Modelos politômicos (para respostas politômicas)

Prof. Caio Azevedo

Aspectos Gerais

- Os modelos policotômicos visam analisar itens que apresentam mais de dois escores possíveis (categorias).
- Basicamente se dividem em duas classes
  - Nominais : as categorias não possuem ordenação entre si - itens de múltipla escolha.
  - Ordinais : as categorias possuem ordenação entre si - itens abertos aos quais são atribuídos algum escore (intervalo $\{0, 1, ..., R\}$, $R \in \mathbb{Z}^+$, escala de Likert etc).

Exemplo de item nominal

QUESTÃO 16

Seja $f(x)$ uma função tal que para todo número real x temos que $xf(x-1) = (x-3)f(x) + 3$. Então, $f(1)$ é igual a

a) 0.

b) 1.

c) 2.

d) 3.

http://www.comvest.unicamp.br/wp-content/uploads/2017/02/

Exemplo de item gradual

14. Sejam c um número real e $f(x) = x^2 - 4x + c$ uma função quadrática definida para todo número real x. No plano cartesiano, considere a parábola dada pelo gráfico de $y = f(x)$.

a) Determine c no caso em que a abscissa e a ordenada do vértice da parábola têm soma nula e esboce o respectivo gráfico para $0 \leq x \leq 4$.

b) Considere os pontos de coordenadas $A = (a, f(a))$ e $B = (b, f(b))$, onde a e b são números reais com $a < b$. Sabendo que o ponto médio do segmento $\overline{AB}$ é $M = (1, c)$, determine a e b.

http://www.comvest.unicamp.br/wp-content/uploads/2017/02/

Examplo de item gradual: escala de likert

Qual é o nível de importância do trabalho em equipe para seu empregador?

- Extremamente importante
- Muito importante
- Um pouco importante
- Pouco importante
- Não é importante

https://help.surveymonkey.com/articles/pt_BR/kb/Likert-Scales

Exemplo de item gradual: escala de likert (altura)

Na cama, eu frequentemente sinto frio nos pés.

- Concordo totalmente
- Concordo parcialmente
- Discordo parcialmente
- Discordo totalmente

Modelo de Resposta Nominal - MRN

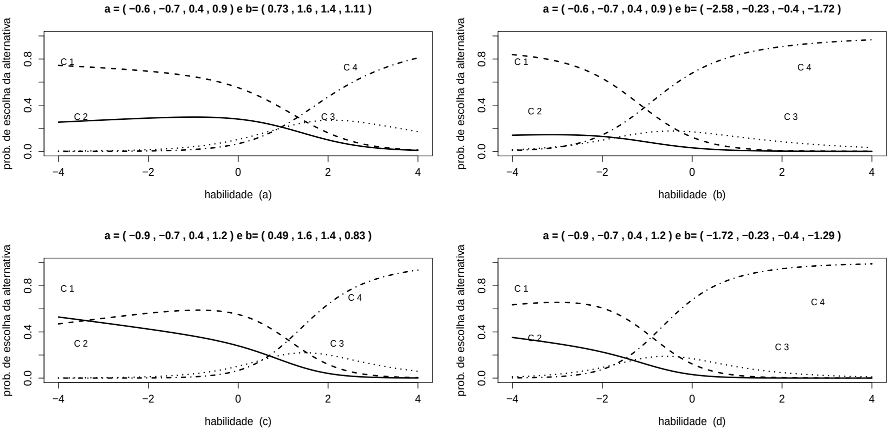
$$P_{ijs} = P\left(Y_{ijs} = 1 | \theta_j, \zeta_i\right) = \frac{e^{a_{is}(\theta_j - b_{is})}}{\displaystyle\sum_{h=1}^{m_i} e^{a_{ih}(\theta_j - b_{ih})}},$$

- $Y_{ijs} = 1$ se o indivíduo j, $(j = 1, 2, .., n)$, escolhe a categoria s, $(s = 1, 2, ..., m_i)$ do item i, $(i = 1, 2, ..., I)$ e 0, caso contrário $(\sum_{h=1}^{m_i} Y_{ijs} = 1$, as categorias são exaustivas e mutuamente excludentes)
- θ_j : traço latente do indivíduo j.

Modelo de Resposta Nominal - MRN

- a_{is} : está associado à discriminação da categoria s.
- b_{is} : está relacionado à dificuldade dsa categoria s.
- $\zeta_i = (a_1, b_{i1}, ..., b_{im_i})'$. Parâmetros resumo: Pereira, S. R. S. (2012) Contribuições ao Estudo do Modelo de Resposta Nominal, Dissertação de Mestrado, IMECC, Unicamp, `http://repositorio.unicamp.br/handle/REPOSIP/306794`.
- Apresentar exemplos de valores para os parâmetros resumo.

