

# Modelos com distribuições assimétricas para os traços latentes

Prof. Caio Azevedo

## ■ Resumo da apresentação

- Breve Introdução.
- Modelo de 2 parâmetros.
- Motivação: distribuições latentes.
- Distribuição normal assimétrica univariada centrada.
- Representações estocásticas e identificabilidade.
- Abordagem hierárquica.
- Estimação bayesiana.
- Estudo de simulação.
- Conclusões e comentários.

## ■ Introdução

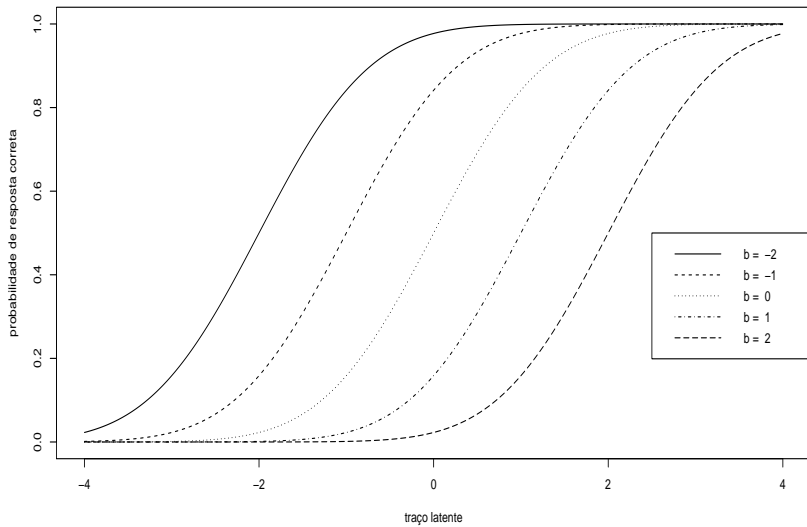
- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e itens de um instrumento de medida (prova, questionário).
- Tal modelagem consiste na probabilidade de obter um certo escore em cada item.
- Existe um grande número de classes de MRI : dicotômicos e policotômicos, um e múltiplos grupos, multidimensionais, longitudinais multiníveis, dentre outros.
- Um MRI apresenta um grande número de parâmetros.

$$P(Y_{ij} = 1 | (\theta_j, \zeta_i)) = \Phi(a_i(\theta_j - b_i^*)) = \Phi(a_i\theta_j - b_i) \quad (1)$$

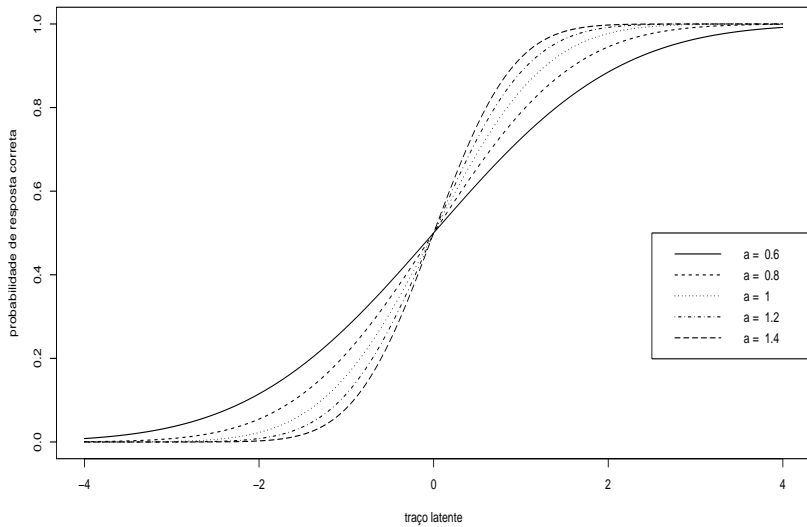
$i = 1, \dots, l$  (item),  $j = 1, \dots, n$  (indivíduo),

- $Y_{ij}$  : é a resposta do indivíduo  $j$  ao item  $i$ . É igual a 1 se o indivíduo responde corretamente e 0 caso contrário.
- $\theta_j$  : é o traço latente (conhecimento, nível de depressão, etc) do indivíduo  $j$ .
- $\zeta_i = (a_i, b_i)^t$ .
- $a_i$  : é o parâmetro de discriminação associado ao item  $i$ .
- $b_i^*$  : é o parâmetro de dificuldade associado ao item  $i$ .
- $b_i = a_i b_i^*$  : é o intercepto associado ao item  $i$ .

### Curvas do modelo L2P



### Curvas do modelo L2P



- Aplicação do modelo de 2 parâmetros:
  - Testes de conhecimento com itens dissertativos corrigidos como certo/errado: ensino fundamental.
  - Questionários de avaliação médica.
    - Psiquiatria.
    - Qualidade de vida.
  - Questionários de avaliação institucional.
    - Gestão pela qualidade total.
    - Gestão escolar.
  - Testes com itens de múltiplas escolha.

