Métodos de estimação: moda marginal a posteriori/marginal-perfilados bayesianos

Prof. Caio Azevedo

Histórico da estimação na TRI

- Características.
  - Número elevados de parâmetros para estimar.
  - A verificação das propriedades dos estimadores é muito complicada.
  - Falta de identificabilidade.
  - Necessidade de utilização de métodos numéricos.
  - Espaços paramétricos restritos.

Histórico da estimação na TRI

- Máxima Verossimilhança Marginal, MVM, (Bock and Lieberman, 1970).
- MVM via pseudo algoritmo EM (Bock and Aitkin, 1981).
- Moda marginal a posteriori via pseudo algoritmo EM (Mislevy, 1986).
- MCMC via dados aumentados (Albert, 1992).
- MCMC via algoritmo de MH (Patz and Junker, 1999).
- Algoritmo EM via expansão paramétrica (Rubin and Thomas, 2000).
- Algoritmo EM estocástico (SEM), (Fox, 2000).
- Algoritmo Robbins-Monro de Metropolis-Hastings (MHRM), Cai (2010).
- Algoritmo CADEM (contidional augmented data EM) (Azevedo and Andrade (2013)).

Modelo logístico unidimensional de 3 parâmetros

$$P(Y_{ij} = 1|(\theta_j, \zeta_i)) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

$i = 1, \ldots, I\,(item)$, $j = 1, \ldots, n\,(individuo)$,

- Y_{ij} : é a resposta do indivíduo j ao item i. É igual a 1 se o indivíduo responde corretamente e 0 caso scontrário.
- θ_j : é o traço latente (conhecimento, nível de depressão, etc) do indivíduo j.
- ζ_i : (a_i, b_i, c_i).
- a_i : é o parâmetro de discriminação associado ao item i.
- b_i : é o parâmetro de dificuldade associado ao item i.
- c_i : é o parâmetro de "acerto casual" associado ao item i.

- Seja θ um parâmetro associado a um determinado modelo $p(\boldsymbol{x}|\theta)$ verossimilhança para o qual assumimos uma priori $p(\theta)$.
- Posteriori: $p(\theta|\boldsymbol{x}) = \dfrac{p(\boldsymbol{x}|\theta)p(\theta)}{\int_{\Theta} p(\boldsymbol{x}|\theta)p(\theta)d\theta}$
- Problema (na TRI): é muito complicado obter as posterioris marginais de interesse, de forma analítica. Se $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ elas seriam dadas por $p(\theta_1|\boldsymbol{x})$ e $p(\theta_2|\boldsymbol{x})$.
- Uma solução maximizar a posteriori (marginal) dos parâmetros dos itens.

Estimação por Máxima Verossimilhança

- Maximizar a verossimilhança:

$$L(\boldsymbol{\theta}, \zeta) = \prod_{i=1}^{I} \prod_{j=1}^{n} P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}}$$

- Não identificabilidade

$$\begin{aligned} P_{ij} &= c_i + \frac{1 - c_i}{1 + e^{-Da_i\left(\theta_j - b_i\right)}} = c_i + \frac{1 - c_i}{1 + e^{-D\frac{a_i}{\alpha}\left(\frac{(\theta_j - \beta)}{\alpha} - \frac{(b_i - \beta)}{\alpha}\right)}} \\ &= c_i + \frac{1 - c_i}{1 + e^{-Da_i^*\left(\theta_j^* - b_i^*\right)}} \end{aligned}$$

- Estimação simultânea.
  - Inversão de matrizes da ordem de n x I.
  - Comprometimento das propriedades assintóticas dos estimadores.
- Alternativa : Estimação marginal perfilada.

- Considera-se uma distribuição de probabilidade para os traços latentes (não necessariamente no sentido bayesiano).
- Multiplica-se a verossimilhança original por essa densidade porposta e então integra-se com respeito aos traços latentes.
- Maximiza-se, então, essa verossimilhança marginal, com relação aos parâmetros dos itens.
- Suposição usual $\theta_j|\boldsymbol{\eta} \sim N(0,1)$, $\boldsymbol{\eta} = (\mu = 0, \psi = 1)$. Altamente questionável.

Construção da Verossimilhança Marginal

- Probabilidade Marginal de Respostal

$$
\begin{aligned}
P\left(\boldsymbol{Y}_{.j} = \boldsymbol{y}_{.j} | \zeta, \boldsymbol{\eta}\right) \equiv P\left(\boldsymbol{Y}_{.j} | \zeta, \boldsymbol{\eta}\right) &= \int_{\Re} P\left(\boldsymbol{Y}_{.j} = \boldsymbol{y}_{.j} | \theta, \zeta\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re} P\left(\boldsymbol{Y}_{.j} = \boldsymbol{y}_{.j} | \theta, \zeta\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re} P\left(\boldsymbol{Y}_{.j} | \theta, \zeta\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \,,
\end{aligned}
$$

em que $P\left(\boldsymbol{Y}_{.j} | \theta, \zeta\right) = \prod_{i=1}^{I} P_i^{y_{ij}} Q_i^{1-y_{ij}}$ e $\boldsymbol{\eta}$ é chamado de vetor de parâmetros populacionais.

Construção da Verossimilhança Marginal

- Verossimilhança marginal

$$
\begin{aligned}
L(\zeta, \boldsymbol{\eta}) &= \prod_{j=1}^{n} P\left(\boldsymbol{Y}_{.j} | \zeta, \boldsymbol{\eta}\right) = \prod_{j=1}^{n} \int_{\Re} P\left(\boldsymbol{Y}_{.j} | \theta, \zeta\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \\
&= \prod_{j=1}^{n} \int_{\Re} \prod_{i=1}^{I} P\left(\boldsymbol{Y}_{ij} | \theta, \zeta_i\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \, .
\end{aligned}
$$

- Posteriori marginal (considerando $\boldsymbol{\eta}$ conhecido)

$$
p(\zeta | \boldsymbol{y}, \boldsymbol{\eta}) = \frac{L(\zeta, \boldsymbol{\eta}) p(\zeta)}{\int L(\zeta, \boldsymbol{\eta}) p(\zeta) d\zeta} \qquad (1)
$$

- Log posteriori marginal

$$l(\zeta,\boldsymbol{\eta}) \;=\; \sum_{j=1}^{n} ln \int_{\Re} \prod_{i=1}^{I} P\left(\boldsymbol{Y}_{ij}|\theta,\zeta_i\right) g\left(\theta,\boldsymbol{\eta}\right) d\theta + \ln p(\zeta) + C \,.$$

- Estimadores moda marginal a posteriori

$$\begin{aligned}
\frac{\partial l\left(\zeta,\boldsymbol{\eta}\right)}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i}\left\{\sum_{j=1}^{n} \ln P\left(\boldsymbol{Y}_{.j}|\zeta,\boldsymbol{\eta}\right)\right\} \\
&= \sum_{j=1}^{n} \frac{1}{P\left(\boldsymbol{Y}_{.j}|\zeta,\boldsymbol{\eta}\right)} \frac{\partial P\left(\boldsymbol{Y}_{.j}|\zeta,\boldsymbol{\eta}\right)}{\partial \zeta_i} + \frac{\ln \partial p(\zeta)}{\partial \zeta} \,.
\end{aligned}$$

