Métodos de Estimação na Teoria da Resposta ao Item

Caio Lucidius Naberezny Azevedo

cnaber@ime.usp.br

IME-USP

Aspectos Gerais

- Os modelos, em geral, apresentam problemas de identificabilidade.
- Necessidade do uso de métodos numéricos.
- Em geral os métodos bayesianos apresentam melhor desempenho e permitem uma análise mais detalhada.
- Existe um razoável número de métodos (combinações) na literatura.

Método de Máxima Verossimilhança

- Método de Máxima Veorossimilhança
  - Consiste em maximizar a verossimilhança com relação aos parâmetros (de interesse).
  - A verossimilhança, basicamente, é a função conjunta da amostra vista como função dos parâmetros.
- Modelo Logístico de 1 parâmetro

$$P(Y_i j = 1|\theta_j, b_i) = \frac{1}{1 + e^{-(\theta_j - b_i)}}$$

Contexto Geral

- Considere que *n* indivíduos respondem a um teste composto por *I* itens.
- Verossimilhança : pela independência entre as respostas dos diferentes indivíduos e pela independência condicional temos,

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

em que $Q_{ij} = 1 - P_{ij}$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$ e $\mathbf{b} = (b_1, \ldots, b_j)'$.

Quantidades necessárias

- logverossimilhança

$$l(\boldsymbol{\theta}, \mathbf{b}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} ln P_{ij} + (1 - y_{ij}) \, ln Q_{ij}\} \, ,$$

Quantidades necessárias

- logverossimilhança

$$l(\boldsymbol{\theta}, \mathbf{b}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} ln P_{ij} + (1 - y_{ij}) \, ln Q_{ij}\} \, ,$$

- Suponha que temos um conjunto de n=1000 indivíduos submetidos a um teste de I = 30 itens.

Quantidades necessárias

- logverossimilhança

$$l(\boldsymbol{\theta}, \mathbf{b}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} lnP_{ij} + (1 - y_{ij})\, lnQ_{ij}\},$$

- Suponha que temos um conjunto de n=1000 indivíduos submetidos a um teste de I = 30 itens.
- Considere que os itens que compõem o instrumento de medida são oriundos de uma banco de itens e, portanto, seus parâmetros são conhecidos.

Quantidades necessárias

- logverossimilhança

$$l(\boldsymbol{\theta}, \mathbf{b}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} ln P_{ij} + (1 - y_{ij})\, ln Q_{ij}\},$$

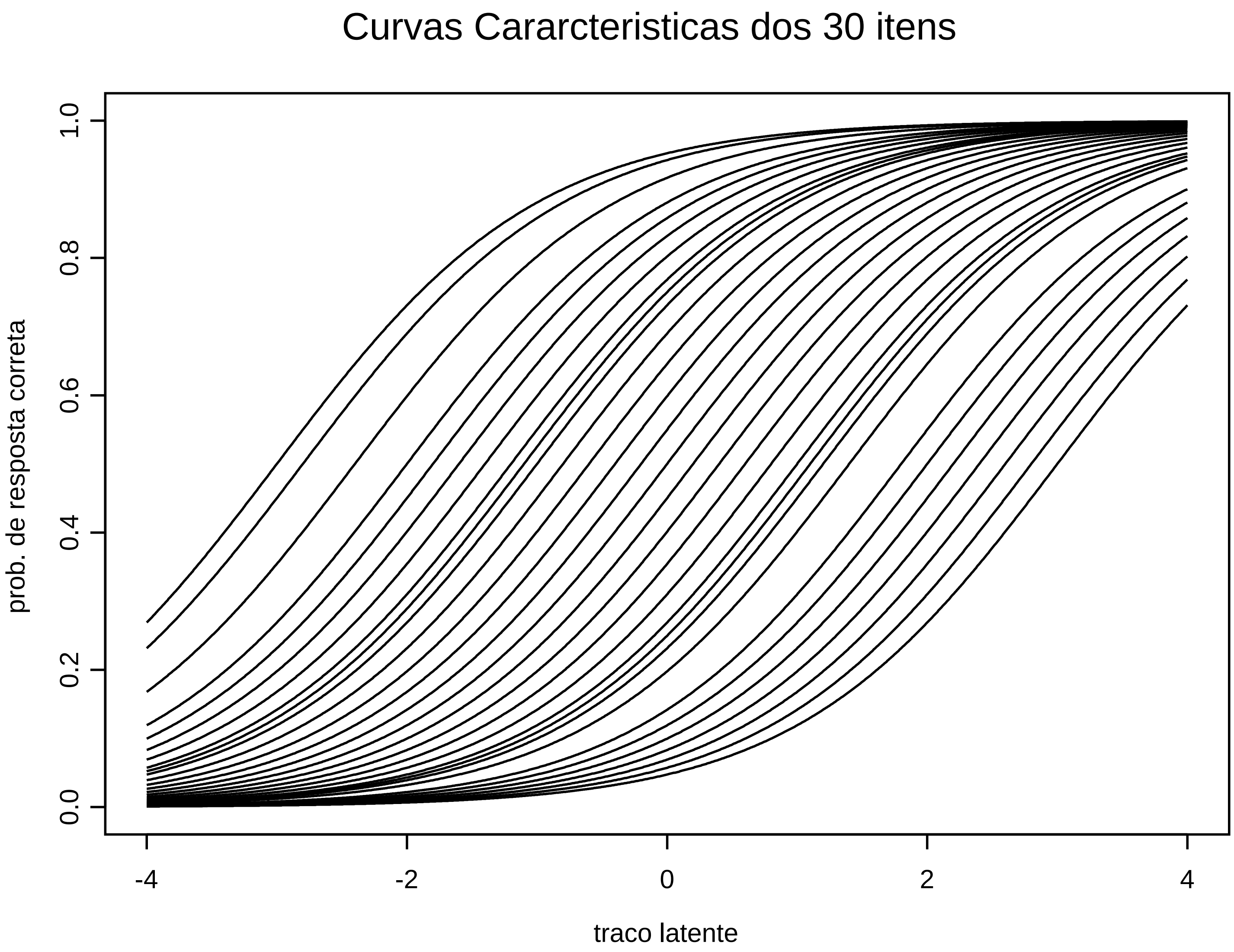
- Suponha que temos um conjunto de n=1000 indivíduos submetidos a um teste de I = 30 itens.
- Considere que os itens que compõem o instrumento de medida são oriundos de uma banco de itens e, portanto, seus parâmetros são conhecidos.
- A verossimilhança passa a ser

$$l(\boldsymbol{\theta}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} ln P_{ij} + (1 - y_{ij})\, ln Q_{ij}\}.$$