- Traços latentes:
  - Características inerentes aos indivíduos.
  - Efeitos fixos: parâmetros.
  - Efeitos aleatórios: variáveis aleatórias.
  - Função de resposta ao item caracteriza completamente um MRI.
  - Atribuir uma distribuição de probabilidade para os traços latentes.
    - Incorpora características do processo de amostragem.
    - Incorpora informações a priori.
    - Permite fazer inferência para indivíduos não observados.
    - Contribui para se obter a identificabilidade do modelo.



- Uma suposição usual consiste em:  $\theta_j | \boldsymbol{\eta}_\theta \sim N(0, 1)$ ,  
 $\boldsymbol{\eta}_\theta = (\mu_\theta = 0, \psi_\theta = 1)$ .
- Esta suposição nem sempre é realista.
- Desvios da suposição de normalidade: assimetria, polimodalidade, caudas mais pesadas ou mais leves.

## ■ Propostas.

- Mistura de normais: Mislevy (1984).
- Beta-Binomial: Mislevy (1984).
- Estimaco de mxima verossimilhana no paramtrica: Mislevy(1984).
- Distribuico t multivariada com graus de liberdade conhecidos: Ghosh et al (2000).
- Normal assimtrica: Bazan, Branco & Bolfarine et al (2006).
- Distribuico t com graus de liberdade conhecidos: Azevedo & Andrade (2007).
- Normal assimtrica sob a parametrizaco centrada: Azevedo, Bolfarine & Andrade (2011 a)

- Foco do trabalho: assimetria.
  - Seleção de indivíduos: maior escore em uma primeira etapa, menor condição sócio-econômica.
  - Programas especiais de ensino longitudinais (maior conhecimento).
  - Natureza dos dados.
- Alternativa

$$\begin{aligned}\theta_j | \boldsymbol{\eta}_\theta &\sim NA(0, 1, \lambda_\theta), \\ \boldsymbol{\eta}_\theta &= (\mu_\theta = 0, \psi_\theta = 1, \lambda_\theta)\end{aligned}\tag{2}$$

- Na TRI tem-se a questão de determinação da escala da proficiência.
- Tornar os resultados não só comparáveis como também interpretáveis.

- Fato: sob a suposição (2):

$$\mathcal{E}(\theta_j|\lambda_\theta) = h(\lambda_\theta) \quad (3)$$

$$\mathcal{V}(\theta_j|\lambda_\theta) = g(\lambda_\theta) \quad (4)$$

- Isto torna o modelo (1) não-identificável.

Note que:

$$\begin{aligned} P(Y_{ij} = 1 | \theta_j, \zeta_i) &= \Phi(a_i(\theta_j - b_i)) = \Phi\left(\frac{a_i}{\alpha}(\alpha\theta_j - \alpha b_i)\right) \\ &= \Phi\left(\frac{a_i}{\alpha}(\alpha\theta_j + \beta - \alpha b_i - \beta)\right) \\ &= \Phi(-a_i^*(\theta_j^* - b_i^*)) \end{aligned} \quad (5)$$

em que  $\theta_j^* | \lambda_\theta \sim NA(\beta, \alpha^2, \lambda_\theta)$ .

Isso ocorre pois o valor esperado e a variância de  $\theta$ , ou seja, a métrica (escala), não está determinada. Isto, por sua vez, induz uma falta de identificabilidade no modelo.

**Identificabilidade  $\leftrightarrow$  métrica está determinada.**

- Solução: utilizar a parametrização centrada definida por Azzalini (1989):

$$\theta_j^{(C)} = \frac{\theta_j - h(\theta)}{\sqrt{g(\theta)}} \quad (6)$$

- Dessa forma,  $\theta_j^{(C)} \sim NA(h_1(\lambda_\theta), h_2(\lambda_\theta), \lambda_\theta)$ , em que,  $\forall \lambda_\theta \in IR$ :

$$\mathcal{E}(\theta_j^{(C)} | \lambda_\theta) = 0$$

$$\mathcal{V}(\theta_j^{(C)} | \lambda_\theta) = 1$$

e, dessa forma, transformações do tipo (5) não são mais possíveis e, portanto, o modelo está identificado (a métrica está definida).

Notação :  $\theta_j^{(C)} \sim NA_C(0, 1, \gamma_\theta)$

## ■ Henzé

$$\theta_j^{(H)} = \delta_\theta H_j + \left( \sqrt{1 - \delta_\theta^2} \right) \xi_j, \quad (7)$$

em que  $H_j \perp \xi_j$ ,  $H_j \sim HN(0, 1)$ ,  $\xi_j \sim N(0, 1) \forall j$ ,  $\delta_\theta = \frac{\lambda_\theta}{1 + \lambda_\theta^2}$ , e:

$$\mathcal{E}(\theta_j^{(H)} | \lambda_\theta) = \delta_\theta \sqrt{\frac{2}{\pi}}; \mathcal{V}(\theta_j^{(H)} | \lambda_\theta) = \left( 1 - \frac{2}{\pi} \delta_\theta^2 \right)$$