- Mas,

$$
\begin{aligned}
\frac{\partial P\left(\boldsymbol{Y}_{.j}|\zeta,\boldsymbol{\eta}\right)}{\partial \zeta_i} &= \frac{\partial}{\partial \zeta_i}\int_{\Re} P\left(\boldsymbol{Y}_{.j}|\theta,\zeta\right) g\left(\theta,\boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re}\left(\frac{\partial}{\partial \zeta_i} P\left(\boldsymbol{Y}_{.j}|\theta,\zeta\right)\right) g\left(\theta,\boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re}\left(\frac{\partial}{\partial \zeta_i}\prod_{h=1}^{I} P\left(\boldsymbol{Y}_{hj}|\theta,b_h\right)\right) g\left(\theta,\boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re}\left(\prod_{h\neq i}^{I} P\left(\boldsymbol{Y}_{hj}|\theta,b_h\right)\right)\left(\frac{\partial}{\partial \zeta_i} P\left(\boldsymbol{Y}_{ij}|\theta,\zeta_i\right)\right) g\left(\theta,\boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re}\left(\frac{\partial P\left(\boldsymbol{Y}_{ij}|\theta,\zeta_i\right)/\partial\zeta_i}{P\left(\boldsymbol{Y}_{ij}|\theta,\zeta_i\right)}\right) P\left(\boldsymbol{Y}_{.j}|\theta,\zeta\right) g\left(\theta,\boldsymbol{\eta}\right) d\theta\,.
\end{aligned}
$$

Além disso,

$$\begin{aligned}\frac{\partial P\left(Y_{ij}|\zeta_i,\theta\right)}{\partial\zeta_i} &= \frac{\partial}{\partial\zeta_i}\left(P_i^{y_{ij}}Q_i^{1-y_{ij}}\right)\\ &= y_{ij}P_i^{y_{ij}-1}\left(\frac{\partial P_i}{\partial\zeta_i}\right)Q_i^{1-y_{ij}} + P_i^{y_{ij}}\left(1-y_{ij}\right)Q_i^{-y_{ij}}\left(\frac{-\partial P_i}{\partial\zeta_i}\right)\\ &= \left[y_{ij}P_i^{y_{ij}-1}Q_i^{1-y_{ij}} - P_i^{y_{ij}}\left(1-y_{ij}\right)Q_i^{-y_{ij}}\right]\left(\frac{\partial P_i}{\partial\zeta_i}\right).\end{aligned}$$

Notemos que o termo entre colchetes vale 1 quando $y_{ij}=1$ e -1 quando $y_{ij}=0$, portanto, podemos reescrevê-lo como $(-1)^{y_{ij}+1}$. Com isso,

$$\frac{\partial P\left(Y_{ij}|\zeta_i,\theta\right)}{\partial\zeta_i} = (-1)^{y_{ij}+1}\left(\frac{\partial P_i}{\partial\zeta_i}\right).$$

Note agora que

$$\frac{(-1)^{y_{ij}+1}P_iQ_i}{P_i^{y_{ij}}Q_i^{1-y_{ij}}} = \begin{cases} Q_i, & \text{se } y_{ij}=1\\ -P_i, & \text{se } y_{ij}=0\end{cases} = [y_{ij}-P_{ij}].$$

- Dessa forma, temos que

$$\frac{1}{P\left(\boldsymbol{Y}_{ij}|\theta,\zeta_i\right)}\frac{\partial}{\partial\zeta_i}P\left(\boldsymbol{Y}_{ij}|\theta,\zeta_i\right)=\frac{(y_{ij}-P_i)}{P_iQ_i}\left(\frac{\partial P_i}{\partial\zeta_i}\right),$$

Logo,

$$\frac{\partial P\left(\boldsymbol{Y}_{.j}|\zeta,\boldsymbol{\eta}\right)}{\partial\zeta_i} = \int_{\Re}\left[\frac{(y_{ij}-P_i)}{P_iQ_i}\left(\frac{\partial P_i}{\partial\zeta_i}\right)\right]P\left(\boldsymbol{Y}_{.j}|\theta,\zeta\right)g\left(\theta,\boldsymbol{\eta}\right)d\theta$$

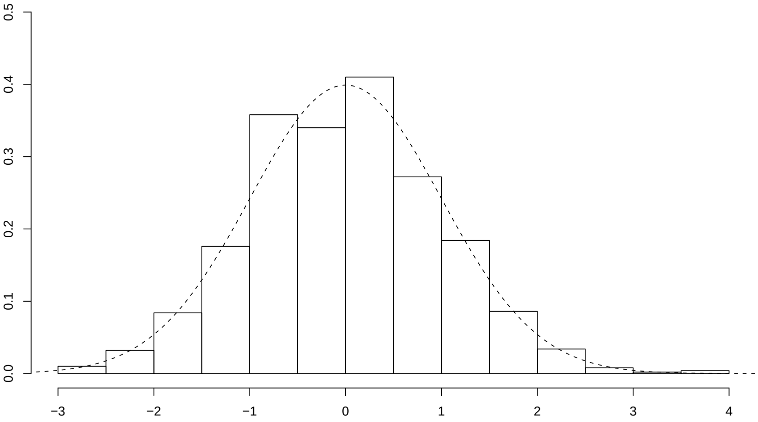
- Adicionalmente, vamos supor que $p(\zeta)=\prod_{i=1}^{I}p(a_i)p(b_i)p(c_i)$.

Portanto,

$$
\begin{aligned}
a_i &: \quad (1-c_i)\sum_{j=1}^{n}\int_{\Re}\left[(y_{ij}-P_i)(\theta-b_i)W_i\right]g_j^{*}(\theta)+\frac{\partial p(a_i)}{\partial a_i}=0\\
b_i &: \quad -a_i(1-c_i)\sum_{j=1}^{n}\int_{\Re}\left[(y_{ij}-P_i)W_i\right]g_j^{*}(\theta)+\frac{\partial p(b_i)}{\partial b_i}=0\\
c_i &: \quad \sum_{j=1}^{n}\int_{\Re}\left[(y_{ij}-P_i)\frac{W_i}{P_i^{*}}\right]g_j^{*}(\theta)+\frac{\partial p(c_i)}{\partial c_i}=0\,,
\end{aligned}
$$

em que,

$$
g_j^{*}(\theta)\equiv g\left(\theta|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}\right)=\frac{P\left(\boldsymbol{Y}_{.j}|\theta,\zeta\right)g\left(\theta|\boldsymbol{\eta}\right)}{P\left(\boldsymbol{Y}_{.j}|\zeta,\boldsymbol{\eta}\right)}.
$$


0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
4

Forma de quadratura

$$a_i \quad : \quad (1-c_i)\sum_{j=1}^{n}\sum_{l=1}^{q}\left[(y_{ij}-P_{il})\left(\overline{\theta}_l-b_i\right)W_{il}\right]g_j^*\left(\overline{\theta}_l\right)+\frac{\partial p(a_i)}{\partial \ln a_i}=0$$

$$b_i \quad : \quad -a_i\left(1-c_i\right)\sum_{j=1}^{n}\sum_{l=1}^{q}\left[(y_{ij}-P_{il})W_{il}\right]g_j^*\left(\overline{\theta}_l\right)+\frac{\partial \ln p(b_i)}{\partial b_i}=0$$

$$c_i \quad : \quad \sum_{j=1}^{n}\sum_{l=1}^{q}\left[(y_{ij}-P_{il})\frac{W_{il}}{P_{il}^*}\right]g_j^*\left(\overline{\theta}_l\right)+\frac{\partial \ln p(c_i)}{\partial c_i}=0$$