Descrição dos itens do teste 1

Item	Dificuldade (b)	Item	Dificuldade (b)	Item	Dificuldade (b)
1	-3,0	1	-0,6	1	1,4
2	-2,8	2	-0,4	2	1,8
3	-2,4	3	-0,2	3	2,0
4	-2,0	4	0,0	4	2,2
5	-1,8	5	0,2	5	2,4
6	-1,6	6	0,4	6	2,6
7	-1,4	7	0,6	7	2,8
8	-1,2	8	0,8	8	3,0
9	-1,0	9	1,0	9	-1,1
10	-0,8	10	1,2	10	1,1

Representação Gráfica dos Itens - teste 1



Desenvolvimento das equações de estimação

- Objetivo

$$S(\theta_j) = \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} lnP_{ij} + (1 - y_{ij})\, lnQ_{ij}\}$$

Desenvolvimento das equações de estimação

- Objetivo

$$S(\theta_j) = \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \sum_{j=1}^{n} \sum_{i=1}^{I} \{ y_{ij} lnP_{ij} + (1 - y_{ij})\, lnQ_{ij} \}$$

- Desenvolvimento

$$\begin{aligned}
\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_j} &= \sum_{i=1}^{I} \left\{ y_{ij} \frac{\partial \ln P_{ij}}{\partial \theta_j} + (1 - y_{ij}) \frac{\partial \ln Q_{ij}}{\partial \theta_j} \right\} \\
&= \sum_{i=1}^{I} \left\{ y_{ij} \frac{1}{P_{ij}} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right) - (1 - y_{ij}) \frac{1}{Q_{ij}} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right) \right\} \\
&= \sum_{i=1}^{I} \left\{ y_{ij} \frac{1}{P_{ij}} - (1 - y_{ij}) \frac{1}{Q_{ij}} \right\} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right) \\
&= \sum_{i=1}^{I} \left\{ \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right\} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right)
\end{aligned}$$

Desenvolvimento das equações de estimação

Mas,

$$\begin{aligned}
\frac{\partial P_{ij}}{\partial \theta_j} &= \frac{e^{-(\theta_j - b_i)}}{[1 + e^{-(\theta_j - b_i)}]^2} = \left\{ \frac{1}{1 + e^{-(\theta_j - b_i)}} \right\} \left\{ \frac{e^{-(\theta_j - b_i)}}{1 + e^{-(\theta_j - b_i)}} \right\} \\
&= P_{ij} Q_{ij},
\end{aligned}$$

dessa forma,

$$S(\theta_j) = \sum_{i=1}^{I} [y_{ij} - P_{ij}] \, .$$

- A equação acima não possui solução explícita.
- Alternativa : uso de algoritmos numéricos (Newton-Raphson/Escore de Fisher).

Elementos dos Processos Iterativos

- Para a utilização dos algoritmos anteriores é necessário o cálculo da função Hessiana / Informação de Fisher.

Elementos dos Processos Iterativos

- Para a utilização dos algoritmos anteriores é necessário o cálculo da função Hessiana / Informação de Fisher.
- Função Hessiana

$$H(\theta_j) = \frac{\partial S(\theta_j)}{\partial \theta_j} = -\sum_{i=1}^{I} \frac{\partial P_{ij}}{\partial \theta_j} = -\sum_{i=1}^{I} P_{ij} Q_{ij}.$$

Elementos dos Processos Iterativos

- Para a utilização dos algoritmos anteriores é necessário o cálculo da função Hessiana / Informação de Fisher.
- Função Hessiana

$$H(\theta_j) = \frac{\partial S(\theta_j)}{\partial \theta_j} = -\sum_{i=1}^{I} \frac{\partial P_{ij}}{\partial \theta_j} = -\sum_{i=1}^{I} P_{ij} Q_{ij}.$$

- Informação de Fisher
  Notando que a função Hessiana é não-estocástica temos,

$$I(\theta_j) = \sum_{i=1}^{I} P_{ij} Q_{ij}.$$

Métodos Iterativos

- Considere $\widehat{\theta}_j^{(t)}$ uma estimativa de θ_j na *t*-ésima iteração.

  Newton-Raphson

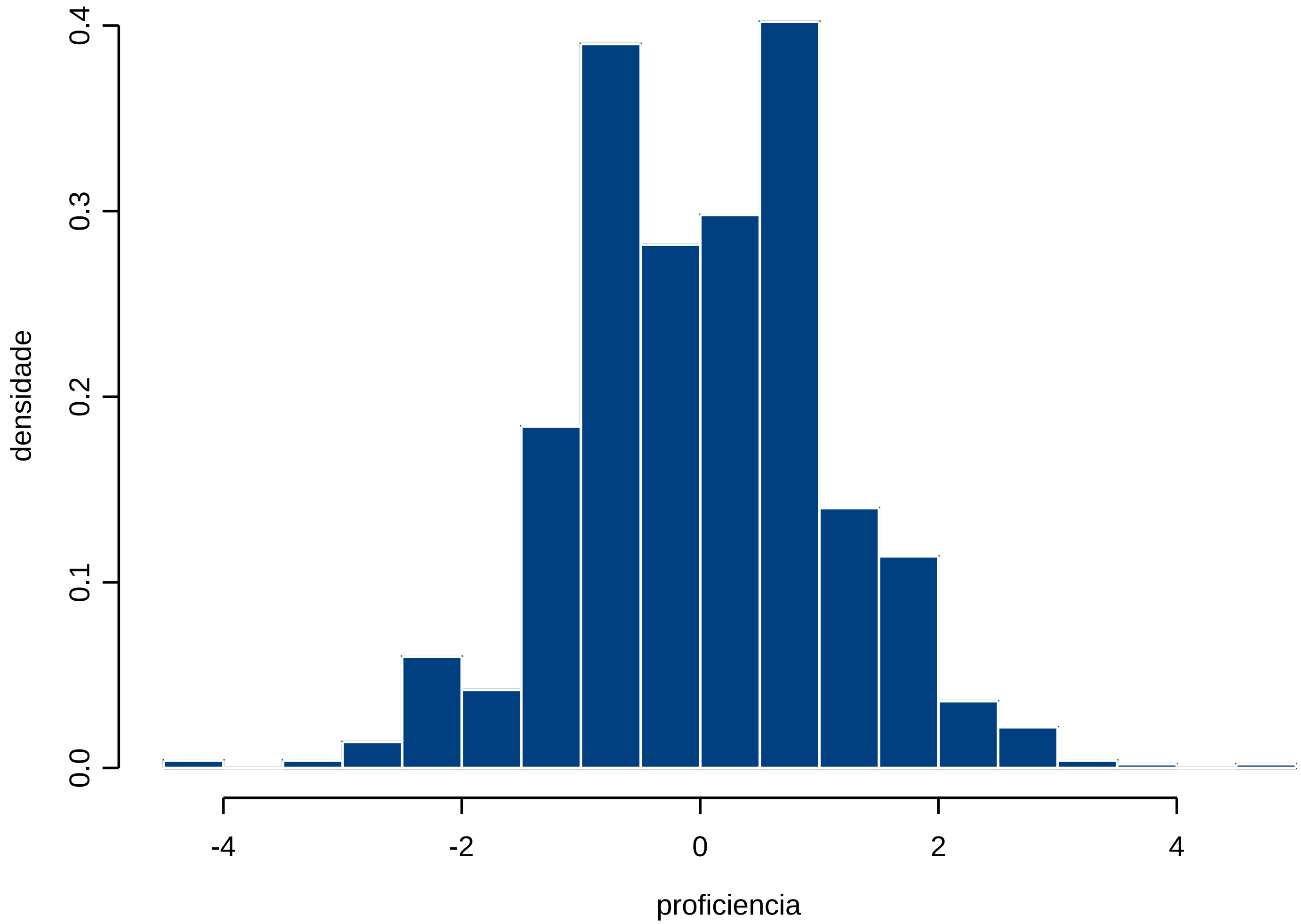
$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} - \left[H\left(\widehat{\theta}_j^{(t)}\right)\right]^{-1} S\left(\widehat{\theta}_j^{(t)}\right).$$