## ■ Sahu

$$\theta_j^{(S)} = \lambda_\theta H_j + \xi_j, \quad (8)$$

em que  $H_j \perp \xi_j$ ,  $H_j \sim HN(0, 1)$ ,  $\xi_j \sim N(0, 1) \forall j$ , e:

$$\mathcal{E}(\theta_j^{(S)} | \lambda_\theta) = \lambda_\theta \sqrt{\frac{2}{\pi}}; \mathcal{V}(\theta_j^{(S)} | \lambda_\theta) = \left( 1 - \frac{2}{\pi} \right) \lambda_\theta^2$$

- Resultado:** Sob a parametrização centralizada, as representações estocásticas de Henzé e Sahu et al são idênticas, isto é, produzem a mesma densidade, dada por:

$$f(\theta_j|\gamma_\theta) = 2(\varsigma_\theta)^{1/2} \phi\left(\frac{\theta_j - \tau_\theta}{\sqrt{\varsigma_\theta}}\right) \Phi\left[v_\theta\left(\frac{\theta_j - \tau_\theta}{\sqrt{\varsigma_\theta}}\right)\right], \quad (9)$$

em que

$$\begin{aligned} \tau_\theta \equiv \tau_\theta(\gamma_\theta) &= -s\gamma_\theta^{1/3}; \varsigma_\theta \equiv \varsigma_\theta(\gamma_\theta) = 1 + s^2\gamma_\theta^{2/3} \\ v_\theta \equiv v_\theta(\gamma_\theta) &= \frac{-s\gamma_\theta^{1/3}}{\sqrt{r^2 + s^2\gamma_\theta^{2/3}(r^2 - 1)}}; s = \sqrt{\frac{2}{\pi}}, r = \left(\frac{2}{4 - \pi}\right)^{1/3}, \end{aligned}$$

em que  $\gamma_\theta$  é o coeficiente de assimetria de Pearson o qual é dado por:

$$\gamma_\theta = \sqrt{\frac{2}{\pi}} \delta_\theta^3 \left[\frac{4}{\pi} - 1\right] \left[1 - \frac{2}{\pi} \delta_\theta^2\right]^{-3/2}, \gamma_\theta \in (-0.99527, 0.99527)$$



- O modelo anterior (Função de Resposta ao Item (1) + distribuição normal assimétrica (6)) foi proposto por Azevedo, Bolfarine and Andrade (2011 a).
- Sob a parametrização centralizada, indexada por  $\gamma_\theta$ , pode-se considerar a seguinte representação estocástica:

$$\theta_j = \left[ v_\theta T_j + \left( \sqrt{1 - v_\theta^2} \right) \xi_j \right] \sqrt{s_\theta} + \tau_\theta, \quad (10)$$

em que  $T_j \sim HN(0, 1)$ ,  $HN(\mu, \psi)$  representa uma distribuição normal truncada à esquerda do 0 com parâmetros  $(\mu, \psi)$ ,  $\xi_j \sim N(0, 1)$  e  $T_j \perp \xi_j, \forall j$ .

- Duas abordagens:
  - Densidade original: Azevedo, Bolfarine and Andrade (2011 a)
  - Representação hierárquica: Azevedo, Bolfarine and Andrade (2011 b)
  - Apresentaremos resultados relativos ao segundo trabalho, embora este verse sobre comparações com os resultados do primeiro.

- Priori

$$p(\boldsymbol{\theta}, \mathbf{t}, \boldsymbol{\zeta}, \gamma_\theta) = \left\{ \prod_{j=1}^n p(\theta_j | t_j, \gamma_\theta) \right\} \left\{ \prod_{j=1}^n p(t_j) \right\} \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_\zeta) \right\} \\ \times p(\gamma_\theta | \boldsymbol{\eta}_\lambda)$$

- Traços latentes condicionado às variáveis latentes  $t_j$ : Normal simétrica.
- Variáveis latentes  $t_j$ : Normal simétrica truncada à esquerda 0.
- Parâmetros dos itens: Normal simétrica bivariada truncada à esquerda do 0 na componenten  $a_i$ .
- Parâmetro de assimetria: beta modificada e Jeffreys.

- Posteriori sem forma analítica tratável, independentemente da escolha das prioris.
- Solução: algoritmos do tipo MCMC. Neste caso, Metropolis-Hastings dentro do amostrador de Gibbs.
- Compararemos dois algoritmos: ADMHWGSR (abordagem utilizando a representação heirárquica da distribuição dos traços latentes) e ADMHWGS (utilizando a densidade original da distribuição dos traços latentes).