Representação gráfica do MRN



Modelo de Resposta Gradual - MRG

- Seja $Y_{ijs} = 1$ se o indivíduo j, ($j = 1, 2, .., n$), escolhe a categoria s, ($s = 0, 1, 2, ..., m_i$) do item i, ($i = 1, 2, ..., I$) e 0, caso contrário ($\sum_{h=1}^{m_i} Y_{ijs} = 1$, as categorias são exaustivas e mutuamente excludentes)
- Probabilidade do indivíduo j escolher (receber o escore s) a categoria s ou outra (outro escore) de maior valor:

$$P_{i,s}^{+}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{is})}};$$

Modelo de Resposta Gradual - MRG

em que (de modo que $\sum_{h=1}^{m_i} P_{ijs} = 1$, P_{ijs} é probabilidade do indivíduo j escolher a categoria s do item i):

$$P_{ijs} = P_{i,s}(\theta_j) = P^{+}_{i,s}(\theta_j) - P^{+}_{i,s+1}(\theta_j)$$

$$P_{i,0}(\theta_j) = P^{+}_{i,0}(\theta_j) - P^{+}_{i,1}(\theta_j) = 1 - P^{+}_{i,1}(\theta_j)$$

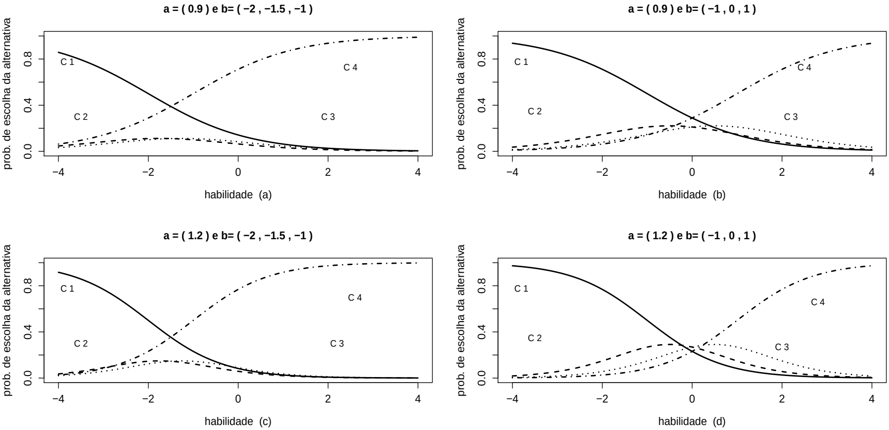
$$P_{i,m_i}(\theta_j) = P^{+}_{i,m_i}(\theta_j) - P^{+}_{i,m_i+1}(\theta_j) = P^{+}_{i,m_i}(\theta_j)$$

cont.

$$P_{ijs} = P(Y_{ijs} = 1|\theta_j, \zeta_i) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{is})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i(s+1)})}}$$

- θ_j : traço latente do indivíduo j.
- a_i : está associado à discriminação do item i.
- b_{is} : está relacionado à dificuldade dsa categoria s.
- $\zeta_i = (a_1, b_{i1}, ..., b_{im_i})$.

Representação gráfica do modelo de resposta gradual



Estimação

- Em geral, há muito mais parâmetros para se estimar do que nos modelos dicotômicos.
- MRN : $2m_i - 2 + n$ (restrições de identificabilidade: $\sum_{h=1}^{m_i} a_{is} = \sum_{h=1}^{m_i} d_{is} = 0$, $d_{is} = -a_{is} b_{is}$.)
- MRG : $m_i + 1 + n$ (restrições de identificabilidade: $b_{i0} = 0$).
- Adicionalmente, $\theta \sim D(\eta_\theta)$, em que $\mathcal{E}(\theta) = 0$, $\mathcal{V}(\theta) = 1$.

Estimação

- Maior número de iterações requeridas nos processos iterativos.
- Mais pontos de quadratura para se obter estimativas acuradas.
- Espaço paramétrico, em geral, apresenta menos restrições.
  - MRN : sem restrições.
  - MRG : $a_i > 0$ e $b_{is} \in \mathbb{R}, b_{i1} \leq \ldots \leq b_{im_i}$ (além das restrições de identificabilidade).

Estimação por máxima verossimilhança

- Para cada indivíduo/item (verossimilhança multinomial (Bernoulli multivariada) (r = 0,1):
  $P(Y_{ijr} = y_{ijr}, ..., Y_{ijm_i} = y_{ijm_i}|\theta_j, \zeta_i) = \prod_{s=r}^{m_i} P_{ijs}^{y_{ijs}}$.
- Maximizar a verosimilhança (conjunta)

$$L(\boldsymbol{\theta}, \zeta) = \prod_{j=1}^{n}\prod_{i=1}^{I}\prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}.$$

- Problemas em se maximizar a verossilhança genuína (essencialmente os mesmos relativos aos modelos dicotômicos).
- Alternativa : maximizar a verossimilhança marginal.

Estimação por máxima verossimilhança marginal

- log-verossimilhança marginal

$$
\begin{aligned}
l(\zeta, \boldsymbol{\eta}) &= \sum_{j=1}^{n} ln \int_{\Re} P(\boldsymbol{Y}_{.j.}|\zeta, \theta)\, g(\theta|\boldsymbol{\eta})\, d\theta \\
&= \sum_{j=1}^{n} lnP(\boldsymbol{Y}_{.j.}|\zeta, \boldsymbol{\eta}) \; .
\end{aligned}
$$

- Função escore

$$
\boldsymbol{S}(\zeta_i) = \sum_{j=1}^{n} \left\{ \frac{1}{P(\boldsymbol{Y}_{.j.}|\zeta, \boldsymbol{\eta})} \left(\frac{\partial}{\partial \zeta_i} P(\boldsymbol{Y}_{.j.}|\zeta, \boldsymbol{\eta}) \right) \right\} ,
$$

Estimação por máxima verossimilhança marginal - cont.

mas,

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

com
$\prod_{i=1}^{I} P_{ij.}^{y_{ij.}} = \prod_{i=1}^{I}\prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}$ e $P_{ijs} = P(Y_{ijs} = 1|\zeta_i, \theta)$.

Estimação por máxima verossimilhança marginal - cont.