  Escore de Fisher

$$\widehat{\theta}_j^{(t+1)} = \widehat{\theta}_j^{(t)} + \left[I\left(\widehat{\theta}_j^{(t)}\right)\right]^{-1} S\left(\widehat{\theta}_j^{(t)}\right).$$

- Estimativas iniciais : escores padronizados.

Resultados da estimação dos traços latentes



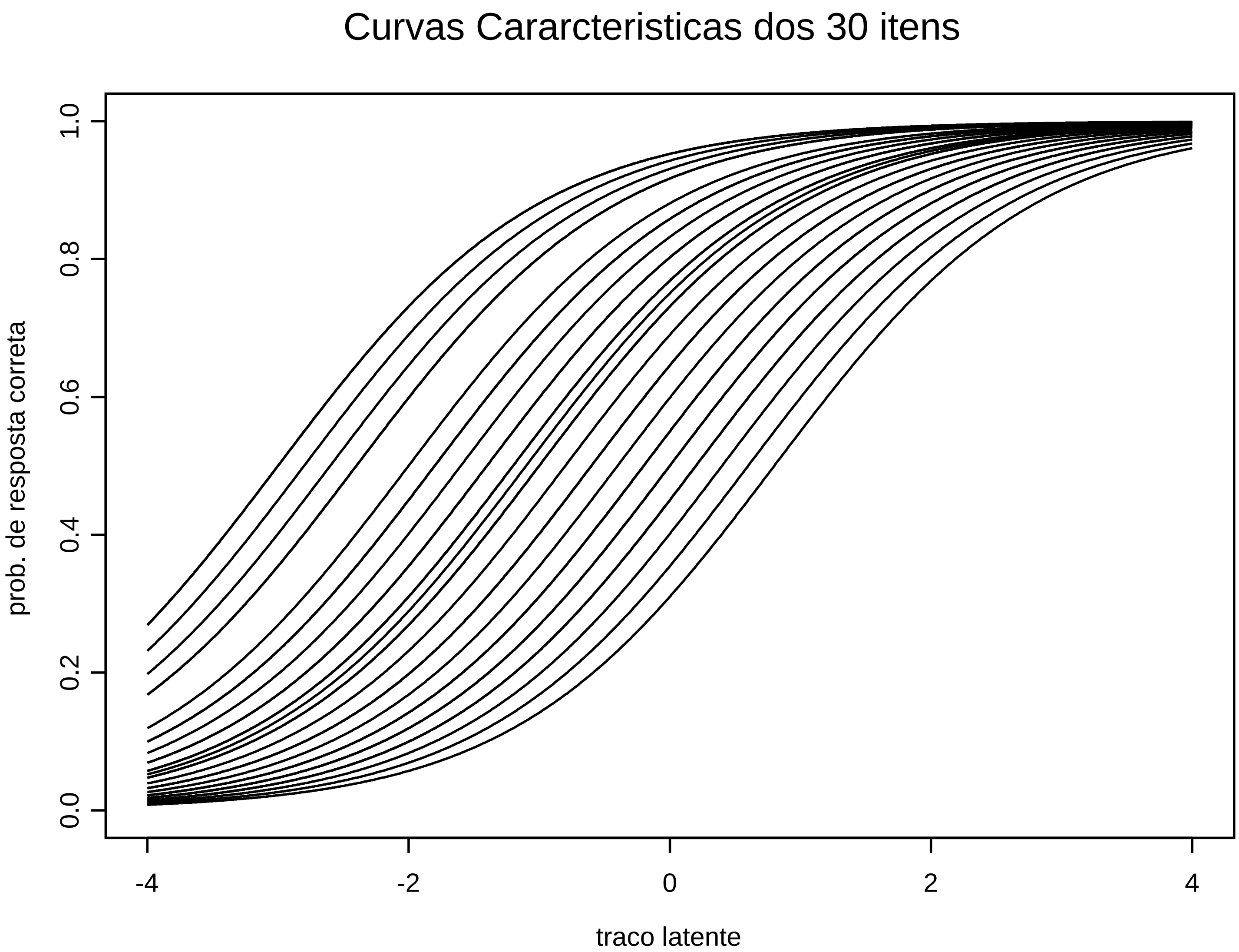
Invariância na estimação das habilidades

- As estimativas obtidas através da TRI possuem a propriedade da invariância, ou seja, uma vez obtidos os seus valores, estes são únicos.
- Considere a estimação das habilidades dos mesmos indivíduos obtidas a través de dois outros conjuntos de itens (provas).
- O primeiro conjunto é constituído de itens, em geral, mais fáceis que o primeiro conjunto. Já o segundo, é constituído de itens mais difíceis.

Descrição dos itens do teste 2

Item	Dificuldade (b)	Item	Dificuldade (b)	Item	Dificuldade (b)
1	-3,0	1	-0,6	1	-2,4
2	-2,8	2	-0,4	2	-1,8
3	-2,4	3	-0,2	3	-1,4
4	-2,0	4	0,0	4	-1,0
5	-1,8	5	0,2	5	-0,6
6	-1,6	6	0,4	6	-0,2
7	-1,4	7	0,6	7	0,2
8	-1,2	8	0,8	8	-2,6
9	-1,0	9	-1,1	9	-1,2
10	-0,8	10	-3,0	10	0,0