Equações de Mislevy

$$
\begin{aligned}
a_i &: \quad (1-c_i)\sum_{l=q}^{n}\left[\left(\overline{r}_{il}-\overline{f}_{il}P_{il}\right)\left(\overline{\theta}_l-b_i\right)W_{il}\right]+\frac{\partial \ln p(a_i)}{\partial a_i}=0 \\
b_i &: \quad -a_i\left(1-c_i\right)\sum_{l=1}^{q}\left[\left(\overline{r}_{il}-\overline{f}_{il}P_{il}\right)W_{il}\right]+\frac{\partial \ln p(b_i)}{\partial b_i}=0 \\
c_i &: \quad \sum_{l=1}^{q}\left[\left(\overline{r}_{il}-\overline{f}_{il}P_{il}\right)\frac{W_{il}}{P^*_{il}}\right]+\frac{\partial \ln p(c_i)}{\partial c_i}=0,
\end{aligned}
$$

em que

$$
\overline{r}_{il}=\sum_{j=1}^{n}y_{ij}g_j^*\left(\overline{\theta}_l\right) \quad , \quad \overline{f}_{il}=\sum_{j=1}^{n}g_j^*\left(\overline{\theta}_l\right).
$$

$$X_{(i)jl} = \begin{cases} 1, & \text{se o indivíduo } j \text{ responde (é submetido) ao} \\ & \text{item } i \text{ e possui habilidade em torno de } \overline{\theta}_l \\ 0, & \text{caso contrário.} \end{cases}$$

$$\mathcal{E}\left(F_{il}|\boldsymbol{Y}_{..},\zeta,\boldsymbol{\eta}\right) = \sum_{j=1}^{n} \mathcal{E}\left(X_{(i)jl}|\boldsymbol{Y}_{..},\zeta,\boldsymbol{\eta}\right) = \sum_{j=1}^{n} g_j^{*}\left(\overline{\theta}_l\right) = \overline{f}_{il}$$

e

$$\mathcal{E}\left(R_{il}|\boldsymbol{Y}_{..},\zeta,\boldsymbol{\eta}\right) = \sum_{j=1}^{n} \mathcal{E}\left(y_{ij}X_{(i)jl}|\boldsymbol{Y}_{..},\zeta,\boldsymbol{\eta}\right) = \sum_{j=1}^{n} y_{ij} g_j^{*}\left(\overline{\theta}_l\right) = \overline{r}_{il}.$$

Pseudo Algoritmo EM (versão bayesiana)

- Calcula a moda a posteriori na presença de dados faltantes (processo iterativo).
- Aplicação na TRI : considerar as proficiências como os dados não observados.
- Implementação do pseudo algoritmo EM

  Seja $p(\zeta|\boldsymbol{Y}_{..},\boldsymbol{\theta})$ a posteriori conjunta do dados completos . Se $\widehat{\zeta}^{(t)}$ é uma estimativa de ζ na iteração t, então os passos EM para obtenção de $\widehat{\zeta}^{(t+1)}$ são

  Passo E: Calcular $E[\ln p(\zeta|\boldsymbol{Y}_{..},\boldsymbol{\theta})|\boldsymbol{Y}_{..},\widehat{\zeta}^{(t)}]$

  Passo M: Obter $\widehat{\zeta}^{(t+1)}$ que maximiza a função do Passo E.
- No passo M a mazimização pode ser feita utilizando o algoritmo Newton-Raphson/Escore de Fisher.

- Considere uma população divididade em q categorias de proficiência e que dela se extrai uma amostra de tamanho n.
- Suponha que as proporções no item anterior são dadas por $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_q)'$.
- Denote por $\boldsymbol{f}_i$=$(f_{i1}, \cdots, f_{iq})'$ a quantidade de indivíduos em cada nível de habilidade e $\boldsymbol{r}_i$ =$(r_{i1}, \cdots, r_{iq})'$ a quantidade daqueles que respondem corretamente ao item i com nível de habilidade l, ambos observados na amostra. Além disso $\boldsymbol{r} = (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_I)'$.

- A probabilidade conjunta que os f_{il} indivíduos tenham habilidades $\overline{\theta}_l$, $l = 1, \cdots, q$, é dada pela distribuição multinomial:

$$P\left(\boldsymbol{F}_i = \boldsymbol{f}_i | \boldsymbol{\pi}\right) \equiv P(\boldsymbol{f}_i | \pi) = \frac{n_{(i)}!}{\prod_{l=1}^{q} f_{il}!} \prod_{l=1}^{q} \pi_j^{f_{il}}, \quad i = 1, \cdots, I,$$

- Dados f_{il} e $\overline{\theta}_l$, a probabilidade de ocorrerem r_{il} acertos ao item i dentre as f_{il} tentativas (respostas) por indivíduos com habilidade $\overline{\theta}_l$ é

$$P\left(R_{il} = r_{il} | f_{il}, \overline{\theta}_l\right) \equiv P(r_{il} | f_{il}, \overline{\theta}_l) = \binom{f_{il}}{r_{il}} P_{il}^{r_{il}} Q_{il}^{f_{il} - r_{il}},$$

- A probabilidade conjunta de $\boldsymbol{f}$ e $\boldsymbol{r}$, dados $\overline{\boldsymbol{\theta}} = (\overline{\theta}_1, \cdots, \overline{\theta}_q)'$ e π, é

$$
\begin{aligned}
P\left(\boldsymbol{F} = \boldsymbol{f}, \boldsymbol{R} = \boldsymbol{r} | \overline{\boldsymbol{\theta}}, \pi\right) & \equiv P(\boldsymbol{f}, \boldsymbol{r} | \overline{\boldsymbol{\theta}}, \pi) = P(\boldsymbol{r} | \boldsymbol{f}, \overline{\boldsymbol{\theta}}, \pi) P(\boldsymbol{f} | \overline{\boldsymbol{\theta}}, \pi) \\
& = P(\boldsymbol{r} | \boldsymbol{f}, \overline{\boldsymbol{\theta}}) P(\boldsymbol{f} | \pi) \\
& = \left\{ \prod_{i=1}^{I} \prod_{l=1}^{q} P(r_{il} | f_{il}, \overline{\theta}_l) \right\} \left\{ \prod_{i=1}^{I} P(\boldsymbol{f}_i | \pi) \right\}
\end{aligned}
$$

- Segue que a log posteriori para os dados completos é :

$$
\begin{aligned}
\ln L(\zeta) &= \ln P(\boldsymbol{f}|\pi) + \sum_{i=1}^{I}\sum_{l=1}^{q} \ln P(r_{il}|f_{il}, \overline{\theta}_l) \\
&= \ln P(\boldsymbol{f}|\pi) + \sum_{i=1}^{I}\sum_{l=1}^{q} \left\{ \ln \binom{f_{il}}{r_{il}} + r_{il} \ln P_{il} + (f_{il} - r_{il}) \ln Q_{il} \right\} \\
&= C + \sum_{l=1}^{q}\sum_{i=1}^{I} \{ r_{il} \ln P_{il} + (f_{il} - r_{il}) \ln Q_{il} \} + \ln p(\zeta),
\end{aligned}
$$

- Tomando a esperança da log posteriori, condicionada a $(\boldsymbol{Y}'_{..}, \zeta')'$, para os dados completos, temos que

$$E[\ln L(\zeta)|(\boldsymbol{Y}'_{..}, \zeta')'] = \overline{C} + \sum_{i=1}^{I} \sum_{l=1}^{q} \left\{ \overline{r}_{il} \ln P_{il} + (\overline{f}_{il} - \overline{r}_{il}) \ln Q_{il} \right\},$$

em que

$$\overline{r}_{il} = E[r_{il}|\boldsymbol{Y}_{..}, \zeta], \quad \overline{f}_{il} = E[f_{il}|\boldsymbol{Y}_{..}, \zeta] \quad \text{e} \quad \overline{C} = E[C|\boldsymbol{Y}_{..}, \zeta].$$

- Dessa forma, os passos E e M são :

Passo E

Usar os pontos de quadratura $\overline{\theta}_l$, os pesos A_l, $l = 1, \cdots, q$ e estimativas iniciais dos parâmetros dos itens, $\widehat{\zeta}_i$, $i = 1, \cdots, I$, para gerar $g_j^*(\overline{\theta}_l)$ e, posteriormente, $\overline{r}_{il}$ e $\overline{f}_{il}$, $i = 1, \cdots, I$ e $l = 1, \cdots, q$.