Seguindo, temos que

$$
\begin{aligned}
\frac{\partial}{\partial \zeta_i} P\left(\boldsymbol{Y}_{.j.} | \zeta, \boldsymbol{\eta}\right) &= \int_{\Re} \left\{ \prod_{i' \neq i} Pij.^{y_{ij.}} \right\} \left\{ \frac{\partial P_{ij.}^{y_{ij.}}}{\partial \zeta_i} \right\} g\left(\theta | \boldsymbol{\eta}\right) d\theta \\
&= \int_{\Re} \left\{ \prod_{i=1}^{I} P_{ij.}^{y_{ij.}} \right\} \left\{ \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}} \right\}^{-1} \left\{ \frac{\partial P_{ij.}^{y_{ij.}}}{\partial \zeta_i} \right\} g\left(\theta | \boldsymbol{\eta}\right) d\theta \, .
\end{aligned}
$$

Por outro lado (y_s, a_s, ϕ) são quantidades genéricas.

$$
\left\{ \prod_s y_s^{a_s} \right\}^{-1} \frac{\partial}{\partial \phi} \left\{ \prod_s y_s^{a_s} \right\} = \sum_s \left\{ \frac{\partial y_s}{\partial \phi} \right\} \left\{ \frac{a_s}{y_s} \right\}, \quad \text{com} \quad y_s = f\left(\phi\right) \, .
$$

Estimação por máxima verossimilhança marginal - cont.

Dessa forma, temos

$$\left\{\prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}\right\}^{-1}\left\{\frac{\partial}{\partial \zeta_i}\left[\prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}}\right]\right\} = \sum_{s=1}^{m_i}\left\{\frac{\partial P_{ijs}}{\partial \zeta_i}\right\}\left\{\frac{y_{ijs}}{P_{ijs}}\right\}.$$

Assim,

$$\frac{\partial}{\partial \zeta_i} P\left(\boldsymbol{Y}_{.j.}|\zeta,\boldsymbol{\eta}\right) = \int_{\Re} P\left(\boldsymbol{Y}_{.j.}|\zeta,\theta\right) g\left(\theta|\boldsymbol{\eta}\right)\left\{\sum_{s=1}^{m_i}\left[\frac{y_{ijs}}{P_{ijs}}\left(\frac{\partial P_{ijs}}{\partial \zeta_i}\right)\right]\right\} d\theta\,.$$

Estimação por máxima verossimilhança marginal - cont.

Desse modo, vem que

$$\boldsymbol{S}\left(\zeta_i\right) = \sum_{j=1}^{n}\left\{\int_{\Re} g_j^*\left(\theta\right)\left\{\sum_{s=1}^{m_i}\left[\frac{y_{ijs}}{P_{ijs}}\left(\frac{\partial P_{ijs}}{\partial \zeta_i}\right)\right]\right\}d\theta\right\}.$$

Em que $g_j^*\left(\theta\right) = \dfrac{P\left(\boldsymbol{Y}_{.j.}|\theta,\boldsymbol{\eta}\right)g\left(\theta|\boldsymbol{\eta}\right)}{\int_{\Re} P\left(\boldsymbol{Y}_{.j.}|\theta,\boldsymbol{\eta}\right)g\left(\theta|\boldsymbol{\eta}\right)}$ o qual compõe os **dados artificiais**.

- Resolução do sistemas de equações de estimação : adaptação do (Pseudo) **algoritmo EM**.

Desenvolvimento da verossimilhança - Algoritmo EM

- 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 $\boldsymbol{r}_i = (\boldsymbol{r}_{i1}, \cdots, \boldsymbol{r}_{iq})'$, $\boldsymbol{r}_{ik} = (r_{ik1}, ..., r_{ikm_i})'$ a quantidade daqueles que escolhem a categoria $s, s = 1, .., m_i$, no item i com nível de traço latente l, ambos observados na amostra. Além disso $\boldsymbol{r} = (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_I)'$.

Desenvolvimento da verossimilhança - Algoritmo EM

- 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_{ilh} escolhas da categoria h, ao item i dentre as f_{il} tentativas (respostas) por indivíduos com traço latente $\overline{\theta}_l$ (multinomial($f_{il}, \boldsymbol{P}_{il}$)), em que $\boldsymbol{P}_{il} = (P_{il1}, ..., P_{ilm_i})'$, é dada por

$$P\left(\boldsymbol{R}_{il} = \boldsymbol{r}_{il} | f_{il}, \overline{\theta}_l\right) \equiv P(r_{il}|f_{il}, \overline{\theta}_l) = \frac{f_{il}}{\prod_{s=1}^{m_i} r_{ils}!} \prod_{s=1}^{m_i} P_{ils}^{r_{ils}},$$

Desenvolvimento da verossimilhança - Algoritmo EM

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

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

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

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

Desenvolvimento da verossimilhança - Algoritmo EM

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

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

em que

$$\overline{r}_{ils} = E[r_{ils}|\boldsymbol{Y}_{..}, \zeta] = \sum_{j=1}^{n} y_{ijs} g_j^*(\overline{\theta}), \quad \overline{f}_{il} = E[f_{il}|\boldsymbol{Y}_{..}, \zeta] = \sum_{j=1}^{n} g_j^*(\overline{\theta})$$