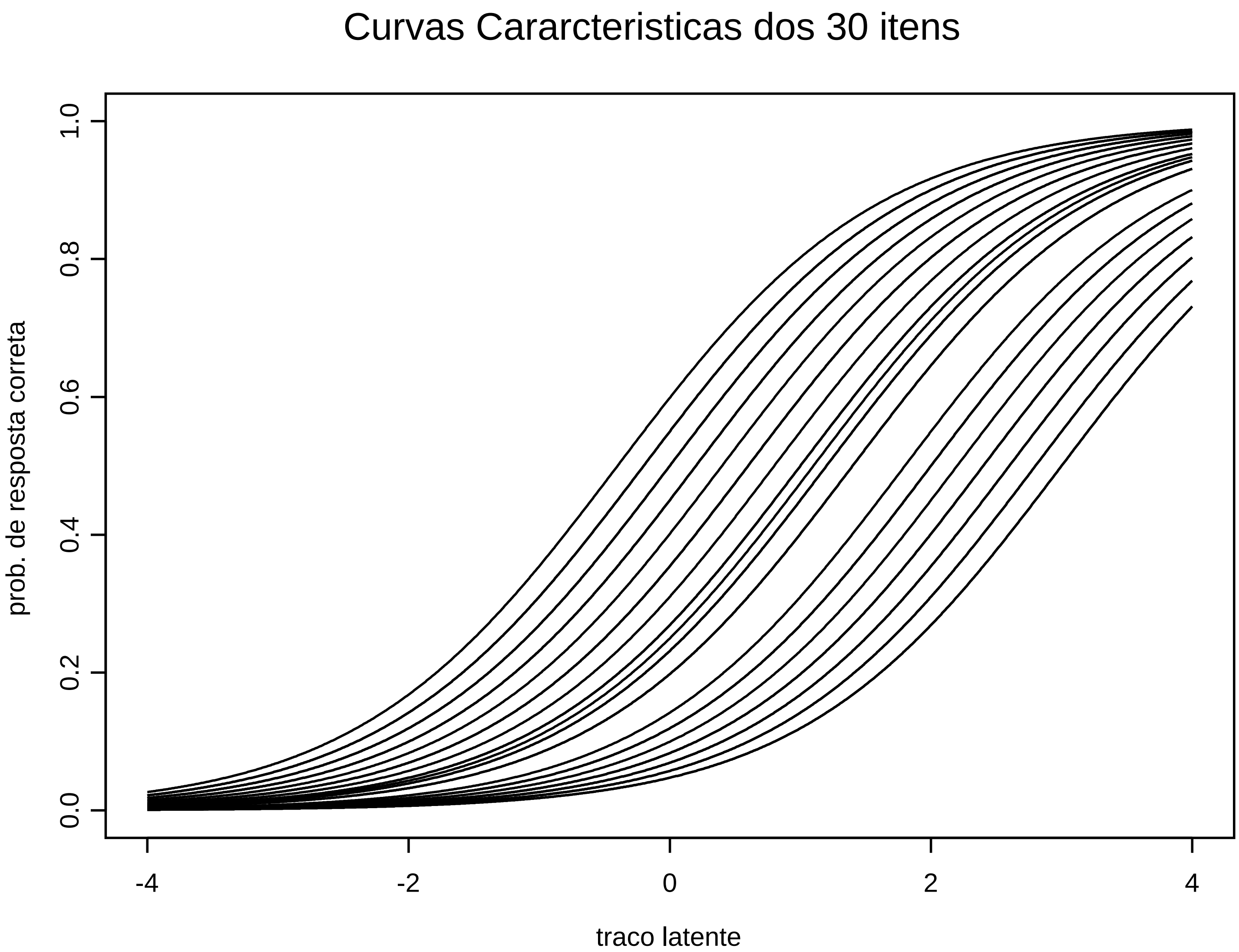
Representação Gráfica dos Itens - teste 2



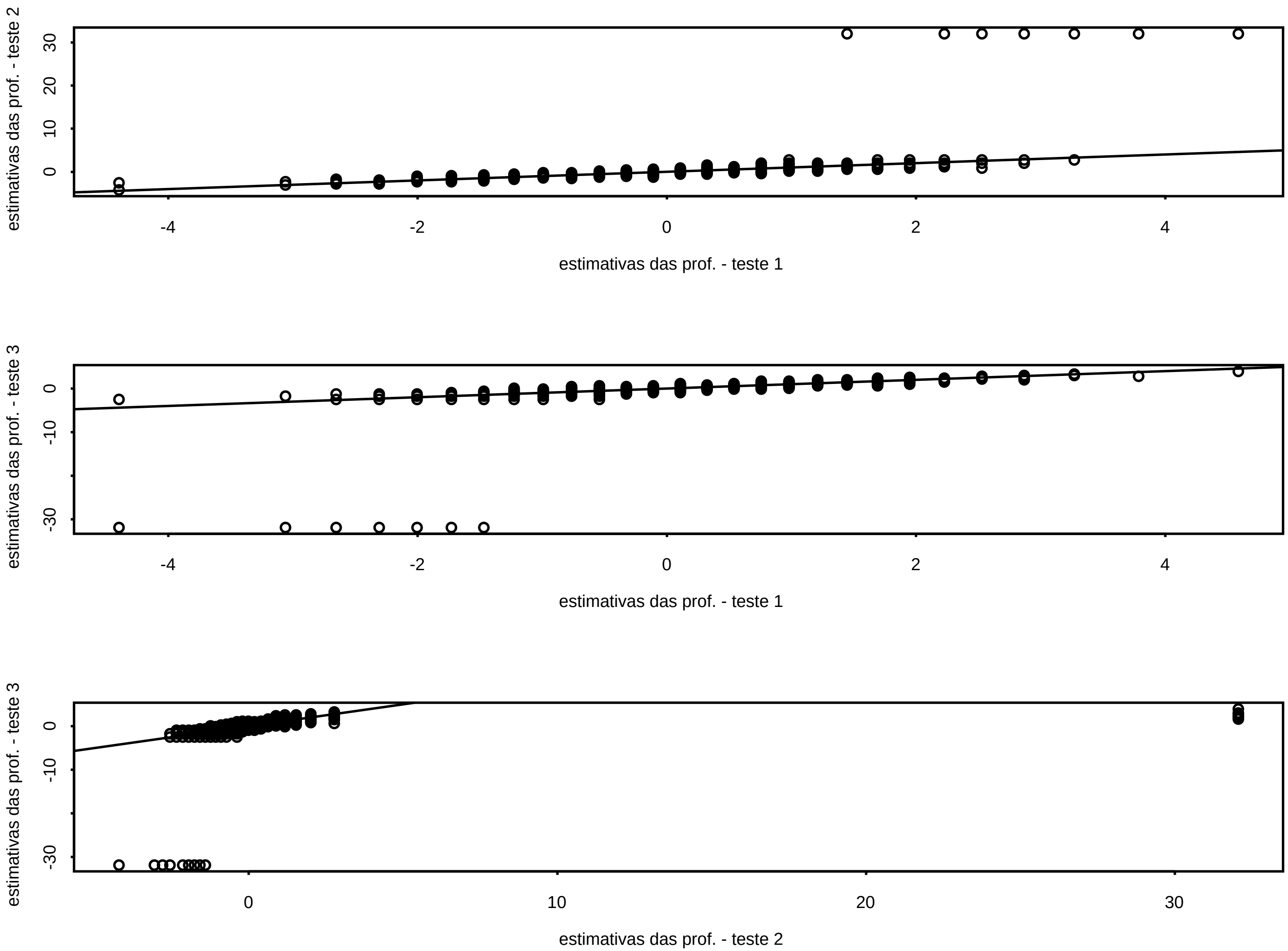
Descrição dos itens do teste 3

Item	Dificuldade (b)	Item	Dificuldade (b)	Item	Dificuldade (b)
1	1,0	1	1,1	1	0,2
2	1,2	2	1,0	2	0,4
3	1,4	3	1,4	3	0,6
4	1,8	4	2,0	4	0,8
5	2,0	5	2,4	5	0,2
6	2,2	6	2,8	6	0,6
7	2,4	7	3,0	7	0,0
8	2,6	8	2,6	8	-0,2
9	2,8	9	1,2	9	-0,2
10	3,0	10	0,0	10	-0,4