Passo M

Com $\boldsymbol{r}$ e $\boldsymbol{f}$ obtidos no Passo E, resolver as equações de estimação para ζ_i, $i = 1, \cdots, I$, usando o algoritmo de Newton-Raphson ou Escore de Fisher.

Abordagem hierárquica

Posteriori marginal

$$
\begin{aligned}
p(\zeta, \boldsymbol{\eta}|\boldsymbol{y}_{..}) \quad &\propto \quad \left\{ \prod_{j=1}^{n} \int_{\Re} \prod_{i=1}^{I} P\left(\boldsymbol{Y}_{ij}|\theta, \zeta_i\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \right\} \left\{ \prod_{i=1}^{I} p(\zeta_i|\boldsymbol{\tau}_i) \right\} \\
&\times \quad \left\{ \prod_{i=1}^{I} p(\boldsymbol{\tau}_i) \right\}
\end{aligned}
$$

Log posteriori marginal

$$
\begin{aligned}
\ln p(\zeta, \boldsymbol{\eta}|\boldsymbol{y}_{..}) \quad &\propto \quad \left\{ \sum_{j=1}^{n} \ln \int_{\Re} \prod_{i=1}^{I} P\left(\boldsymbol{Y}_{ij}|\theta, \zeta_i\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \right\} \left\{ \sum_{i=1}^{I} \ln p(\zeta_i|\boldsymbol{\tau}_i) \right\} \\
&\left\{ \sum_{i=1}^{I} \ln p(\boldsymbol{\tau}_i) \right\}
\end{aligned}
$$

Abordagem hierárquica

Maximizar a log posteriori marginal

$$
\begin{aligned}
\frac{\partial}{\partial \zeta_i} \ln p(\zeta, \boldsymbol{\eta} | \boldsymbol{y}_{..}) \quad &\propto \quad \left\{ \sum_{j=1}^{n} \frac{\partial}{\partial \zeta_i} \ln \int_{\Re} \prod_{i=1}^{I} P\left(\boldsymbol{Y}_{ij} | \theta, \zeta_i\right) g\left(\theta, \boldsymbol{\eta}\right) d\theta \right\} \\
&\times \quad \left\{ \frac{\partial}{\partial \zeta_i} \ln p(\zeta_i | \boldsymbol{\tau}_i) \right\}
\end{aligned}
$$

Parâmetro a

$$p(a_i|\mu_{a_i}, \sigma^2_{a_i}) = \frac{1}{\sqrt{2\pi} a_i \sigma_{a_i}} \exp\left[-\frac{1}{2\sigma^2_{a_i}}(\ln a_i - \mu_{a_i})^2\right].$$

Parâmetro b

$$p(b_i|\mu_{b_i}, \sigma^2_{b_i}) = \frac{1}{\sqrt{2\pi}\sigma_{b_i}} \exp\left\{\frac{-(b_i - \mu_{b_i})^2}{2\sigma^2_{b_i}}\right\}$$

Parâmetro c

$$p(c_i|\alpha_i, \beta_i) = \frac{\Gamma(\alpha_i + \beta_i - 2)}{\Gamma(\alpha_i - 1)\Gamma(\beta_i - 1)} c_i^{\alpha_i - 2}(1 - c_i)^{\beta_i - 2}.$$

$$\boldsymbol{S}(\zeta_i)_B = \sum_{l=1}^{q} \left(\overline{r}_{il} - \overline{f}_{il} P_{il}\right) W_{il} \boldsymbol{h}_{il} + \boldsymbol{\lambda}_i \,,$$

com

$$\boldsymbol{h}_{ij} = (P_{il}^{*} Q_{il}^{*})^{-1} \left(\frac{\partial P_{il}}{\partial \zeta_i}\right) = \begin{pmatrix} D\,(1 - c_i)\left(\overline{\theta}_l - b_i\right) \\ -D a_i\,(1 - c_i) \\ \dfrac{1}{P_{il}^{*}} \end{pmatrix}$$

$$\boldsymbol{\lambda}_i = \left[\frac{1}{a_i}\left[1 + \frac{\ln a_i - \mu_{a_i}}{\sigma_{a_i}^2}\right] ; -\frac{(b_i - \mu_{b_i})}{\sigma_{b_i}^2} ; \frac{\alpha_i - 2}{c_i} - \frac{\beta_i - 2}{1 - c_i}\right]' .$$

$$\frac{\partial^2}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \ln p(\boldsymbol{\zeta}, \boldsymbol{\eta} | \boldsymbol{y}_{..}) \propto \left\{ \sum_{j=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \ln \int_{\Re} \prod_{i=1}^{I} P\left(\boldsymbol{Y}_{ij} | \theta, \boldsymbol{\zeta}_i \right) g\left(\theta, \boldsymbol{\eta} \right) d\theta \right\}$$

$$\left\{ \frac{\partial^2}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} \ln p(\boldsymbol{\zeta}_i | \boldsymbol{\tau}_i) \right\}$$

$$\boldsymbol{I}(\zeta_i)_{BM} = \sum_{l=1}^{q} \overline{f}_{il} P_{il}^{*} Q_{il}^{*} \boldsymbol{h}_{il} \boldsymbol{h}_{il}^{'} - \boldsymbol{\Lambda}_i \, .$$

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \frac{\left[\sigma_{a_i}^2 + ln\, a_i - \mu_{a_i} - 1\right]}{\mu_{a_i}^2 \sigma_{a_i}^2} & . & . \\ 0 & -\frac{1}{\sigma_{b_i}^2} & . \\ 0 & 0 & -\left[\frac{\alpha_i - 2}{c_i^2}\right] + \frac{\beta_i - 2}{(1 - c_i)^2} \end{bmatrix} .$$

Máxima Verossimilhança Marginal - MVM:

- ⊕ Possui propriedades assintóticas: as estimativas dos parâmetros a_i, b_i e c_i são consistentes;
- ⊕ Uma vez estimados os parâmetros dos itens, pode-se estimar as habilidades através de métodos simples;
- ⊕ Permite resolver o problema de indeterminação (métrica) relativo ao modelo,uma vez que se atribui um parâmetro de escala e de posição para a distribuição das habilidades;
- ⊕ Permite caracterizar empiricamente a distribuição dos traços latentes;
- ⊖ Não está definido para itens com acerto total ou erro total;
- ⊖ É bastante trabalhoso computacionalmente;
- ⊖ Apresenta problemas na estimação do parâmetro c_i em alguns casos; deve ser usado somente com um número suficientemente grande de respondentes.

