Z \sim N(0, 1)$
- $\delta > 0 (\delta < 0)$ assimetria positiva (negativa)
- unimodal

a) Maxima verossimilhança: Singularidade da IF, viés e não existência do emv para α .

Reparametrização (CP): $(\psi, \tau^2, \alpha) \rightarrow (\theta_1, \theta_2, \gamma_1)$.

$$X = \theta_1 + \theta_2 \left(\frac{Z - E[Z]}{V[Z]} \right)$$

$$\gamma_1 = \frac{1}{2}(4 - \pi) \text{sinal}(\alpha) \left(\frac{\alpha^2}{\frac{\pi}{2} + (\frac{\pi}{2} - 1)\alpha^2} \right)^{3/2}$$

onde γ_1 é o coeficiente de assimetria da SN.

Sob a parametrização CP o problema de singularidade da matriz é resolvido. No entanto, o problema de viés do emv para α permanece.

IB e MCCM

Toda inferência é baseada na distribuição a posteriori.

$$f(\theta | x) = \frac{L(\theta; x)f(\theta)}{C}$$

IB e MCCM

Toda inferência é baseada na distribuição a posteriori.

$$f(\theta | x) = \frac{L(\theta; x)f(\theta)}{C}$$

Teoria assintótica × métodos de simulação.

- Na primeira temos que aumentar o tamanho da nossa amostra (n) para garantir o bom funcionamento da teoria. No segundo o trabalho consiste em aumentar o tamanho da amostra simulada (um problema computacional).
- Se pudermos simular uma amostra da distribuição a posteriori, obtemos uma boa aproximação dessa distribuição e também dos seus momentos e quantis (Monte Carlo).

- A dificuldade de simular de θ aumenta com o aumento da sua dimensão.
- Uma alternativa para esse problema de dimensionalidade é o uso das distribuições condicionais completas (menor dimensão). Obtemos um algoritmo aproximado baseado em Cadeias de Markov (MCCM).

- A dificuldade de simular de θ aumenta com o aumento da sua dimensão.
- Uma alternativa para esse problema de dimensionalidade é o uso das distribuições condicionais completas (menor dimensão). Obtemos um algoritmo aproximado baseado em Cadeias de Markov (MCCM).
- *"The impact of Gibbs sampling and MCMC was to, almost instantaneously, change our entire methods of thinking and attacking problems, representing a paradigm shift in the words of the historian of science Thomas Kuhn. "'* by Robert and Casella, 2008.

Exemplo simples (continua):

b) Inferência Bayesiana objetiva (Bayes and Branco, 2007, BJPS)

- Representação Hierárquica da SN:

$$X | \psi, \tau, \delta, W \sim N(\psi + \tau\delta W, \tau^2(1 - \delta^2))$$

$$W \sim NT(0, 1)I_{(0, \infty)}$$

- Especificações a priori: $f(\psi, \tau, \alpha) \propto \frac{1}{\tau} f(\alpha)$
- A distribuição à priori de Jeffreys para α é muito bem aproximada por uma t-Student centrada em zero com escala $\frac{\pi}{2}$ e 0.5 graus de liberdades.
- Problemas de existência da média a posteriori sob priori de Jeffreys.

- As distribuições a posteriori condicionais completas podem ser obtidas usando uma parametrização conveniente, resultando em normais e gamas. Um algoritmo do tipo MCCM (Gibbs) pode ser construído para estimação das distribuições à posteriori. Alternativamente, pode-se usar o WinBUGS apenas indicando a representação hierárquica.
- Reparametrização: $\beta = \tau\delta$ e $\eta = \tau\sqrt{1 - \delta^2}$

$$X \mid \psi, \beta, \eta, W \sim N(\psi + \beta W, \eta^2)$$

$$W \sim NT(0, 1)I_{(0, \infty)}$$

- O MAP sob priori de Jeffreys apresenta "melhor" comportamento que o EMV. Critérios clássicos de comparação (eqm e viés).
- A média a posteriori sob priori uniforme para $\delta \in [-1, 1]$ apresentou o melhor comportamento no geral.
- Obteve-se também intervalos HDR e teste de Hipóteses baseado no Fator de Bayes

Aplicação 1: MAE-116

Notas da primeira prova (87 alunos) da disciplina Noções de Estatística, 1º semestre de 2006, IME-USP.