$$\begin{aligned}
p(\mathbf{z}_{..}, \boldsymbol{\theta}, \mathbf{t}, \boldsymbol{\zeta}, \gamma_{\theta} | \mathbf{y}_{..}, \boldsymbol{\eta}_{\zeta}, \boldsymbol{\eta}_{\gamma}) &\propto \left\{ \prod_{j=1}^n \prod_{i \in \mathcal{I}_j} p(z_{ij} | \theta_j, \zeta_i, y_{ij}) \right\} \left\{ \prod_{j=1}^n p(\theta_j | t_j, \gamma_{\theta}) \right\} \\
&\times \left\{ \prod_{j=1}^n p(t_j) \right\} \left\{ \prod_{i=1}^l p(\zeta_i | \boldsymbol{\eta}_{\zeta}) \right\} \{ p(\gamma_{\theta} | \boldsymbol{\eta}_{\gamma}) \} \\
&\propto \prod_{j=1}^n \prod_{i \in \mathcal{I}_j} \exp \left[ -0.5 (z_{ij} - a_i \theta_j + b_i)^2 \right] \mathbb{1}_{(y_{ijk}, z_{ijk})} \\
&\times \left\{ \prod_{j=1}^n \exp \left\{ -\frac{1}{2\psi_{\theta}} [\theta_j - \mu_{\theta} t_j - \alpha_{\theta}]^2 \right\} \right\} (\psi_{\theta})^{-n/2} \\
&\times \left\{ \prod_{i=1}^l \exp \left[ (\zeta_i - \boldsymbol{\mu}_{\zeta})^t \boldsymbol{\Psi}_{\zeta}^{-1} (\zeta_i - \boldsymbol{\mu}_{\zeta}) \right] \mathbb{1}_{(a_i > 0)} \right\} \\
&\times \left\{ \prod_{j=1}^n \exp \left\{ -0.5 t_j^2 \right\} \mathbb{1}_{(t_j > 0)} \right\} \\
&\times \left( 1 + \frac{2\lambda_{\theta}^2}{\pi^2/4} \right)^{-3/2} \mathbb{1}_{(\gamma_{\theta} \in A_{\gamma_{\theta}})},
\end{aligned}$$

- $\lambda_\theta = (-3, -2, -1, 0, 1, 2, 3) \rightarrow$   
 $\gamma_\theta = (-0, 67; -0.45; -0, 14; 0, 00; 0.14; 0, 45; 0, 67).$
- $n \in \{500, 100\}$  indivíduos e  $l \in \{24, 36\}$ .
- Itens com parâmetros tais que:  $a \in (0.8, 1.4)$  e  $b \in (-2, 2)$ .
- Um único conjunto de respostas, para cada uma das três situações. Em cada uma delas consideramos:  $\mu_\zeta = (1, 0)$ ,  $\Psi_\zeta = \text{diag}(0.5, 3)$  e  $p(\gamma_\theta | \eta_\gamma)$
- Priori de Jeffreys e priori beta modificada com diferentes valores para os hiperparâmetros.
- Burn-in: 5000, simulando mais 3500, retendo os valores a cada 30 iterações, gerando um total de 1000 valores.
- Densidades propostas:  $\gamma_\theta^{(t)} | \gamma_\theta^{(t-1)} \sim U(\gamma_\theta^{(t-1)} - \Delta_\gamma, \gamma_\theta^{(t-1)} + \Delta_\gamma).$
- Um total de  $R = 10$  replicas foram geradas.

Method	$\alpha_{\gamma_1} = \alpha_{\gamma_2}$	$\Delta_\gamma$	Method	$\alpha_{\gamma_1}$	$\Delta_\gamma$
ADMHWGSR1	0.5	0.01	ADMHWGSR16	10	0.01
ADMHWGSR2	0.5	0.02	ADMHWGSR17	10	0.02
ADMHWGSR3	0.5	0.03	ADMHWGSR18	10	0.03
ADMHWGSR4	0.5	0.04	ADMHWGSR19	10	0.04
ADMHWGSR5	0.5	0.05	ADMHWGSR20	10	0.05
ADMHWGSR6	1	0.01	ADMHWGSR21	Jeffreys prior	0.01
ADMHWGSR7	1	0.02	ADMHWGSR22	Jeffreys prior	0.02
ADMHWGSR8	1	0.03	ADMHWGSR23	Jeffreys prior	0.03
ADMHWGSR9	1	0.04	ADMHWGSR24	Jeffreys prior	0.04
ADMHWGSR10	1	0.05	ADMHWGSR25	Jeffreys prior	0.05
ADMHWGSR11	2	0.01			
ADMHWGSR12	2	0.02			
ADMHWGSR13	2	0.03			
ADMHWGSR14	2	0.04			
ADMHWGSR15	2	0.05			

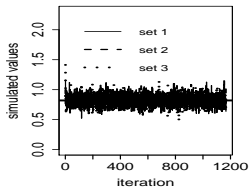
# Estudos de simulação

- Próximos slides:
- 1- 25 ADMHWGSR (com representação estocástica).
- 26- 50 ADMHWGS (sem representação estocástica).