  Necessidade do estabelecimento de uma distribuição para θ;

Moda marginal a Posteriori - MMAP:

- ⊕ Definido para qualquer padrão de resposta;
- ⊕ Uma vez estimados os parâmetros dos itens, pode-se estimar as habilidades através de métodos simples;
- ⊕ Permite resolver o problema de indeterminação (métrica) relativo ao modelo,uma vez que se atribui um parâmetro de escala e de posição para a distribuição das habilidades;
- ⊕ Permite caracterizar empiricamente a distribuição das habilidades;
- ⊖ É mais trabalhoso computacionalmente do que o MVM;

  Necessidade de distribuições a priori para os parâmetros dos itens.

- É possível atualizar a distribuição dos traços latentes.
- Método de máxima verossimilhança não-paramétrica (Mislevy (1982)).
- Utiliza as quantidades $\overline{f}_{il}$ devidamente padronizadas.
- Especificar uma priori em forma de histograma.
- Mistura de normais.
- Os erros-padrão pode ser obtidos através da inversa da Informação de Fisher. Para outras opções veja o manual do mirt.

De posse das estimativas dos parâmetros dos itens constroi-se uma verossimilhança perfilada para estimar as proficiências

$$L(\boldsymbol{\theta},\widehat{\boldsymbol{b}}) = P(\boldsymbol{Y}_{.j}|\theta_j,\widehat{\boldsymbol{\zeta}}) = \prod_{i=1}^{I}\prod_{j=1}^{n}\widehat{P}_{ij}^{y_{ij}}\widehat{Q}_{ij}^{1-y_{ij}},$$

Log posteriori perfilada

$$\ln g_j^*(\theta_j) = Const + \ln P(\boldsymbol{Y}_{.j}|\theta_j,\widehat{\boldsymbol{\zeta}}) + \ln g(\theta_j|\boldsymbol{\eta}).$$

Equação de estimação bayesiana

$$\frac{\partial \ln g_j^*(\theta_j)}{\partial\theta_j} = \frac{\partial \ln P(\boldsymbol{Y}_{.j}|\theta_j,\widehat{\boldsymbol{\zeta}})}{\partial\theta_j} + \frac{\partial \ln g(\theta_j|\boldsymbol{\eta})}{\partial\theta_j} = 0.$$

Informação de Fisher

$$I(\theta_j) = \sum_{i=1}^{I} \widehat{P}_{ij}^{*} \widehat{Q}_{ij}^{*} \widehat{W_{ij}} \widehat{h}_{ij}^{2} - \frac{1}{\sigma^2}.$$

Algoritmo Escore de Fisher

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} + I(\theta_j^{(t)})^{-1} S(\theta_j^{(t)})$$

em que $W_{ij} = \frac{P_{ij}^{*} Q_{ij}^{*}}{P_{ij} Q_{ij}}$ e $h_{ij} = a_i(1 - c_i)$

Erros-padrão: inversa da informação de Fisher.

$$g(\theta|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}) = \frac{P(\boldsymbol{y}_{.j}|\theta,\zeta)g(\theta|\boldsymbol{\eta})}{P(\boldsymbol{y}_{.j}|\zeta,\boldsymbol{\eta})}.$$

Segue que a esperança da posteriori é

$$\widehat{\theta_j} \equiv E[\theta|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}] = \frac{\int_{\Re} \theta g(\theta|\boldsymbol{\eta})P(\boldsymbol{y}_{.j}|\theta,\zeta)d\theta}{\int_{\Re} g(\theta|\boldsymbol{\eta})P(\boldsymbol{y}_{.j}|\theta,\zeta)d\theta}.$$

$$\begin{aligned}
\mathcal{E}\left[\theta_j|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}\right] &\approx \mathcal{E}\left[\overline{\theta}_j|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}\right] = \frac{\sum_{l=1}^{q}\overline{\theta}_l P(\boldsymbol{Y}_{.j}|\overline{\theta}_l,\zeta)g(\overline{\theta}_l|\boldsymbol{\eta})}{\sum_{l=1}^{q} P(\boldsymbol{Y}_{.j}|\overline{\theta}_l,\zeta)g(\overline{\theta}_l|\boldsymbol{\eta})} \\
&= \frac{\sum_{l=1}^{q}\overline{\theta}_l P(\boldsymbol{Y}_{.j}|\overline{\theta}_l,\zeta)A_l}{\sum_{l=1}^{q} P(\boldsymbol{Y}_{.j}|\overline{\theta}_l,\zeta)A_l}.
\end{aligned}$$

E a variância a posteriori

$$Var\left(\theta_j\right) \equiv Var[\theta|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}] = \frac{\int_{\Re}\left(\theta - E\left(\theta\right)\right)^2 g(\theta|\boldsymbol{\eta})P(\boldsymbol{y}_{.j}|\theta,\zeta)d\theta}{\int_{\Re} g(\theta|\boldsymbol{\eta})P(\boldsymbol{y}_{.j}|\theta,\zeta)d\theta}.$$

$$Var\left[\overline{\theta}_j|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}\right] = \frac{\sum_{l=1}^{q}\left\{\overline{\theta}_l - \mathcal{E}\left[\overline{\theta}_j|\boldsymbol{y}_{.j},\zeta,\boldsymbol{\eta}\right]\right\}^2 P(\boldsymbol{Y}_{.j}|\overline{\theta}_l,\zeta)g(\overline{\theta}_l|\boldsymbol{\eta})}{\sum_{l=1}^{q} P(\boldsymbol{Y}_{.j}|\overline{\theta}_l,\zeta)g(\overline{\theta}_l|\boldsymbol{\eta})}$$

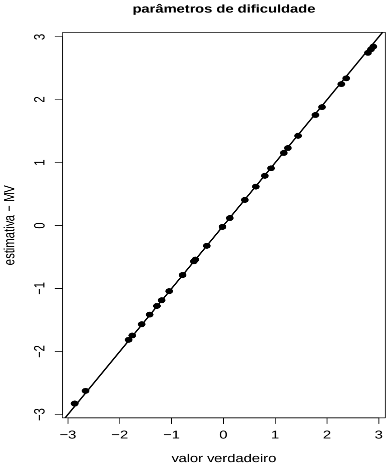
Erros-padrão: raiza quadrada da variância a posteriori.

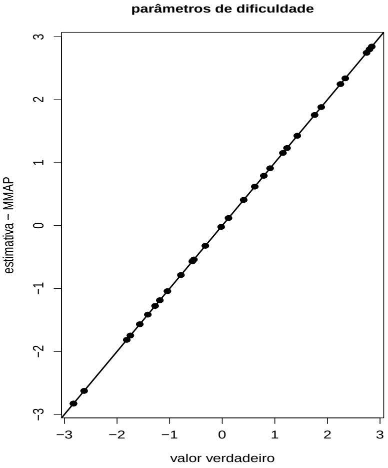
- Máxima Verossimilhança - MV :
  - $\oplus$ Para testes “longos” produz estimadores não viciados;
  - $\ominus$ Não está definido para alguns padrões de resposta.
- Bayesiano - EAP :
  - $\oplus$ Definido para qualquer padrão de resposta;
  - $\oplus$ Possui o menor erro médio;
  - $\ominus$ Viciado;
  - $\ominus$ Exige cálculos mais complexos do que o método de MV;
  - $\ominus$ Necessidade de uma distribuição a priori para θ.