Representação Gráfica dos Itens - teste 3



Resultados da estimação dos traços latentes



Habilidades conhecidas - Parâmetros dos itens desconhecidos

- Logverossimilhança

$$l(\mathbf{b}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{y_{ij} lnP_{ij} + (1 - y_{ij})\, lnQ_{ij}\}\,.$$

Habilidades conhecidas - Parâmetros dos itens desconhecidos

- Logverossimilhança

$$l(\mathbf{b}) = \sum_{j=1}^{n} \sum_{i=1}^{I} \{ y_{ij} lnP_{ij} + (1 - y_{ij})\, lnQ_{ij} \} .$$

- Equações de estimação

$$\begin{aligned}
\frac{\partial l(b)}{\partial b_i} &= \sum_{j=1}^{n} \left\{ y_{ij} \frac{\partial\,(lnP_{ij})}{\partial b_i} + (1 - y_{ij})\, \frac{\partial\,(lnQ_{ij})}{\partial b_i} \right\} \\
&= \sum_{j=1}^{n} \left\{ y_{ij} \frac{1}{P_{ij}} \left(\frac{\partial P_{ij}}{\partial b_i} \right) - (1 - y_{ij})\, \frac{1}{Q_{ij}} \left(\frac{\partial P_{ij}}{\partial b_i} \right) \right\} \\
&= \sum_{j=1}^{n} \left\{ y_{ij} \frac{1}{P_{ij}} - (1 - y_{ij}) \frac{1}{Q_{ij}} \right\} \left(\frac{\partial P_{ij}}{\partial b_i} \right) \\
&= \sum_{j=1}^{n} \left\{ \frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right\} \left(\frac{\partial P_{ij}}{\partial b_i} \right) .
\end{aligned}$$

Desenvolvimento das equações de estimação

Mas,

$$\frac{\partial P_{ij}}{\partial b_i} = \frac{-e^{-(\theta_j - b_i)}}{[1+e^{-(\theta_j - b_i)}]^2} = \left\{ \frac{-1}{1+e^{-(\theta_j - b_i)}} \right\} \left\{ \frac{e^{-(\theta_j - b_i)}}{1+e^{-(\theta_j - b_i)}} \right\}$$
$$= -P_{ij} Q_{ij}$$

dessa forma,

$$S(b_i) = -\sum_{j=1}^{n} [y_{ij} - P_{ij}] .$$

- A equação acima não possui solução explícita.
- Alternativa : Uso de algoritmos numéricos (Newton-Raphson/Escore de Fisher).

Elementos dos Processos Iterativos

- Para a utilização dos algoritmos anteriores é necessário o cálculo da função Hessiana / Informação de Fisher.

Elementos dos Processos Iterativos

- Para a utilização dos algoritmos anteriores é necessário o cálculo da função Hessiana / Informação de Fisher.
- Função Hessiana

$$H(b_i) = \frac{\partial S(b_i)}{\partial b_i} = -\sum_{j=1}^{n} \frac{\partial P_{ij}}{\partial b_i} = -\sum_{i=j}^{n} P_{ij} Q_{ij}.$$

Elementos dos Processos Iterativos

- Para a utilização dos algoritmos anteriores é necessário o cálculo da função Hessiana / Informação de Fisher.
- Função Hessiana

$$H(b_i) = \frac{\partial S(b_i)}{\partial b_i} = -\sum_{j=1}^{n} \frac{\partial P_{ij}}{\partial b_i} = -\sum_{i=j}^{n} P_{ij} Q_{ij}.$$

- Informação de Fisher
  Notando que a função Hessiana é não-estocástica temos,

$$I(b_i) = \sum_{j=1}^{n} P_{ij} Q_{ij}.$$

Métodos Iterativos

- Considere $\widehat{b}_i^{(t)}$ uma estimativa de θ_j na *t*-ésima iteração.

  Newton-Raphson

$$\widehat{b}_i^{(t+1)} = \widehat{b}_i^{(t)} - \left[H\left(\widehat{b}_i^{(t)}\right)\right]^{-1} S\left(\widehat{b}_i^{(t)}\right).$$

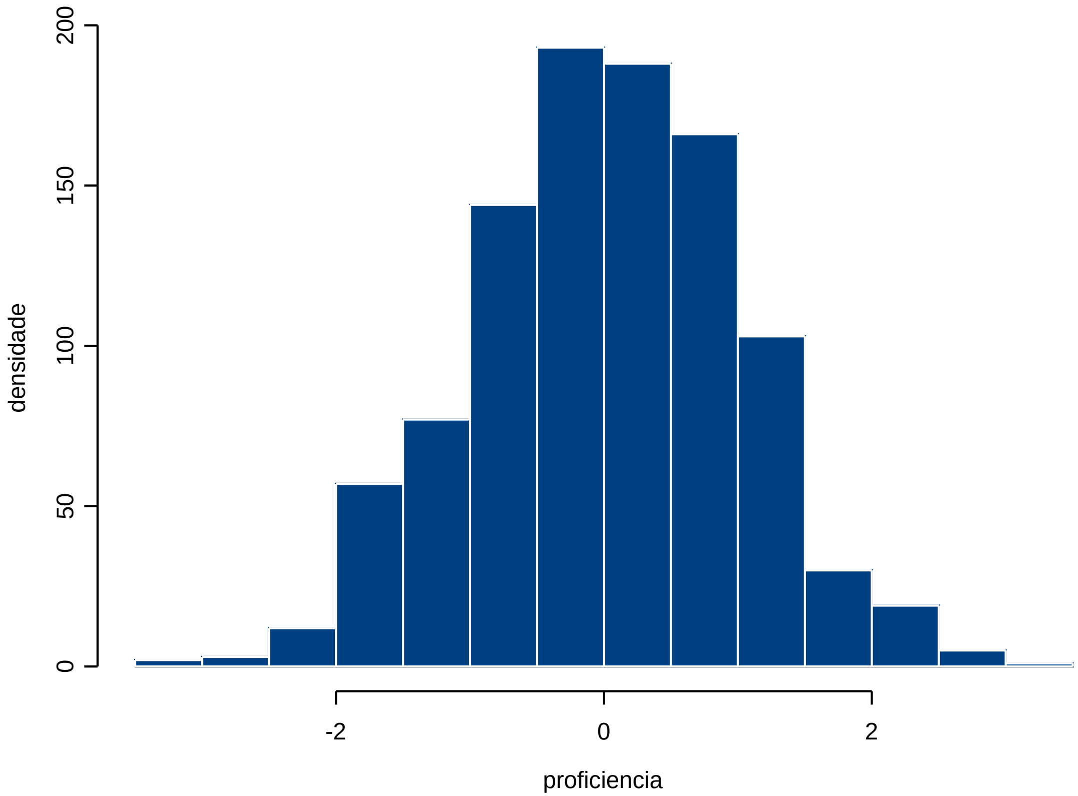
  Escore de Fisher

$$\widehat{b}_i^{(t+1)} = \widehat{b}_i^{(t)} + \left[I\left(\widehat{b}_i^{(t)}\right)\right]^{-1} S\left(\widehat{b}_i^{(t)}\right).$$