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

lembrando que $\sum_{s=1}^{m_i}\overline{r}_{ils} = \overline{f}_{il}$ e

$$g_j^*(\overline{\theta}) = \frac{P(\boldsymbol{Y}_{.j.}|\theta_l, \boldsymbol{\eta})\, g(\overline{\theta}_l|\boldsymbol{\eta})}{\sum_{l=1}^{q} P(\boldsymbol{Y}_{.j.}|\theta_l, \boldsymbol{\eta})\, g(\overline{\theta}_l|\boldsymbol{\eta})}.$$

Aplicação do Pseudo algoritmo EM

A proposta de Bock constitui-se no seguinte procedimento

Passo E

Usar os pontos de quadratura $\overline{\theta}_l$, os pesos $A_l^{(t)}$, $l = 1, \cdots, q$ e estimativas provisórias dos parâmetros dos itens, $\widehat{\zeta}_i^{(t)}$, $i = 1, \cdots, I$, para gerar $g_j^*(\overline{\theta}_l)^{(t)}$ e, posteriormente, $\overline{r}_{ils}^{(t)}$ e $\overline{f}_{il}^{(t)}$, $i = 1, \cdots, I$, $l = 1, \cdots, q$ e $s = 0, 1, \ldots, m_i$.

Passo M

Com $\overline{\boldsymbol{r}}^{(t)}$ e $\overline{\boldsymbol{f}}^{(t)}$ 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.

Expressões para o MRN

- Inserir as restrições de identificabilidade dos itens (slide 13). Há várias formas de fazê-lo.
- Define-se os parâmetros irrestritos ($\mathbf{\Gamma}_i$), $\alpha_{ih} = a_{i1} - a_{i(h+1)}$, $\delta_{ih} = d_{i1} - d_{i(h+1)}$, com base nos parâmetros transformados (ζ_i^*), $d_{is} = -a_{is} b_{is}$.
- As derivadas necessárias são calculadas em função de $\mathbf{\Gamma}_i = \zeta_i^* \boldsymbol{S}_i$.
- Mais detalhes veja Azevedo, C. L. N. (2003). Métodos de estimação na Teoria da Resposta ao Item. Dissertação de Mestardo, IMECC, Unicamp.

  `http://www.ime.unicamp.br/~cnaber/Disser_METRI.pdf`

Expressões para o MRN

- Vetor escore

$$\boldsymbol{S}(\boldsymbol{\Gamma}_i) = \sum_{l=1}^{q} \left\{ \begin{bmatrix} 1 \\ \overline{\theta}_l \end{bmatrix} \otimes \boldsymbol{T}_i \left[\overline{\boldsymbol{r}}_{il.} - \overline{f}_{il} \boldsymbol{P}_{il.} \right] \right\}$$

- Matriz Hessiana (= - Informação de Fisher)

$$\boldsymbol{H}(\boldsymbol{\Gamma}_i) = \frac{\partial l(\boldsymbol{\Gamma}, \boldsymbol{\eta})}{\partial \boldsymbol{\Gamma}_i \partial \boldsymbol{\Gamma}_i'} = -\sum_{l=1}^{q} \left\{ \begin{bmatrix} 1 & \overline{\theta}_l \\ \overline{\theta}_l & \overline{\theta}_l^2 \end{bmatrix} \otimes \left[\boldsymbol{T}_i \boldsymbol{W}_{il} \boldsymbol{T}_i' \right] \right\}$$

- Erros-padrão (assintóticos): inversa da Informação de Fisher.

Expressões para o MRN (versão bayesiana)

- priori: $\mathbf{\Gamma}_i \sim N_{m_i-1}(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i)$, $\boldsymbol{V}_i = \mathbf{\Sigma}_i^{-1}$.
- Vetor escore

$$\boldsymbol{S}(\mathbf{\Gamma}_i) = \sum_{l=1}^{q} \left\{ \begin{bmatrix} 1 \\ \overline{\theta}_l \end{bmatrix} \otimes \boldsymbol{T}_i \left[\overline{\boldsymbol{r}}_{il.} - \overline{f}_{il} \boldsymbol{P}_{il.} \right] \right\} - \boldsymbol{V}_i (\mathbf{\Gamma}_i - \mu_i)$$

- Matriz Hessiana (= - Informação de Fisher)

$$\boldsymbol{H}(\mathbf{\Gamma}_i) = \frac{\partial l(\mathbf{\Gamma}, \boldsymbol{\eta})}{\partial \mathbf{\Gamma}_i \partial \mathbf{\Gamma}_i'} = -\sum_{l=1}^{q} \left\{ \begin{bmatrix} 1 & \overline{\theta}_l \\ \overline{\theta}_l & \overline{\theta}_l^2 \end{bmatrix} \otimes \left[\boldsymbol{T}_i \boldsymbol{W}_{il} \boldsymbol{T}_i' \right] \right\} - \boldsymbol{V}_i$$

- Erros-padrão (assintóticos): inversa da Informação de Fisher.

Estimação dos traços latentes

- Máxima verossimilhança perfilada: P_{ijs} avaliada nas estimativas obtidas para ζ_i.
- Verossimilhança perfilada:

$$L(\boldsymbol{\theta}, \widehat{\zeta}) = \prod_{j=1}^{n} \prod_{i=1}^{I} \prod_{s=1}^{m_i} P_{ijs}^{y_{ijs}} .$$

- Função Escore

$$S(\theta_j) = \sum_{i=1}^{I} \boldsymbol{\alpha}_i' \boldsymbol{T}_i \left[\boldsymbol{y}_{ij.} - \boldsymbol{P}_{ij.} \right] .$$

- Função hessiana

$$H(\theta_j) = -\sum_{i=1}^{I} \left\{ \boldsymbol{\alpha}_i' \boldsymbol{T}_i \boldsymbol{W}_{ij} \boldsymbol{T}_i' \boldsymbol{\alpha}_i \right\} .$$

Estimação dos traços latentes

- Algoritmo de Newton -Raphson.