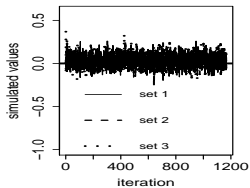


## Gráficos de trajetória para alguns parâmetros

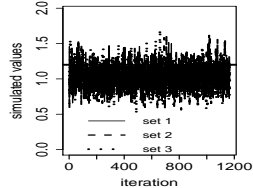
Item 7: discrimination parameter



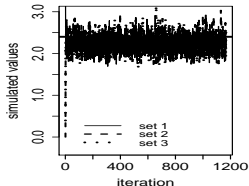
Item 7: difficulty parameter



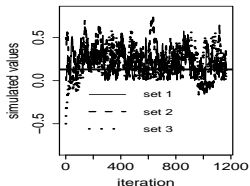
Item 12: discrimination parameter



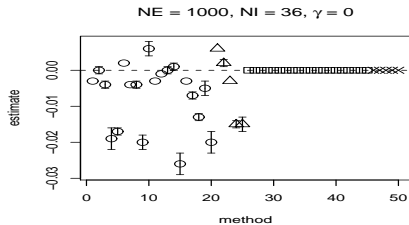
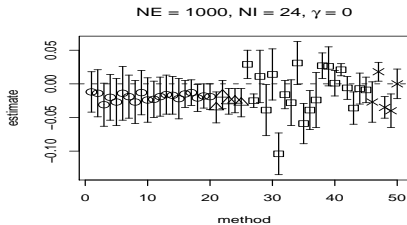
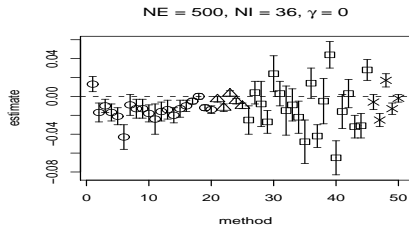
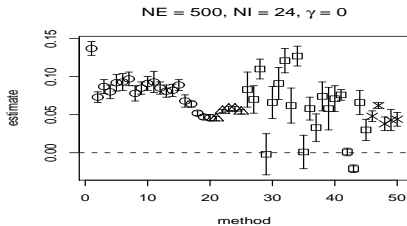
Item 12: difficulty parameter



asymmetry coefficient

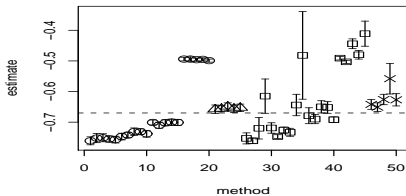


# Estimativas de $\gamma_\theta$ , sob ( $\gamma_\theta = 0$ )

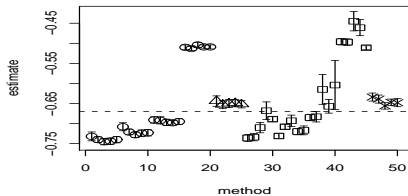


# Estimativas de $\gamma_\theta$ , sob ( $\gamma_\theta = -0.67$ )

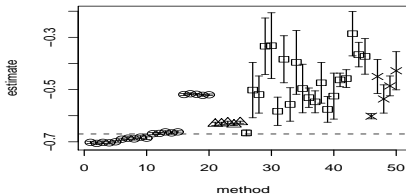
NE = 500, NI = 24,  $\gamma = -0.67$



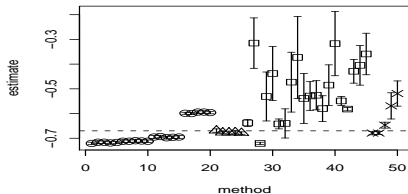
NE = 500, NI = 36,  $\gamma = -0.67$



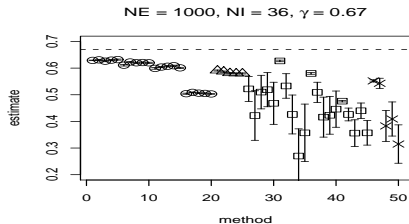
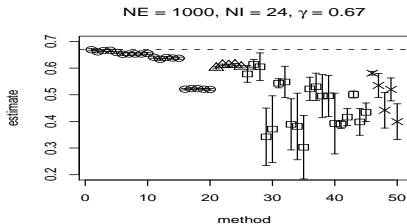
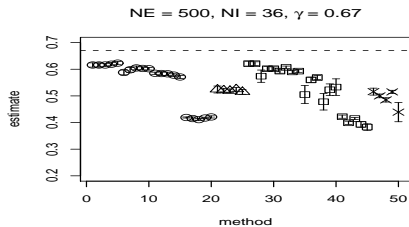
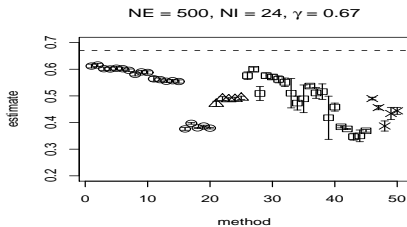
NE = 1000, NI = 24,  $\gamma = -0.67$



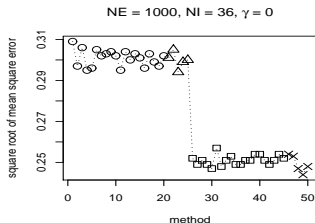
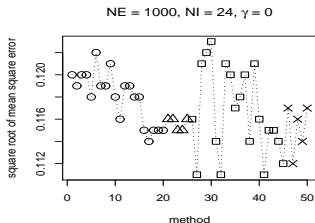
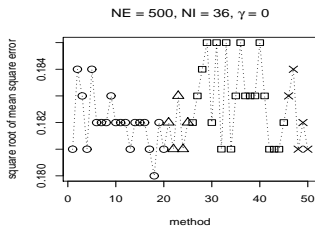
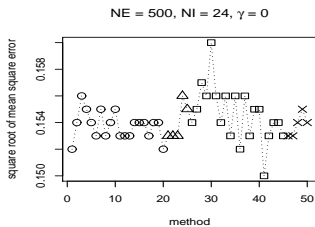
NE = 1000, NI = 36,  $\gamma = -0.67$