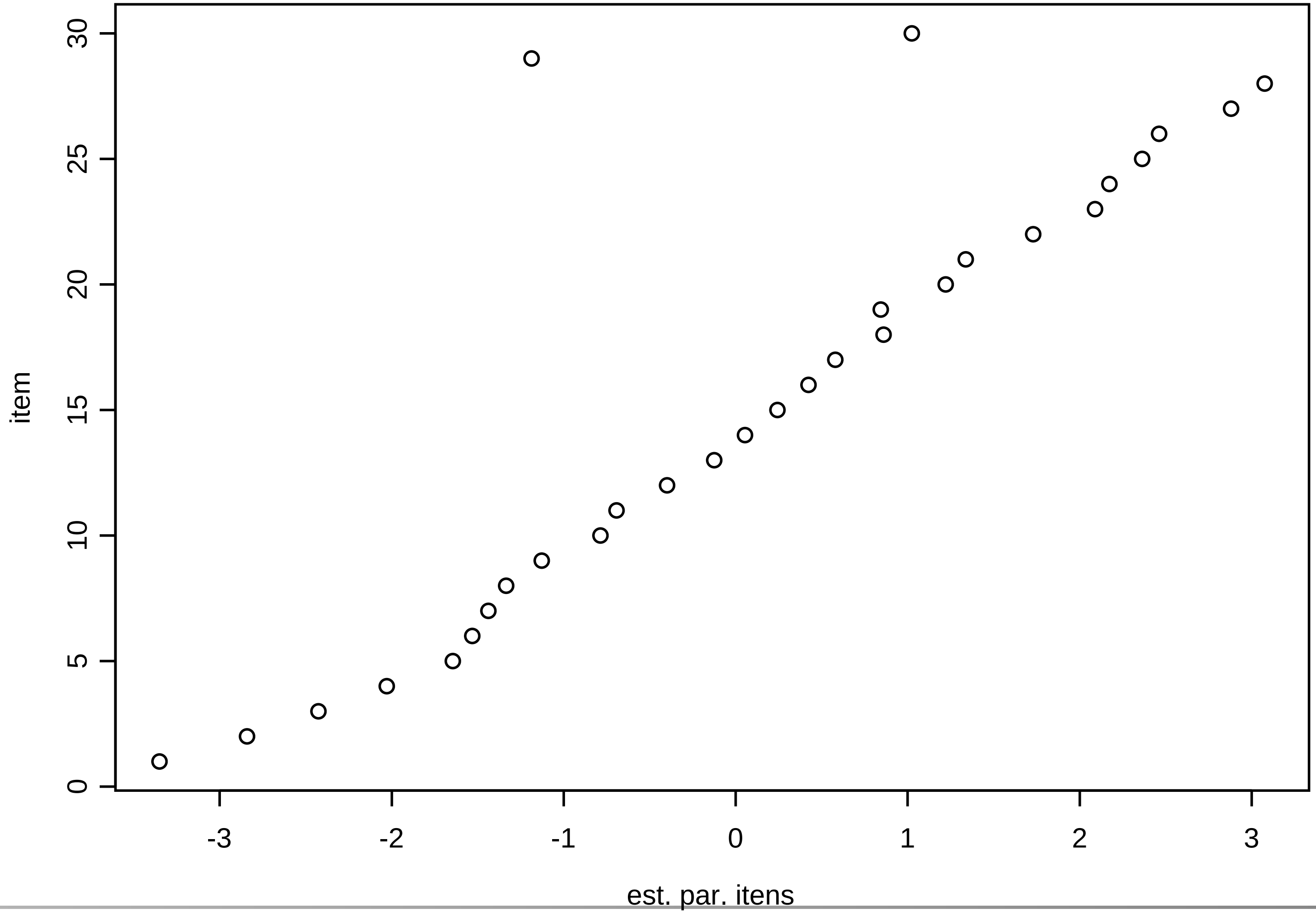
- Estimativas iniciais : proporção de erros padronizadas.

Descrição do teste 4

- Considere um grupo de n=1000 indivíduos, com as proficiências conhecidas, submetido a uma prova com I = 30 itens (com os parâmetros desconhecidos).