  $\theta_j^{(t)} = \theta_j^{(t-1)} - H(\theta_j^{(t-1)})^{-1} S\left(\theta_j^{(t-1)}\right)$

- Erros-padrão (assintóticos): inversa da Informação de Fisher.

Estimação dos traços latentes

- Priori: $\theta_j \sim N(0,1)$. Posteriori perfilada: P_{ijs} avaliada nas estimativas obtidas para ζ_i.
- Função escore bayesiana

$$S(\theta_j)_B = \sum_{i=1}^{I} \boldsymbol{\alpha}_i' \boldsymbol{T}_i \left[\boldsymbol{y}_{ij.} - \boldsymbol{P}_{ij.} \right] - \theta_j \,.$$

- Função hessiana bayesiana

$$H(\theta_j)_B = -\sum_{i=1}^{I} \left\{ \boldsymbol{\alpha}_i' \boldsymbol{T}_i \boldsymbol{W}_{ij} \boldsymbol{T}_i' \boldsymbol{\alpha}_i \right\} - 1.$$

- Algoritmo de Newton -Raphson
  $\theta_j^{(t)} = \theta_j^{(t-1)} - H(\theta_j^{(t-1)})_B^{-1} S\left(\theta_j^{(t-1)}\right)_B$
- Erros-padrão (assintóticos): inversa da Informação de Fisher.

Estimação dos traços latentes

- Esperança a posteriori

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

- Variância a posteriori.

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

- Erros-padrão: raiz quadrada da variância a posteriori.

Expressões para o MRG

- Várias quantidades são iguais ou semelhantes àquelas obtidas para o MRN.
- Vetor escore $\boldsymbol{S}(\zeta_i) = (S(a_i), S(b_{i0}), ..., S(b_{im_i}))'$

$$S(a_i) \quad = \quad \sum_{l=1}^{q}\sum_{s=0}^{m_i}\left\{\frac{\overline{r}_{ils}}{P_{ils}}\left[\left(\overline{\theta}_j - b_{is}\right)W_{ils}^{+} - \left(\overline{\theta}_j - b_{i(s+1)}\right)W_{il(s+1)}^{+}\right]\right\}$$

e

$$S(b_{ih}) = a_i \sum_{l=1}^{q}\left\{W_{ilh}^{+}\left[\frac{\overline{r}_{il(h-1)}}{P_{il(h-1)}} - \frac{\overline{r}_{il(h)}}{P_{il(h)}}\right]\right\}$$

Expressões para o MRG

- Informação de Fisher

$$
\boldsymbol{I}(\zeta_i) = \begin{bmatrix}
I(b_{i1}, b_{i1}) & I(b_{i1}, b_{i2}) & 0 & \ldots & \ldots & I(b_{i1}, a_i) \\
I(b_{i2}, b_{i1}) & I(b_{i2}, b_{i2}) & I(b_{i2}, b_{i3}) & 0 & \ldots & I(b_{i2}, a_i) \\
0 & I(b_{i3}, b_{i2}) & I(b_{i3}, b_{i3}) & I(b_{i3}, b_{i4}) & \ldots & I(b_{i3}, a_i) \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & I(b_{im_i}, b_{im_i}) & I(b_{im_i}, a_i) \\
I(a_i, b_{i1}) & I(a_i, b_{i2}) & I(a_i, b_{i2}) & \ldots & I(a_i, b_{im_i}) & I(a_i, a_i)
\end{bmatrix}
$$

Expressões para o MRG

$$
\begin{aligned}
I\left(a_i, a_i\right) &= D^2 \sum_{l=1}^{q} \sum_{s=0}^{m_i} \left\{ \frac{f_{il}}{P_{ils}} \left[\left(\overline{\theta}_l - b_{is}\right) W^{+}_{ils} - \left(\overline{\theta}_l - b_{i(s+1)}\right) W^{+}_{il(s+1)} \right]^2 \right\}, \\
I\left(b_{ih}, b_{ih}\right) &= D^2 a_i^2 \sum_{l=1}^{q} \left\{ f_{il} \left[\frac{1}{P_{il(h-1)}} + \frac{1}{P_{ilh}} \right] \left(W^{+}_{ilh}\right)^2 \right\}, \\
I\left(b_{ih}, b_{i(h-1)}\right) &= D^2 a_i^2 \sum_{l=1}^{q} \left\{ \frac{f_{il} W^{+}_{ilh} W^{+}_{il(h-1)}}{P_{il(h-1)}} \right\}, \\
I\left(b_{ih}, b_{i(h+1)}\right) &= D^2 a_i^2 \sum_{l=1}^{q} \left\{ \frac{f_{il} W^{+}_{ilh} W^{+}_{il(h+1)}}{P_{ilh}} \right\}
\end{aligned}
$$

Expressões para o MRG

$$I\left(a_i, b_{ih}\right) = \sum_{l=1}^{q} f_{il} W_{ilh}^{+} \left\{ \frac{1}{P_{il(h-1)}} \left[\left(\overline{\theta}_l - b_{i(h-1)}\right) W_{il(h-1)}^{+} - \left(\overline{\theta}_l - b_{ih}\right) W_{ilh}^{+} \right] \right.$$
$$\left. - \frac{1}{P_{ilh}} \left[\left(\overline{\theta}_l - b_{ih}\right) W_{ilh}^{+} - \left(\overline{\theta}_l - b_{i(h+1)}\right) W_{il(h+1)}^{+} \right] \right\}.$$