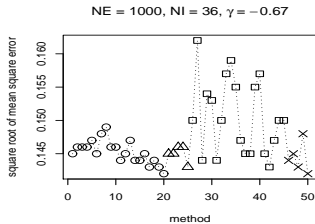
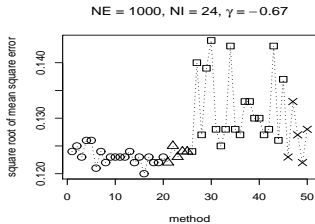
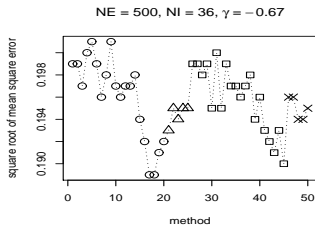
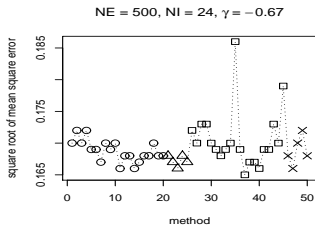
# Estimativas de $\gamma_\theta$ , sob ( $\gamma_\theta = 0.67$ ).



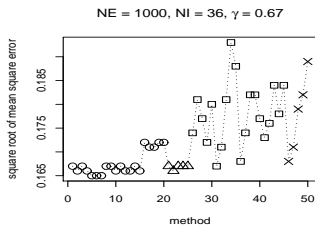
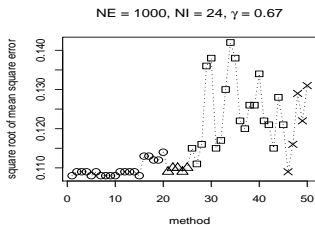
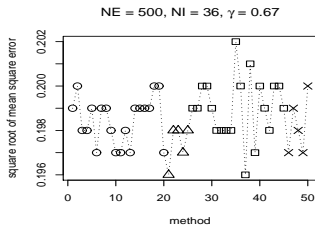
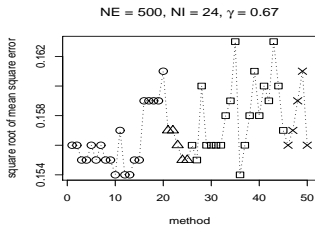
# RMSE para os param. de discriminação, sob $\gamma_\theta = 0$



RMSE para os parâm. de discriminação, sob  $\gamma_\theta = -0.67$ .

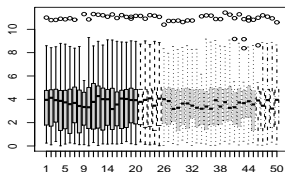


RMSE para o parâmetro de discriminação, sob  $\gamma_\theta = 0.67$ .