- Bayesiano - MAP :

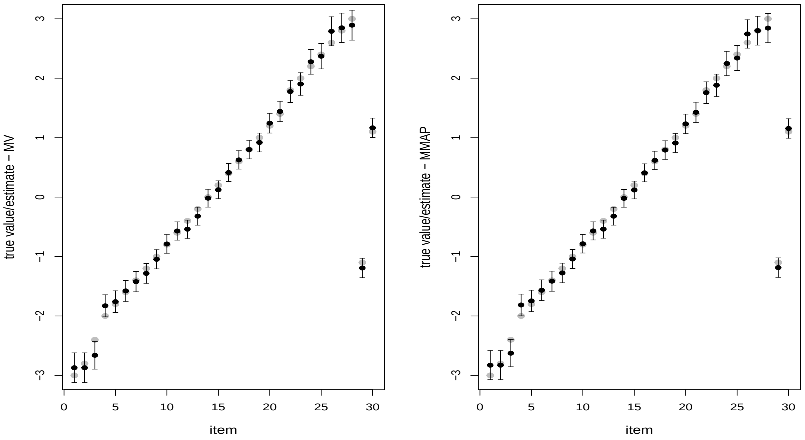
  - ⊕ Definido para qualquer padrão de resposta;
  - ⊖ Viciado.
  - ⊖ Exige cálculos mais complexos do que o método de MV;
  - ⊖ Necessidade de uma distribuição a priori para θ.