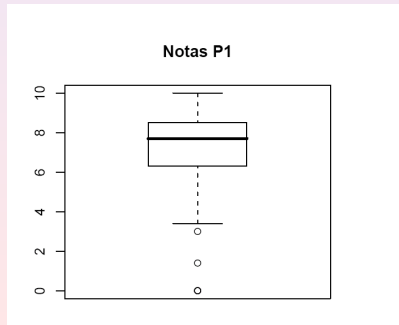
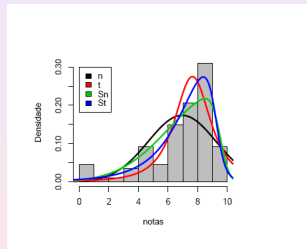


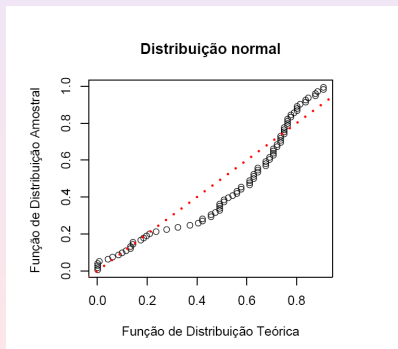
Tabela: Medidas-resumo

| | Min. | 1° Quartil | Mediana | Média | 3° Quartil | Máx. | var |
|-------|-------|------------|---------|-------|------------|--------|-------|
| RP | 0.000 | 4.650 | 7.100 | 6.618 | 8.400 | 9.700 | 5.564 |
| CB | 0.000 | 6.700 | 7.800 | 7.063 | 8.500 | 10.000 | 5.282 |
| Total | 0.000 | 6.300 | 7.700 | 6.951 | 8.500 | 10.000 | 5.327 |

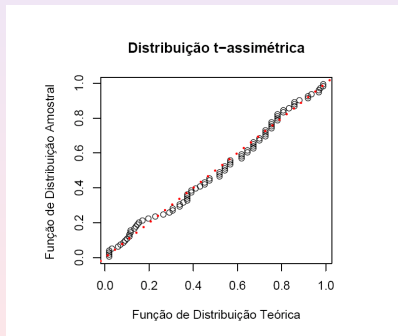
Aplicação: Histograma e curvas de densidades associadas



Aplicação: Gráfico de probabilidades-acumuladas



Aplicação: Gráfico de probabilidades-acumuladas



Aplicação 1: Estimativas Bayesianas

| Priori ν | Parâ | Média | Desvio | 2.5% | mediana | 97.5% | Erro MCMC |
|---------------|----------|--------|--------|--------|---------|--------|-----------|
| U(1, 30) | ψ | 9.270 | 0.256 | 8.732 | 9.286 | 9.748 | 0.008 |
| | τ^2 | 2.613 | 2.880 | 1.332 | 2.598 | 3.539 | 0.104 |
| | ν | 9.131 | 7.226 | 1.799 | 6.334 | 27.21 | 0.183 |
| | α | -4.783 | 1.947 | -9.261 | -4.543 | -1.751 | 0.096 |
| Exp(0.5)I(1,) | ψ | 9.108 | 0.239 | 8.628 | 9.119 | 9.552 | 0.008 |
| | τ^2 | 2.038 | 1.841 | 1.167 | 1.978 | 2.865 | 0.071 |
| | ν | 3.188 | 1.408 | 1.514 | 2.862 | 6.959 | 0.042 |
| | α | -3.836 | 1.530 | -7.461 | -3.579 | -1.517 | 0.076 |

Estimativas de máxima verossimilhança:

$$\hat{\psi} = 9.238, \hat{\tau}^2 = 2.308, \hat{\alpha} = -4.194, \hat{\nu} = 3.653$$

$$\hat{\mu} = 6.933, \hat{\sigma}^2 = 2.798 \text{ (média e variância).}$$

Modelos Mistos com erros t-assimétricos

Modelos lineares mistos são muito utilizados na análise de dados com medidas repetidas porque permitem modelar a correlação entre sujeitos frequentemente presente neste tipo de dados.