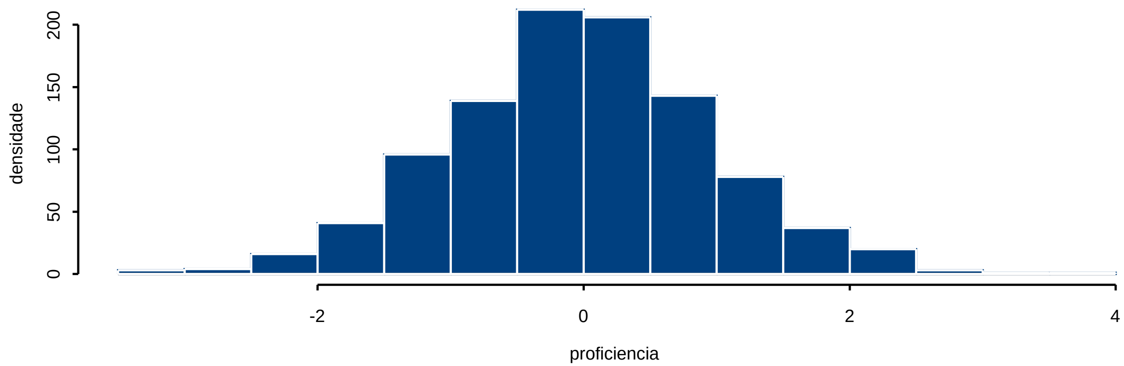


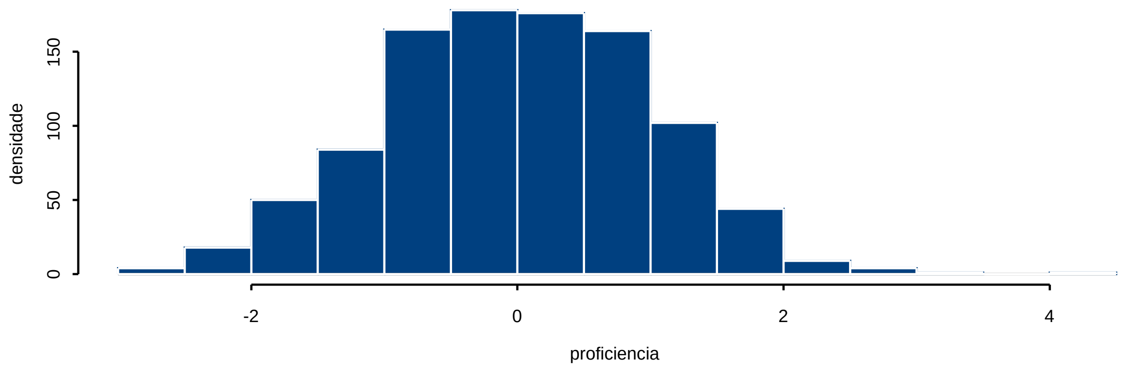
Resultados da estimação dos parâmetros dos itens - teste 4



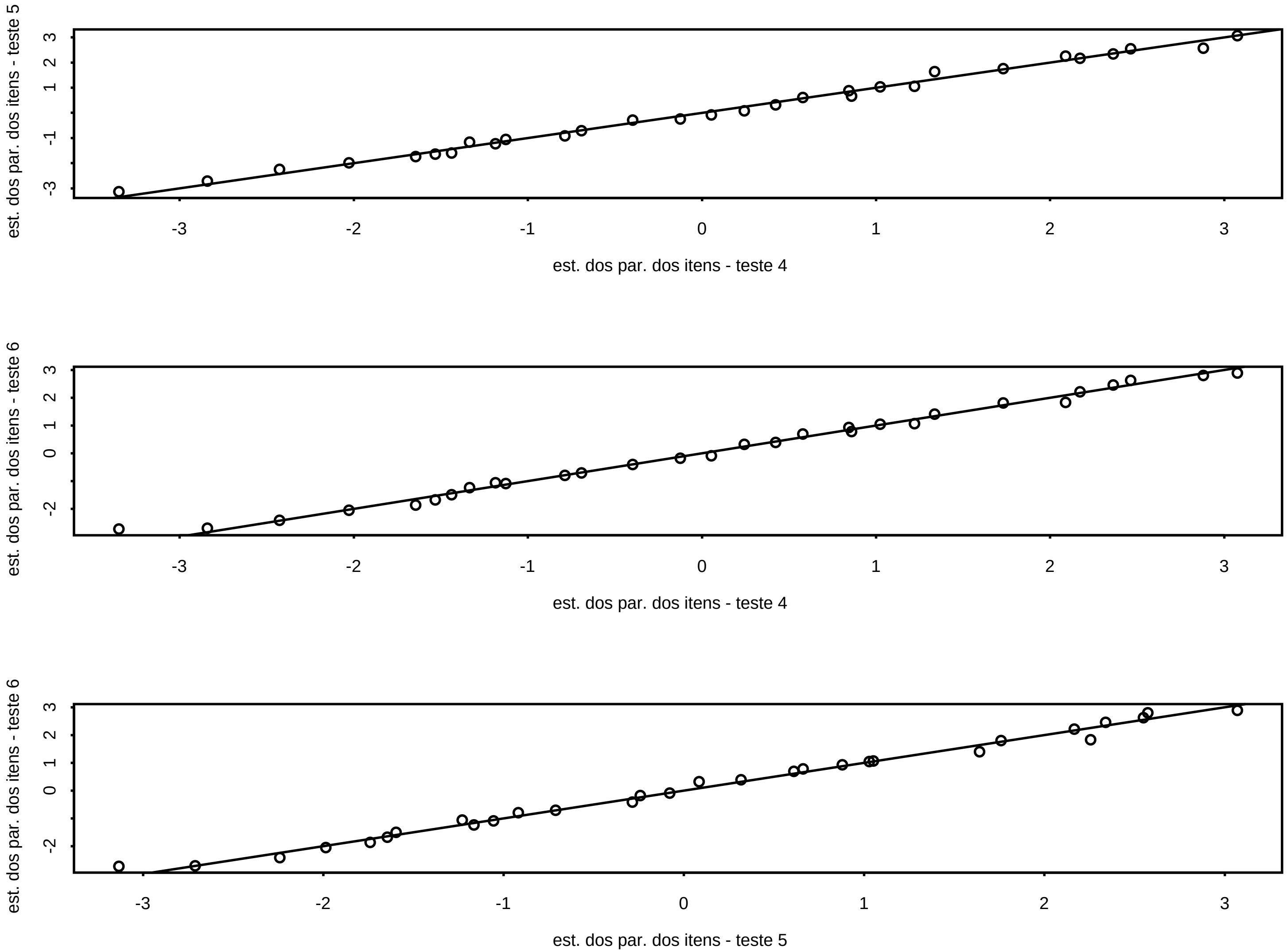
Invariância da estimação dos parâmetros dos itens

- Considere mais dois outros conjuntos de indivíduos (n=1000 cada um), com traços latentes diferentes (do conjunto anterior testes anteriores e entre si).





Resultados da estimação dos parâmetros dos itens - 3 testes



Estimação dos dois conjuntos de parâmetros

- Suponha que temos um conjunto de n=1000 indivíduos submetidos a um teste de I = 30 itens,em que todos os parâmetros são desconhecidos.
- Temos que estimar, neste caso, 1030 parâmetros.
- Note que, agora, há a necessidade de se estabelecer uma métrica em que os parâmetros irão ser estimados.
- Verossimilhança

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

Considerações do processo de estimação - Máxima Verossimilhança

- Não identificabilidade

$$P_{ij} = \frac{1}{1+e^{-D(\theta_j - b_i)}} = \frac{1}{1+e^{-D((\theta_j - d)-(b_i - d))}}$$
$$= \frac{1}{1+e^{-D\left(\theta_j^* - b_i^*\right)}}$$

- Estimação simultânea
  - Inversão de matrizes da ordem de n x I.
  - Comprometimento das propriedades assintóticas dos estimadores.
- Alternativa : Estimação por Máxima Verossimilhança Marginal.

Estimação por Máxima Verossimilhança Marginal

- 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.