Resultados: modelo 1

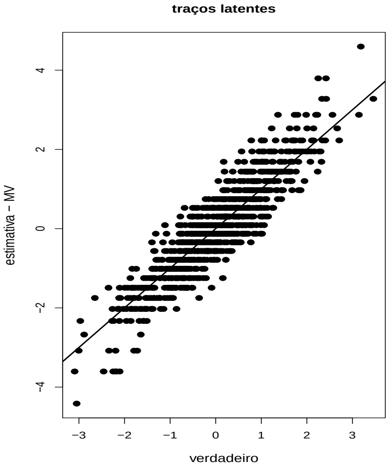


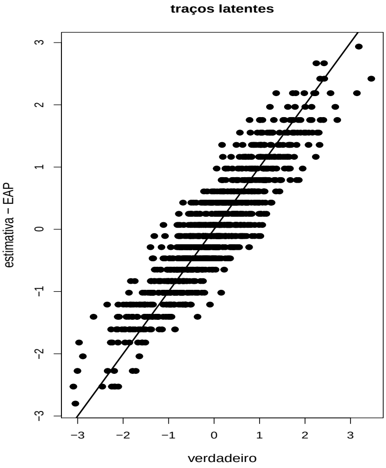


Resultados: modelo 1

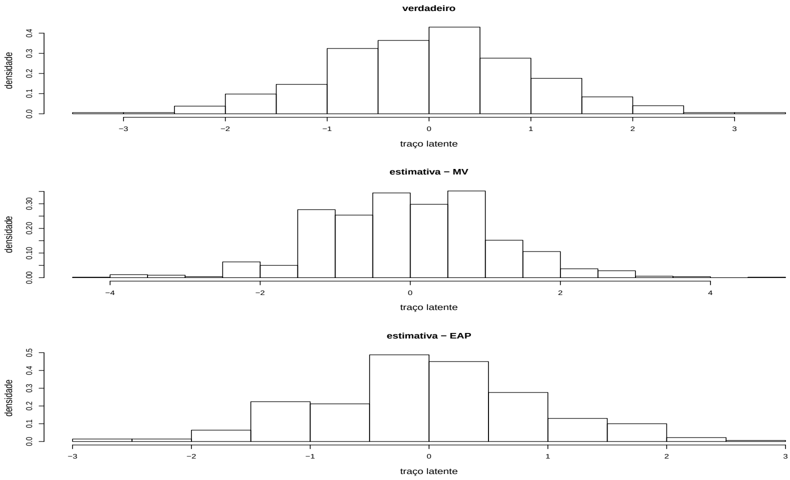


Resultados: modelo 1

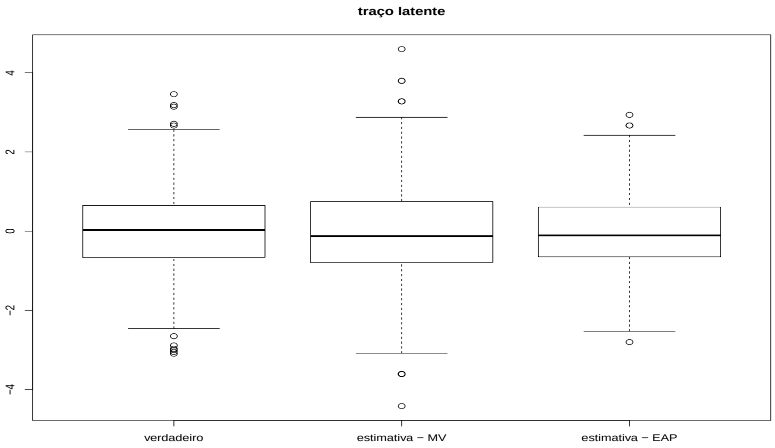




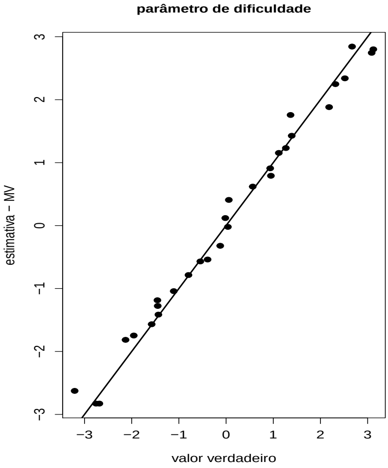
Resultados: modelo 1

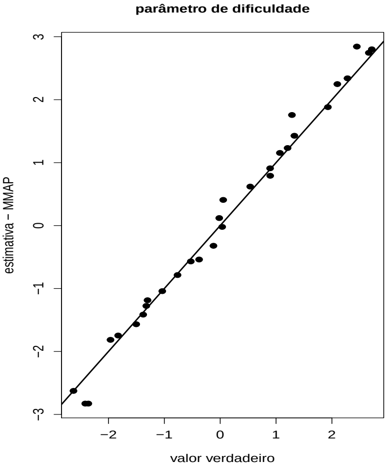


Resultados: modelo 1

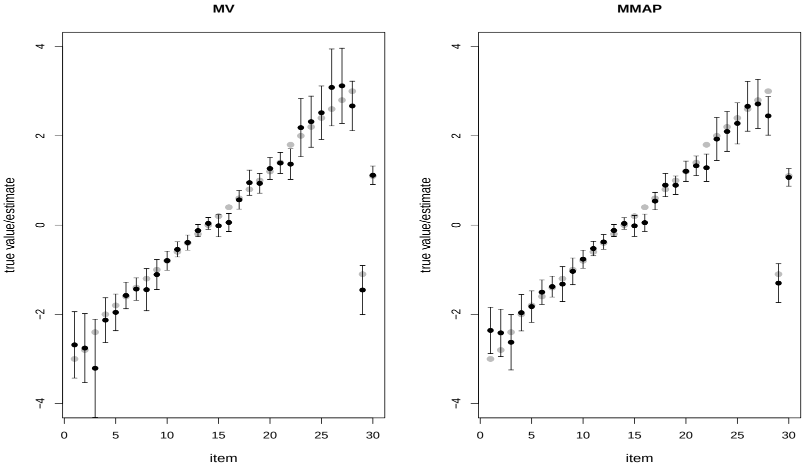


Resultados: modelo 2

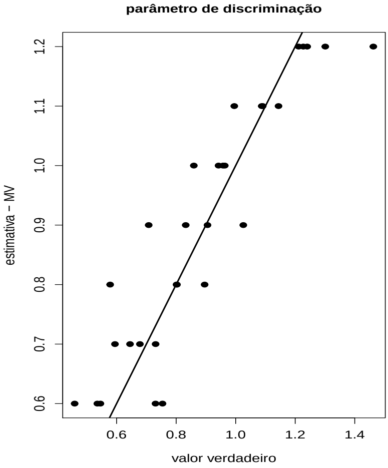


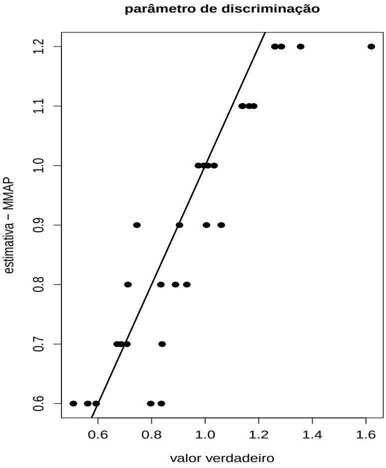


Resultados: modelo 2

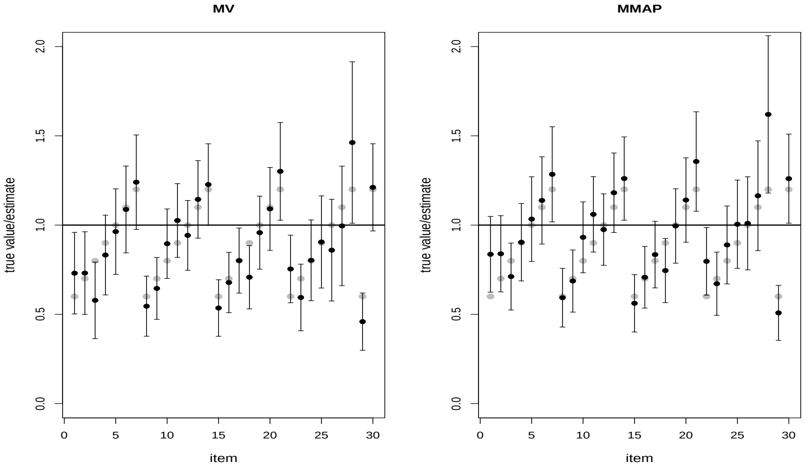


Resultados: modelo 2

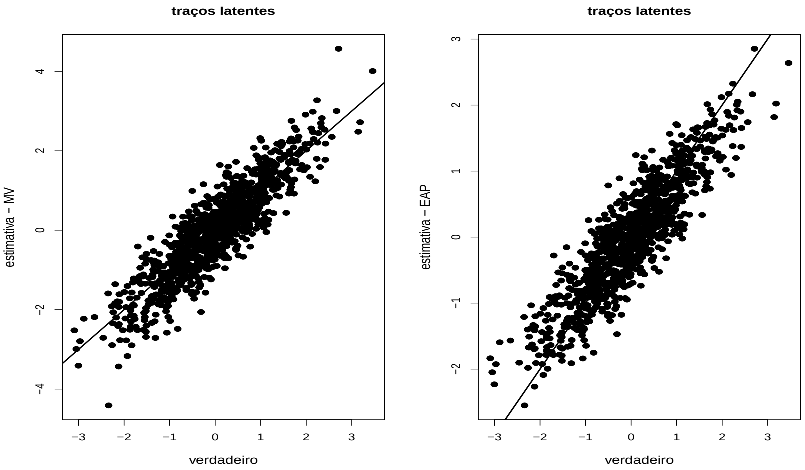




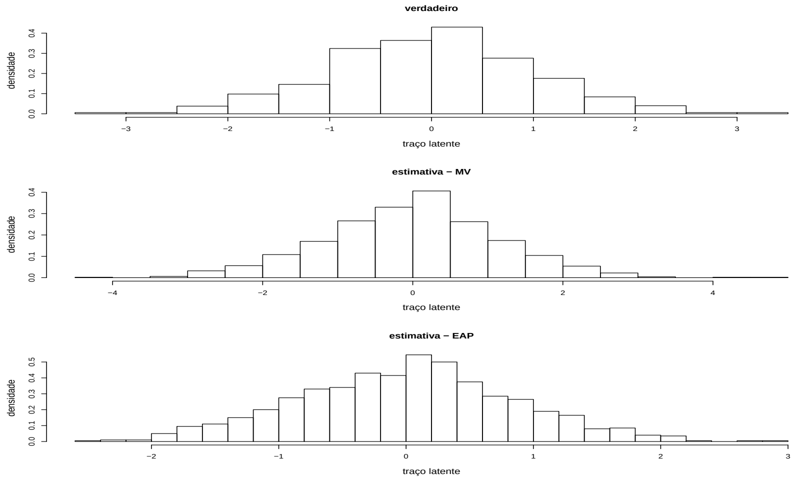