Estimação dos traços latentes

- Máxima verossimilhança perfilada: P_{ijs} avaliada nas estimativas obtidas para ζ_i.
- Função escore.

$$
\begin{aligned}
S(\theta_j) &= \sum_{i=1}^{I}\sum_{s=0}^{m_i}\left\{\frac{y_{ijs}}{P_{ijs}}\left(\frac{\partial P_{ijs}}{\partial \theta_j}\right)\right\} \\
&= \sum_{i=1}^{I}\sum_{s=0}^{m_i}\left\{\frac{a_i y_{ijs}\left(W_{ijs}^{+} - W_{ij(s+1)}^{+}\right)}{P_{ijs}}\right\}.
\end{aligned}
$$

Estimação dos traços latentes

- Informação de Fisher.

$$I(\theta_j) = a_i^2 \sum_{j=1}^{n} \sum_{s=0}^{m_i} \left\{ \frac{\left(W_{ijs}^{+} - W_{ij(s+1)}^{+} \right)^2}{P_{ijs}} \right\}.$$

- Algoritmo Escore de Fisher $\theta_j^{(t)} = \theta_j^{(t-1)} + I(\theta_j^{(t-1)})^{-1} S\left(\theta_j^{(t-1)}\right)$
- Erros-padrão (assintóticos): inversa da Informação de Fisher.

Recursos computacionais

- Multilog.
  - Modelos dicotômicos e modelo de resposta nominal.
  - MVM, MV, MAP.
  - Facilidades semelhantes às do BILOG-MG.
  - Em sua versão mais recente permite a análise de grupos múltiplos.

Recursos computacionais

- Pacote mirt (MRN e MRG): opções de parametrização para o MRN.
- $a_{is}(\theta_j - b_{is}) = as_{s-1} * (a_1 * \theta_j) + d_{s-1}$, em que $as_0 = 0$, $as_{(s-1)} = s - 1$ e $d_0 = 0$.

Recursos computacionais

- No MRN esta parametrização ajuda a identificar o ordenamento empírico das categorias inspecionando os valores de ak. Valores maiores indicam que a categoria do item está relacionada mais positivamente com a(s) característica(s) do(s) traço(s) latente(s) medidos. Por exemplo, se um item for verdadeiramente ordinal (como numa escala de Likert) e tivesse 4 categorias de resposta, esperaríamos ter $ak_0 < ak_1 < ak_2 < ak_3$, em termos de suas estimativas. Se, por outro lado, $ak_0 > ak_1$, pareceria que a segunda categoria está menos relacionada ao traço latente do que a primeira e, portanto, a segunda categoria deve ser entendida como a "pontuação mais baixa".

Estimação bayesiana plena

- Via dados aumentados: Ferreira, E. V. (2014). Modelos da Teoria de Resposta ao Item assimétricos de grupos múltiplos para respostas politômicas nominais e ordinais sob um enfoque bayesiano, Dissertação de Mestrado, Unicamp. http://repositorio.unicamp.br/bitstream/REPOSIP/306788/1/Ferreira_EduardoVargas_M.pdf
- Via verossimilhança original: Azevedo (2003)

Modelo de resposta nominal

Simular $\theta_j^{(t)} \sim g(\theta_j | \boldsymbol{\Gamma}^{(t-1)}, \boldsymbol{y}_{\ldots})$(condicional completa), para $j = 1, \ldots, n$ independentemente, considerando como priori uma $N\left(\theta_j | \mu_{\theta_j}, \sigma^2_{\theta_j}\right)$ através de :

(a) Simular $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \psi_{\theta_j})$

(b) Calcular o vetor de probabilidades de aceitação $\theta_j^{(t)} = \theta_j^{(*)}$

$$\pi_j\left(\theta_j^{(t-1)}, \theta_j^{(*)}\right) = min\left\{\frac{L(\boldsymbol{\Gamma}^{(t-1)}, \theta_j^{(*)})exp\left\{-\frac{\left(\theta_j^{(*)} - \mu_{\theta_j}\right)^2}{2\sigma^2_{\theta_j}}\right\}}{L(\boldsymbol{\Gamma}^{(t-1)}, \theta_j^{(t-1)})exp\left\{-\frac{\left(\theta_j^{(t-1)} - \mu_{\theta_j}\right)^2}{2\sigma^2_{\theta_j}}\right\}}, 1\right\}$$

(c) Aceitar cada $\theta_j^{(t)} = \theta_j^{(*)}$ com probabilidade π_j, caso contrário $\theta_j^{(t)} = \theta_j^{(t-1)}$

Modelo de resposta nominal

Simular $\boldsymbol{\Gamma}_i^{(t)} \sim g(\boldsymbol{\Gamma}_i|\boldsymbol{\theta}^{(t)}, \boldsymbol{y}_{\dots})$(condicional completa), para $i = 1, \dots, I$ independentemente, considerando como priori $f(\boldsymbol{\Gamma}_i|\boldsymbol{\tau}_i) \equiv N_{2m_i-2}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ através de

(a) Simular $\boldsymbol{\Gamma}_i^{(*)} \sim N_{2m_i-2}(\boldsymbol{\Gamma}_i^{(t-1)}, \boldsymbol{\Psi}_i)$