MLM proposto por Laird e Ware(1982)

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i, \quad i = 1, \dots, m$$

- $Y_i(n_i \times 1)$ vetor de respostas,
- $\beta(p \times 1)$ efeito fixo,
- $X_i(n_i \times p)$ matriz de delineamento do efeito fixo,
- $Z_i(n_i \times q)$ matriz de delineamento do efeito aleatório,
- $b_i(q \times 1)$ efeito aleatório,
- $\epsilon_i(n_i \times 1)$ vetor de erros aleatórios.

A suposição usual é

$$\begin{pmatrix} b_i \\ \epsilon_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Psi & 0 \\ 0 & \Sigma \end{pmatrix} \right). \quad (4)$$

$i = 1, \dots, m$ independentes. Além disso, a suposição acima implica em independência entre erros e efeitos aleatórios.

Fora da normalidade existem diferentes maneiras de especificar o MLM.

- Sob o modelo t-Student: (a) mantendo-se a estrutura anterior obtém-se correlação nula, mas não independência; (b) supondo independência entre erro e efeito aleatório, perde-se a estrutura conjunta.
- Sob o modelo SN de SDB, obtém-se conclusão similar ao normal. Sob outras versões da SN isso não é válido.
- Sob o modelo ST, a estrutura anterior não implica nem mesmo em não correlação entre erros e efeitos aleatórios.
- Consideramos b_i q -dimensional skew-t com $E[b_i] = 0$, e cada ϵ_{ij} univariado skew-t com $E[\epsilon_{ij}] = 0$, independentes.

Aplicação 2

The data are from a reproductive toxicology study in rats, the outcome variable is birth weights (in grams) of rat pups. Three equal size treatment groups: control, a low dose, and a high dose of the test substance.

$$y_{ij} = \beta_0 + \beta_1 LD_i + \beta_2 HD_i + \gamma L_i + \delta S_{ij} + b_i + e_{ij}, \quad (5)$$

where y_{ij} is birth weights (in grams) of rat pup j of litter i ; LD_i and HD_i are dummy variables for the low and high dose of the test substance for the litter i ; L_i is the size of litter i ; S_{ij} is an indicator variable for male rat pups; b_i is the random effect for the litter i ; and e_{ij} is the within-litter error, $i = 1, \dots, 47$, $j = 1, \dots, L_i$.

Four distributions are considered: skew- t (St), skew-normal (SN), t and normal(N). We computed the pseudo Bayes factor (Gelfand, 1996), Expected Akaike Information Criterion (EAIC) and Expected Bayesian Information Criterion (EBIC) proposed in Brooks (2002) for each model relative to the normal model.

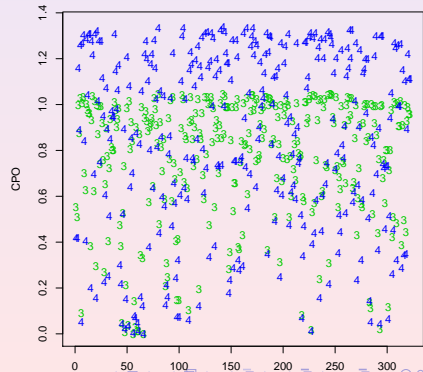
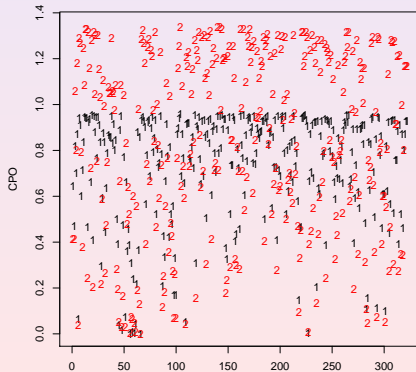
Tabela: Model comparison measures for the rats data set

| Random Effects | Residuals | pBF | EAIC | EBIC |
|----------------|-----------|-----------------------|--------|--------|
| N | St | 2.32×10^{21} | 311.32 | 458.53 |
| N | SN | 8.57×10^9 | 361.80 | 505.23 |
| N | t | 4.07×10^{19} | 316.61 | 460.05 |
| N | N | 1.00 | 404.70 | 544.36 |