Resultados: modelo 2



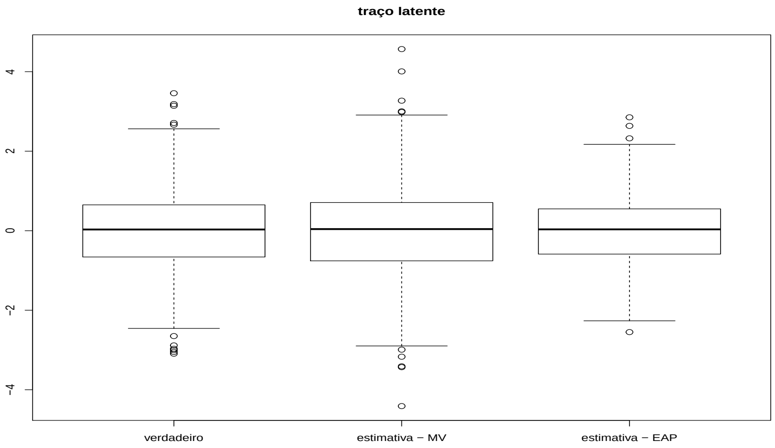
Resultados: modelo 2



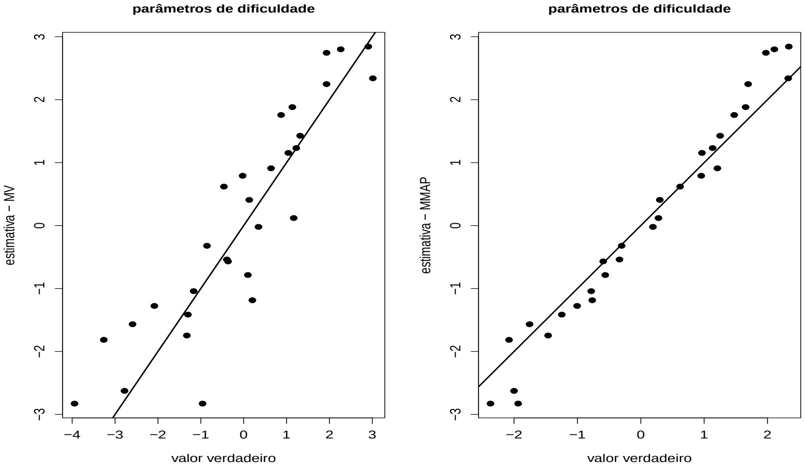
Resultados: modelo 2



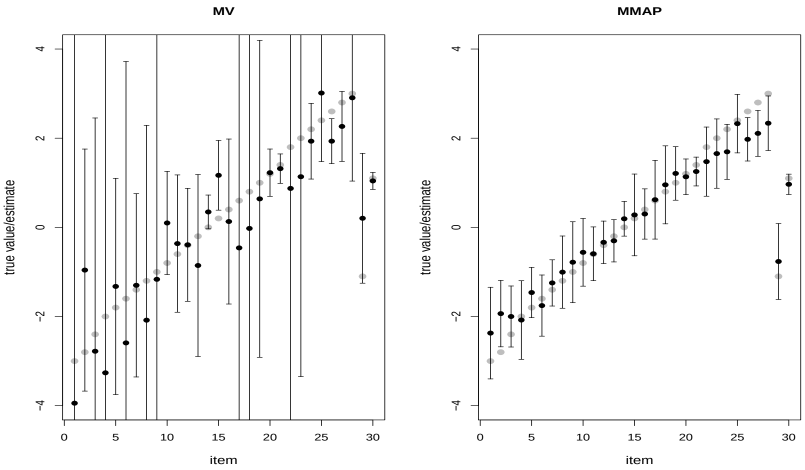
Resultados: modelo 2



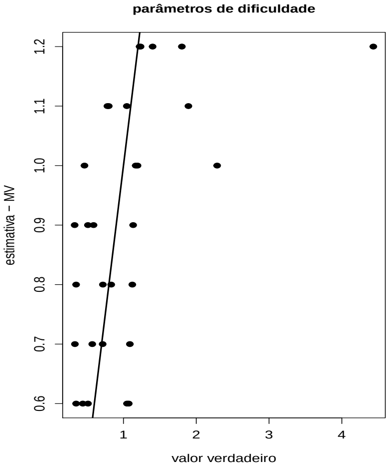
Resultados: modelo 3

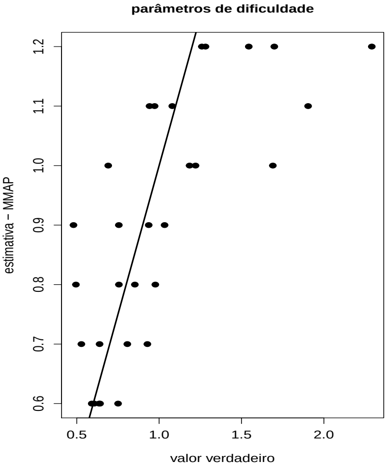


Resultados: modelo 3

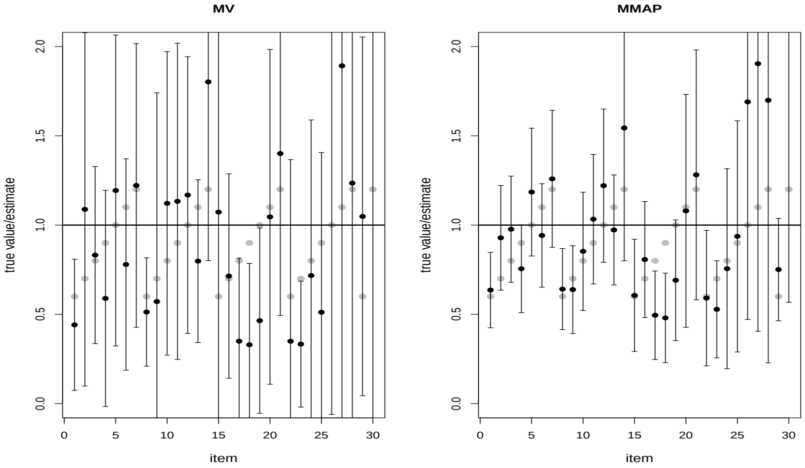


Resultados: modelo 3

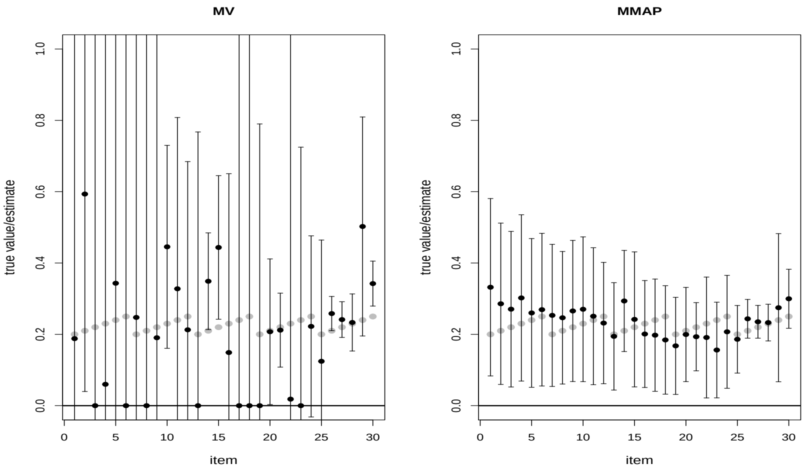




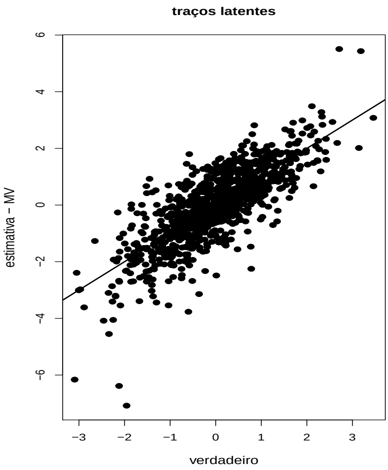
Resultados: modelo 3

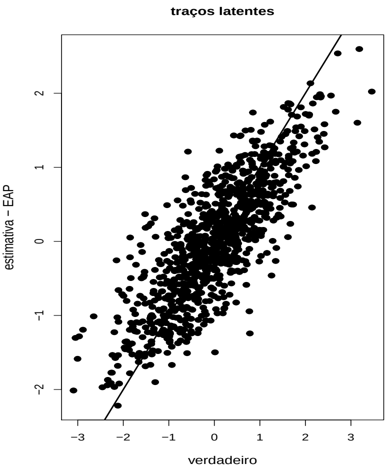


Resultados: modelo 3

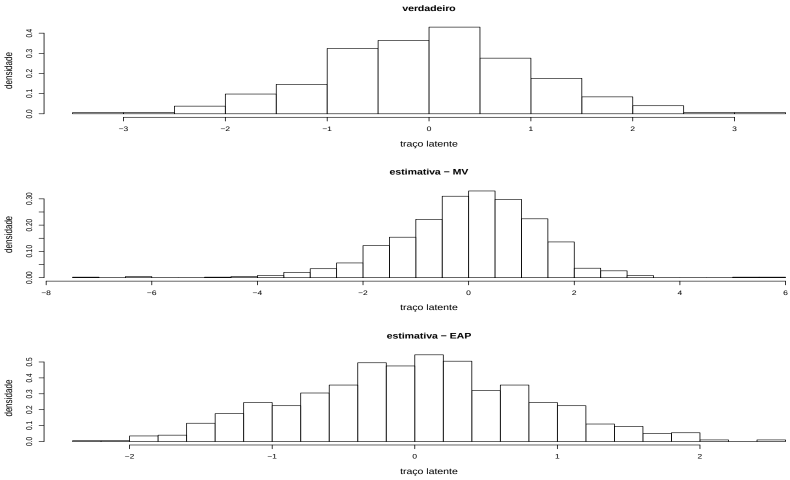


Resultados: modelo 3





Resultados: modelo 3



Resultados: modelo 3