(b) Calcular o vetor de probabilidades de aceitação $\boldsymbol{\Gamma}_i^{(t)} = \boldsymbol{\Gamma}_i^{(*)}$

$$\pi_i\left(\boldsymbol{\Gamma}_i^{(t-1)}, \boldsymbol{\Gamma}_i^{(*)}\right) = min\left\{R_{\boldsymbol{\Gamma}_i}, 1\right\}$$

com

$$R_{\boldsymbol{\Gamma}_i} = \frac{L(\boldsymbol{\Gamma}_i^{(*)}, \boldsymbol{\theta}^{(t)})exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{\Gamma}_i^{(*)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i}\right)' \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{\Gamma}_i^{(*)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i}\right)\right]\right\}}{L(\boldsymbol{\Gamma}_i^{(t-1)}, \boldsymbol{\theta}^{(t)})exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{\Gamma}_i^{(t-1)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i}\right)' \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{\Gamma}_i^{(t-1)} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}_i}\right)\right]\right\}}$$

(c) Aceitar cada $\boldsymbol{\Gamma}_i^{(t)} = \boldsymbol{\Gamma}_i^{*}$ com probabilidade π_i, caso contrário $\boldsymbol{\Gamma}_i^{(t)} = \boldsymbol{\Gamma}_i^{(t-1)}$

Modelo de resposta gradual

Simular $\theta_j^{(t)} \sim g(\theta_j|\zeta^{(t-1)}, \boldsymbol{y})$(condicional completa), para $j = 1, \ldots, n$ independentemente, considerando como priori uma $N\left(\theta_j|\mu_{\theta_j}, \sigma^2_{\theta_j}\right)$ através de :

(a) Simular $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \psi_{\theta_j})$

(b) Calcular o vetor de probabilidades de aceitação $\theta_j^{(t)} = \theta_j^{(*)}$

$$\pi_j\left(\theta_j^{(t-1)}, \theta_j^{(*)}\right) = min\left\{\frac{L(\zeta^{(t-1)}, \theta_j^{(*)})exp\left\{-\frac{\left(\theta_j^{(*)} - \mu_{\theta_j}\right)^2}{2\sigma^2_{\theta_j}}\right\}}{L(\zeta^{(t-1)}, \theta_j^{(t-1)})exp\left\{-\frac{\left(\theta_j^{(t-1)} - \mu_{\theta_j}\right)^2}{2\sigma^2_{\theta_j}}\right\}}, 1\right\}$$

(c) Aceitar cada $\theta_j^{(t)} = \theta_j^{(*)}$ com probabilidade π_j, caso contrário $\theta_j^{(t)} = \theta_j^{(t-1)}$

Simular $\zeta_i^{(t)} \sim g(\zeta_i|\boldsymbol{\theta}_i^{(t)}, \boldsymbol{y})$(condicional completa), para $i = 1, \ldots, I$ independentemente, considerando como priori $f(\boldsymbol{\zeta}_i|\boldsymbol{\tau}_i) \equiv log-normal(a_i|\mu_{a_i}, \sigma^2_{a_i}) \times N_{m_i}(\boldsymbol{b}_i|\boldsymbol{\mu}_{\boldsymbol{b}_i}, \Sigma_{\boldsymbol{b}_i})$ através de:

(a) Simular $a_i^{(*)} \sim log-normal(a_i^{(*)}|a_i^{(t-1)}, \psi_{a_i})$, e $\boldsymbol{b}_i^{(*)} \sim N_{m_i}(\boldsymbol{b}_i^{(*)}|\boldsymbol{b}_i^{(t-1)}, \boldsymbol{\Psi}_{\boldsymbol{b}_i})$

(b) Calcular o vetor de probabilidades de aceitação $\zeta_i^{(t)} = \zeta_i^{(*)}$

$$\pi_i\left(\zeta^{(t-1)}, \zeta^{(*)}\right) = min\left\{R_{\zeta_i}, 1\right\}$$

com

$$R_{\zeta_i} = \frac{L(\zeta_i^{(*)}, \boldsymbol{\theta}^{(t)}) exp\left\{-\frac{\left(lna_i^{(*)} - \mu_{a_i}\right)^2}{2\sigma^2_{a_i}}\right\} exp\left\{-\frac{\left(lna_i^{(*)} - a_i^{(t-1)}\right)^2}{2\psi_{a_i}}\right\}\left[a_i^{(*)}\right]^2}{L(\zeta_i^{(t-1)}, \boldsymbol{\theta}^{(t)}) exp\left\{-\frac{\left(lna_i^{(t-1)} - \mu_{a_i}\right)^2}{2\sigma^2_{a_i}}\right\} exp\left\{-\frac{\left(lna_i^{(t-1)} - a_i^{(*)}\right)^2}{2\psi_{a_i}}\right\}\left[a_i^{(t-1)}\right]^2}$$

$$\times \frac{exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{b}_i^{(*)} - \boldsymbol{\mu}_{\boldsymbol{b}_i}\right)' \boldsymbol{\Sigma}_{\boldsymbol{b}_i}^{-1}\left(\boldsymbol{b}_i^{(*)} - \boldsymbol{\mu}_{\boldsymbol{b}_i}\right)\right]\right\}}{exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{b}_i^{(t-1)} - \boldsymbol{\mu}_{\boldsymbol{b}_i}\right)' \boldsymbol{\Sigma}_{\boldsymbol{b}_i}^{-1}\left(\boldsymbol{b}_i^{(t-1)} - \boldsymbol{\mu}_{\boldsymbol{b}_i}\right)\right]\right\}}$$