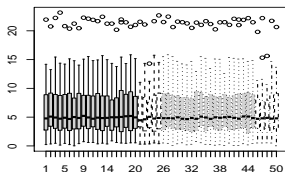


# AVRB para os parâmetros de discriminação, sob $\gamma_\theta = 0$

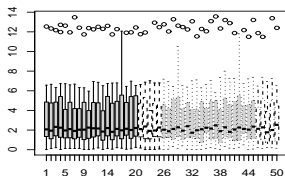
NE = 500, NI = 24,  $\gamma = 0$



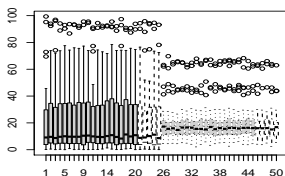
NE = 500, NI = 36,  $\gamma = 0$



NE = 1000, NI = 24,  $\gamma = 0$



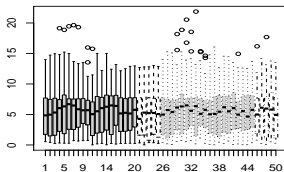
NE = 1000, NI = 36,  $\gamma = 0$



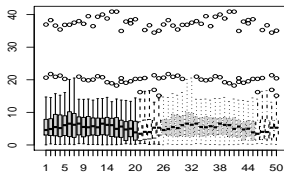


# AVRB para o parâmetro de discriminação, sob $\gamma_\theta = -0.67$

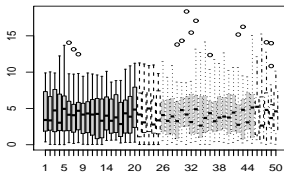
NE = 500, NI = 24,  $\gamma = -0.67$



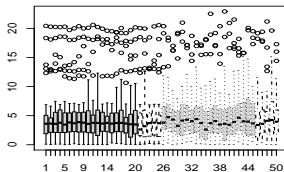
NE = 500, NI = 36,  $\gamma = -0.67$



NE = 1000, NI = 24,  $\gamma = -0.67$

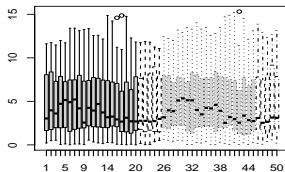


NE = 1000, NI = 36,  $\gamma = -0.67$

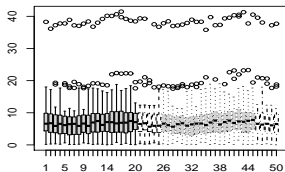


## AVRB para o parâmetro de discriminação, sob $\gamma_\theta = 0.67$

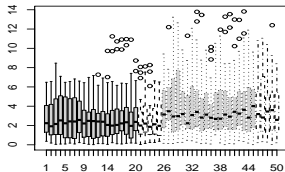
NE = 500, NI = 24,  $\gamma = 0.67$



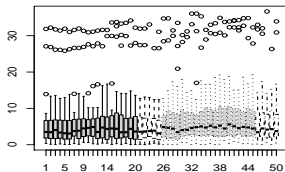
NE = 500, NI = 36,  $\gamma = 0.67$



NE = 1000, NI = 24,  $\gamma = 0.67$



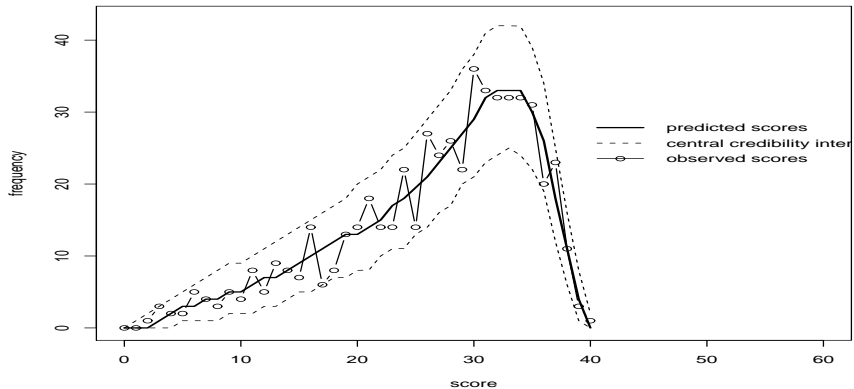
NE = 1000, NI = 36,  $\gamma = 0.67$



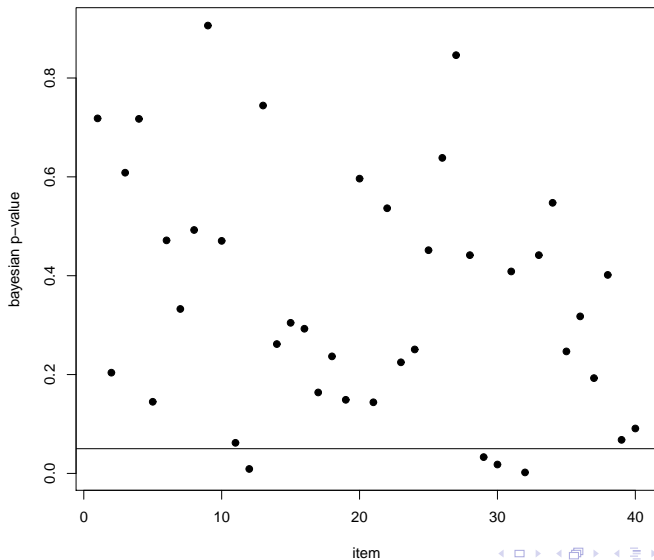
- Análise de dados reais
- Um grupo de 557 crianças (do segundo ano do ensino básico) respondeu uma prova de Matemática 40 itens dissertativos corrigidos como certo/errado.
- O Conjunto de dados em questão é parte de um estudo longitudinal.
- Estimativa do coeficiente de assimetria  $\gamma_\theta$ .

Estimativa	EPP	HPD
-0,81	0,07	[-0,68; -0,95]

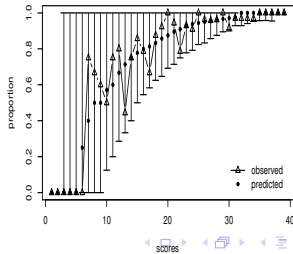
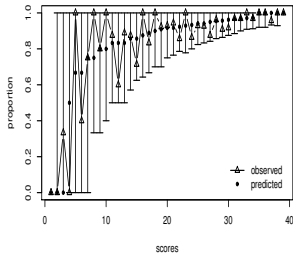
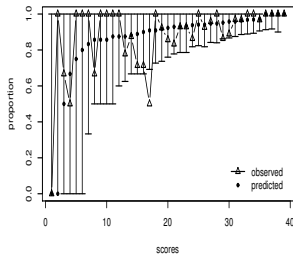
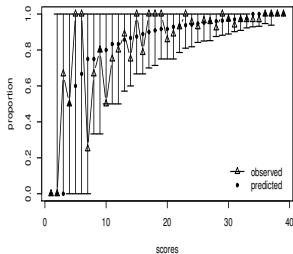
## Escores observados e preditos e intervalos de credibilidade de 95%