Tabela: Posterior mean and 95%–HPD interval

| Parameter | N – St | St–St |
|-------------|--------------------------|-------------------------|
| β_1 | -0.52 (-0.85, -0.17) | -0.52 (-0.88, -0.18) |
| β_2 | -0.94 (-1.34, -0.53) | -0.94 (-1.35, -0.52) |
| $Var[b_i]$ | 0.14 (0.06, 0.23) | 0.17 (0.06, 0.32) |
| λ_e | -0.77 (-1.58, -0.003) | -0.73 (-1.57, 0.06) |
| ν_e | 4.13 (3.00, 5.81) | 4.14 (3.00, 5.85) |
| λ_b | - | 0.04 (-1.93, 1.85) |
| ν_b | - | 86.45 (3.00, 105.10) |

Gráficos CPO: (1) N-N, (2) N-t, (3) N-SN e (4) N-St



Regressão binária ou binomial

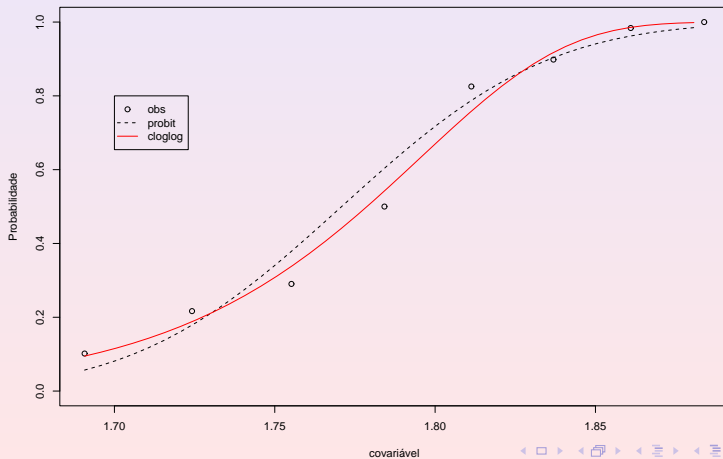
Considere

- $y = (y_1, \dots, y_n)^\top$ um conjunto de variáveis Bernoulli (0 – 1) independentes;
- $x_i = (x_{i1}, \dots, x_{ip})^\top$ um conjunto de quantidades previamente fixadas.
- $p_i = p(y_i = 1) = F(x_i^\top \beta)$.
- F^{-1} lineariza a relação entre a probabilidade de sucesso e as covariáveis. Exemplos: logito, probito e cloglog .
- Usualmente a inferência é feita via máxima verossimilhança, considerando a teoria assintótica.

Aplicação 3: Dados dos Besouros

481 besouros foram expostos a diferentes doses do gás CS_2 (sulfato de carbono gasoso), e observou-se o número de besouros mortos. Deseja-se estudar a relação entre dose do gás e morte dos besouros (modelo de dose/resposta).

| Dose $\log CS_2$ | Besouros expostos | Besouros mortos | Valores ajustados | | |
|---------------------|----------------------|--------------------|-------------------|--------|--------|
| | | | Cloglog | Probit | Logito |
| 1.69 | 59 | 06 | 5.7 | 3.4 | 3.5 |
| 1.72 | 60 | 13 | 11.3 | 10.7 | 9.8 |
| 1.75 | 62 | 18 | 20.9 | 23.4 | 22.4 |
| 1.78 | 56 | 28 | 30.3 | 33.8 | 33.9 |
| 1.81 | 63 | 52 | 47.7 | 49.6 | 50.0 |
| 1.84 | 59 | 53 | 54.2 | 53.4 | 53.3 |
| 1.86 | 62 | 61 | 61.1 | 59.7 | 59.2 |
| 1.88 | 60 | 60 | 59.9 | 59.2 | 58.8 |



A ligação probito-assimétrica

Considera

$$p_i = F_\lambda(x_i^\top \beta) \quad i = 1, \dots, n. \quad (6)$$

F_λ é a distribuição normal-assimétrica e λ é um parâmetro que controla a assimetria.