(c) Aceitar cada $\zeta_i^{(t)} = \zeta_i^{*}$ com probabiliade π_i, caso contrário $\zeta_i^{(t)} = \zeta_i^{(t-1)}$

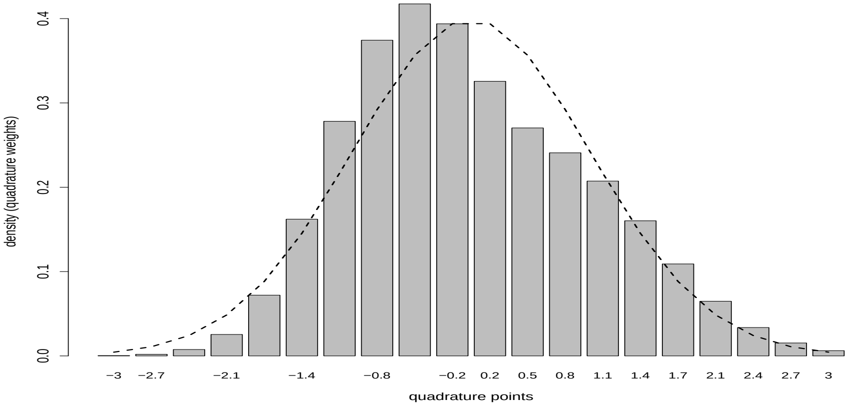
Referências bibliográficas

- Azevedo, C.L.N. (2003). *Métodos de Estimação na Teoria da Resposta ao Item*. Dissertação de Mestrado. IME-USP.
- Baker, F.B. & Kim, Seock-Ho (2004). *Item Response Theory: Parameter Estimation Techniques*, 2nd edition. New York: Marcel Dekker.

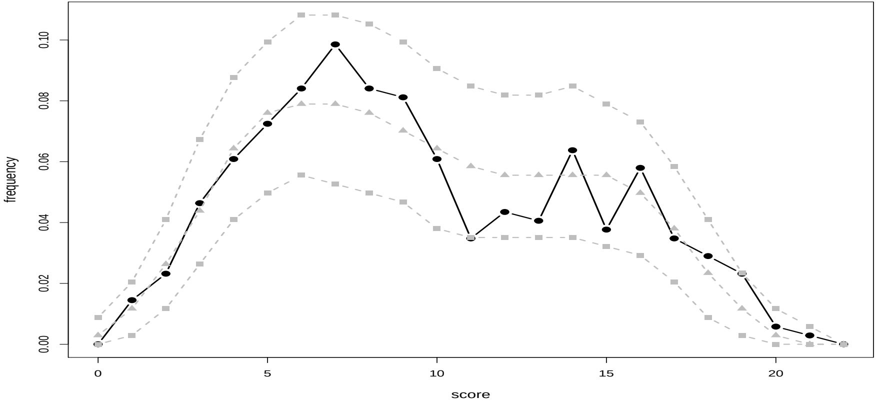
Voltando ao Exemplo 1

- Voltando ao Exemplo 1: itens de múltipla escolha com cinco alternativas.
- Ajustou-se o modelo de resposta nominal, via MIRT, utilizando o método marginal-perfilado (frequentista para os parâmetros dos itens e via EAP para os traços latentes).

Distribuição dos traços latentes



Escores observados e preditos



Verificação da qualidade de ajuste do modelo

- Curvas características do itens:

  http://www.ime.unicamp.br/~cnaber/CCIMRN.pdf

- Curvas características dos itens com bandas de confiança:

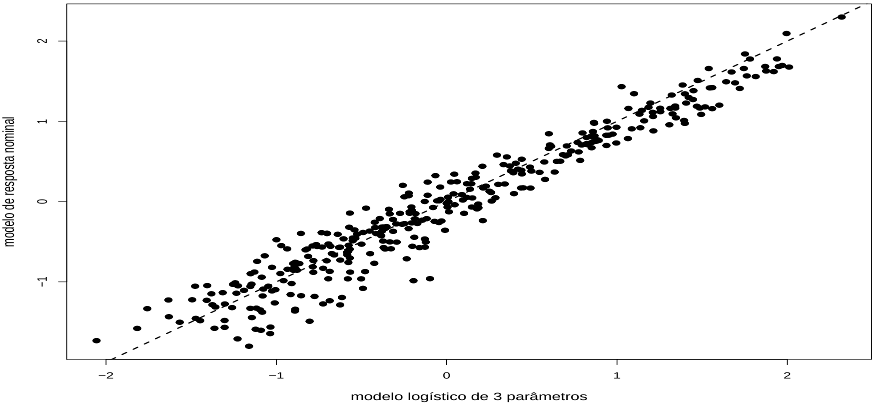
  http://www.ime.unicamp.br/~cnaber/CCICBMRN.pdf

- proporções observadas de escolhas por categoria:http://www.ime.unicamp.br/~cnaber/proportionMRN.pdf

Verificação da qualidade de ajuste do modelo

- Proporções observadas e preditas de escolhas por categorias: `http://www.ime.unicamp.br/~cnaber/proportionCIMRN.pdf`
- curvas características dos itens com probabilidades esperadas:`http://www.ime.unicamp.br/~cnaber/CCIADMRN.pdf`
- curvas características dos itens com probabilidades esperadas com intervalos de confiança:`http://www.ime.unicamp.br/~cnaber/CCIADCIMRN.pdf`

Comparação entre as estimativas dos traços latentes



Comparação entre as estimativas dos traços latentes