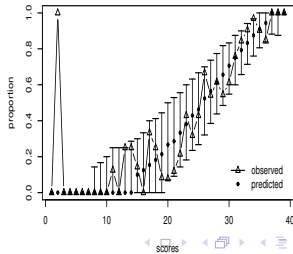
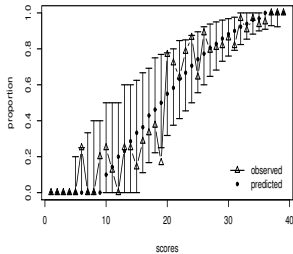
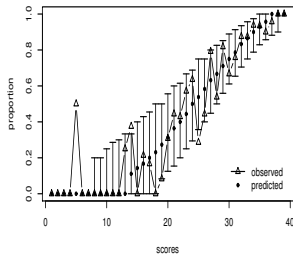
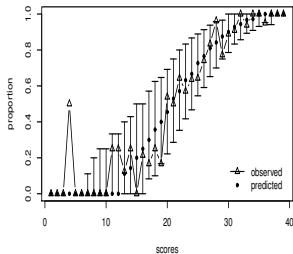
## p-valor bayesiano para os itens



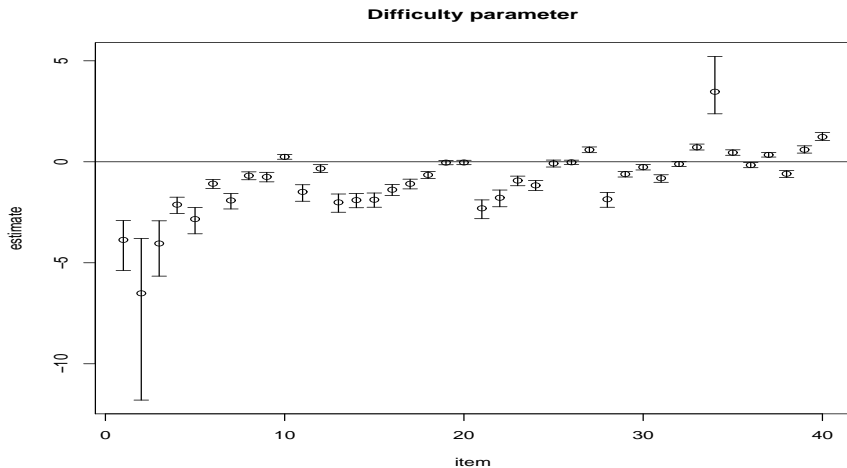
## Gráficos de ajuste por item



## Gráficos de ajuste por item

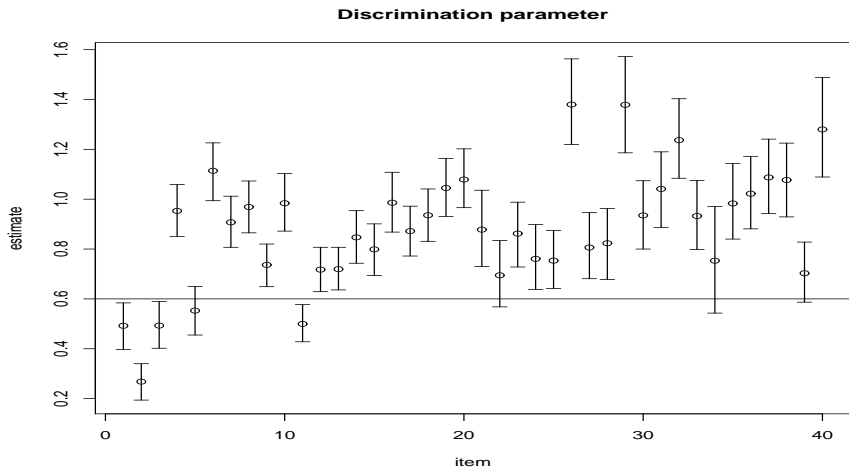


# Estimativas dos parâmetros de dificuldade





## Estimativas dos parâmetros de discriminação



## ■ Comentários

- Parametrização centralizada garante a identificabilidade : a métrica está definida.
- O método de estimação mostrou-se capaz de recuperar os verdadeiros valores dos parâmetros.
- O algoritmo proposto é mais rápido e estima melhor os parâmetros do que o seu predecessor, em algumas situações.
- O modelo capta desvios de normalidade (simetria) de uma forma simples e interpretável.
- As ferramentas de diagnóstico mostraram-se importantes na análise do conjunto de dados.

- C. L. N. Azevedo, H. Bolfarine and D. F. Andrade (2011 a), *Bayesian inference for a skew-normal IRT model under the centred parameterization*. *Computational Statistics & Data Analysis*, 55, 353-365.
- C. L. N. Azevedo, H. Bolfarine and D. F. Andrade (2011 b), *Parameter recovery for a skew normal IRT model under a Bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size*, *Journal of Statistical Computation and Simulation*, 82, 11, 1679-1699.
- J. L. Bazán, M. D. Branco and H. Bolfarine (2006), *A Skew Item Response Model*, *Bayesian analysis*, 861-892.
- N. Henze (1986), *A probalistic representation of the skew-normal distribution*, *Scandinavian journal of statistics*, 13, 271-275 .
- S. K. Sahu, D. K. Dey and M. D. Branco (2003), *A new class of multivariate skew distributions with applications to Bayesian regression models*, *The Canadian Journal of Statistics*, 31, 129-150.