- Se $\lambda = 0$ obtemos a usual ligação probito.
- Se F_λ é a normal-assimétrica padrão de SDB(2003), segue a ligação proposta por Chen, Dey e Shao (1999).
- Se F_λ é a normal-assimétrica padrão de Azzalini (1985), segue a ligação proposta por Bázan, Branco e Bolfarine (2006).
- Relação entre essas duas propostas pode ser encontrada em Bázan, Bolfarine e Branco (2008).

Análise dos dados dos Besouros

Tabela: Models comparison considering different links for the beetle mortality data (n=481)

| class | type | models | thin * | time ** (sec.) | <i>EBIC</i> | <i>EAIC</i> |
|----------------|------------|-----------------|---------------|-------------------|-------------|-------------|
| non- <i>sp</i> | symmetric | logit | 5 | 37 | 386.8 | 378.4 |
| | | asymmetric | cloglog | 5 | 41 | 379.1 |
| | asymmetric | scobit | 65 | 48 | 387.7 | 375.2 |
| | | power-logit | 40 | 48 | 386.1 | 373.6 |
| | | quadratic logit | 15000 | 57 | 392.3 | 379.8 |
| <i>sp</i> | symmetric | probit | 5 | 39 | 385.8 | 377.4 |
| | | asymmetric | <i>CDS sp</i> | 610 | 199 | 178.9 |
| | asymmetric | <i>BBB sp</i> | 470 | 165 | 313.7 | 301.2 |

Tabela: Posterior summaries for the skewness parameters in skew-probit models using beetle mortality data

| models | parameter | mean | median | 95 % HPD | | MCerror |
|---------------|-----------|-------|--------|-------------|-------------|---------|
| | | | | Lower Bound | Upper Bound | |
| <i>CDS sp</i> | λ | -4.19 | -4.02 | -7.82 | -1.43 | 0.07 |
| <i>BBB sp</i> | δ | -0.48 | -0.70 | -0.99 | 0.46 | 0.02 |
| | α | -1.33 | -0.99 | -4.46 | 0.87 | 0.05 |

Referências Importantes

- Azzalini and Capitanio (1999). JRSS-B. Skew-normal.
- Branco and Dey (2001). JMA. Skew-elliptica.
- Arnold and Beaver (2002). Test. First review.
- Azzalini and Capitanio (2003). JRSS-B. Skew-t .
- Sahu, Dey and Branco, M.D. (2003). CJS. A new class with Bayesian regression.
- Arellano and Genton (2005). JMA. Fundamental SS.
- Arellano and Azzalini (2006). Scand.in.Journal. Unification.
- Arellano-Valle, Branco and Genton (2006). CJS. Selection distribution.

Referências Importantes

- Azzalini and Capitanio (1999). JRSS-B. Skew-normal.
- Branco and Dey (2001). JMA. Skew-elliptica.
- Arnold and Beaver (2002). Test. First review.
- Azzalini and Capitanio (2003). JRSS-B. Skew-t .
- Sahu, Dey and Branco, M.D. (2003). CJS. A new class with Bayesian regression.
- Arellano and Genton (2005). JMA. Fundamental SS.
- Arellano and Azzalini (2006). Scandian.Journal. Unification.
- Arellano-Valle, Branco and Genton (2006). CJS. Selection distribution.
- Genton (2004). A Journey Beyond Normality. Chapman and Hall/CRC.

Referências Iniciais

- Cristian Bayes (2005). Dissertação de mestrado. IME-USP
- Luciana Godoi (2007). Dissertação de mestrado. IME-USP.
- Rafael Farias (2007). Dissertação de mestrado. IME-USP
- Bayes e Branco (2007). BJPS, v.21,2.