Construção da Verossimilhança Marginal

- Probabilidade Marginal de Resposta

$$
\begin{aligned}
P\left(\mathbf{Y}_{.j}=\mathbf{y}_{.j}|\mathbf{b},\eta\right) \equiv P\left(\mathbf{Y}_{.j}|\mathbf{b},\eta\right) &= \int_{I\!R} P\left(\mathbf{Y}_{.j}=\mathbf{y}_{.j}|\theta,\mathbf{b}\right) g\left(\theta,\eta\right) d\theta \\
&= \int_{I\!R} P\left(\mathbf{Y}_{.j}=\mathbf{y}_{.j}|\theta,\mathbf{b}\right) g\left(\theta,\eta\right) d\theta \\
&= \int_{I\!R} P\left(\mathbf{Y}_{.j}|\theta,\mathbf{b}\right) g\left(\theta,\eta\right) d\theta\,,
\end{aligned}
$$

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

- Verossimilhança marginal

$$
\begin{aligned}
L\left(\mathbf{b},\eta\right) &= \prod_{j=1}^{n} P\left(\mathbf{Y}_{.j}|\mathbf{b},\eta\right) = \prod_{j=1}^{n} \int_{I\!R} P\left(\mathbf{Y}_{.j}|\theta,\mathbf{b}\right) g\left(\theta,\eta\right) d\theta \\
&= \prod_{j=1}^{n} \int_{I\!R} \prod_{i=1}^{I} P\left(\mathbf{Y}_{ij}|\theta,b_i\right) g\left(\theta,\eta\right) d\theta\,.
\end{aligned}
$$

Desenvolvimento das expressões

- logverossimilhança

$$l(\mathbf{b}, \eta) \quad = \quad \sum_{j=1}^{n} ln \int_{I\!R} \prod_{i=1}^{I} P\left(\mathbf{Y}_{ij} | \theta, b_i\right) g\left(\theta, \eta\right) d\theta \,.$$

Desenvolvimento das expressões

- logverossimilhança

$$l(\mathbf{b}, \eta) \quad = \quad \sum_{j=1}^{n} ln \int_{I\!R} \prod_{i=1}^{I} P(\mathbf{Y}_{ij} | \theta, b_i)\, g(\theta, \eta)\, d\theta\,.$$

- Estimadores de Máxima verossimilhança (Marginal)

$$\begin{aligned} \frac{\partial l(\mathbf{b}, \eta)}{\partial \mathbf{b}_i} &= \frac{\partial}{\partial \mathbf{b}_i} \left\{ \sum_{j=1}^{n} lnP(\mathbf{Y}_{.j} | \mathbf{b}, \eta) \right\} \\ &= \sum_{j=1}^{n} \frac{1}{P(\mathbf{Y}_{.j} | \mathbf{b}, \eta)} \frac{\partial P(\mathbf{Y}_{.j} | \mathbf{b}, \eta)}{\partial b_i}\,. \end{aligned}$$

Desenvolvimento das expressões

- Mas,

$$\begin{aligned}
\frac{\partial P\left(\mathbf{Y}_{.j}|\mathbf{b},\eta\right)}{\partial \mathbf{b}_i} &= \frac{\partial}{\partial \mathbf{b}_i}\int_{I\!R} P\left(\mathbf{Y}_{.j}|\theta,\mathbf{b}\right) g\left(\theta,\eta\right) d\theta \\
&= \int_{I\!R}\left(\frac{\partial}{\partial b_i} P\left(\mathbf{Y}_{.j}|\theta,\mathbf{b}\right)\right) g\left(\theta,\eta\right) d\theta \\
&= \int_{I\!R}\left(\frac{\partial}{\partial b_i}\prod_{h=1}^{I} P\left(\mathbf{Y}_{hj}|\theta,b_h\right)\right) g\left(\theta,\eta\right) d\theta \\
&= \int_{I\!R}\left(\prod_{h\neq i}^{I} P\left(\mathbf{Y}_{hj}|\theta,b_h\right)\right)\left(\frac{\partial}{\partial b_i} P\left(\mathbf{Y}_{ij}|\theta,b_i\right)\right) g\left(\theta,\eta\right) d\theta \\
&= \int_{I\!R}\left(\frac{\partial P\left(\mathbf{Y}_{ij}|\theta,b_i\right)/\partial b_i}{P\left(\mathbf{Y}_{ij}|\theta,b_i\right)}\right) P\left(\mathbf{Y}_{.j}|\theta,\mathbf{b}\right) g\left(\theta,\eta\right) d\theta\,.
\end{aligned}$$

Desenvolvimento das expressões

- Além disso,

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

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,

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

Note agora que

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

Desenvolvimento das expressões

- Dessa forma, temos que

$$\frac{1}{P\left(\mathbf{Y}_{ij}|\theta, b_i\right)} \frac{\partial}{\partial b_i} P\left(\mathbf{Y}_{ij}|\theta, b_i\right) = \frac{(y_{ij} - P_i)}{P_i Q_i} \left(\frac{\partial P_i}{\partial b_i}\right) ,$$

Logo,

$$\begin{aligned} \frac{\partial P\left(\mathbf{Y}_{.j}|\mathbf{b}, \eta\right)}{\partial b_i} &= \int_{I\!R} \left[\frac{(y_{ij} - P_i)}{P_i Q_i} \left(\frac{\partial P_i}{\partial b_i}\right)\right] P\left(\mathbf{Y}_{.j}|\theta, \mathbf{b}\right) g\left(\theta, \eta\right) d\theta \\ &= -\int_{I\!R} \left[(y_{ij} - P_i)\right] P\left(\mathbf{Y}_{.j}|\theta, \mathbf{b}\right) g\left(\theta, \eta\right) d\theta , \end{aligned}$$

portanto,

$$S(b_i) = \frac{1}{P\left(\mathbf{Y}_{.j}|\mathbf{b}, \eta\right)} \frac{\partial P\left(\mathbf{Y}_{.j}|\mathbf{b}, \eta\right)}{\partial b_i} = -\sum_{j=1}^{n} \int_{I\!R} \left[y_{ij} - P_i\right] g_j^*\left(\theta\right) ,$$

em que,

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

Desenvolvimento das expressões

- Forma de quadratura

$$S(b_i) = \sum_{j=1}^{n} \sum_{l=1}^{q} \left[(y_{ij} - P_{il}) \right] g_j^* \left(\overline{\theta}_l \right) \ ,$$

Desenvolvimento das expressões

- Forma de quadratura

$$S(b_i) = \sum_{j=1}^{n} \sum_{l=1}^{q} \left[(y_{ij} - P_{il}) \right] g_j^* \left(\overline{\theta}_l \right) \ ,$$

- Equação de Bock & Aitkin

$$\begin{aligned} S(b_i) \quad &= \quad \sum_{l=1}^{q} \Bigg[\Bigg(\sum_{j=1}^{n} y_{ij} g_j^* \left(\overline{\theta}_l \right) - P_{il} \sum_{j=1}^{n} g_j^* \left(\overline{\theta}_l \right) \Bigg) \Bigg] \\ &= \quad \sum_{l=1}^{q} \left[\left(\overline{r}_{il} - \overline{f}_{il} P_{il} \right) \right] \ , \end{aligned}$$

em que

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

Aplicação do Algoritmo EM

- Calcula estimativas de máxima verossimilhança na presença de dados faltantes (processo iterativo).
- Aplicação na TRI : considerar as proficiências como os dados não observados.
- Implementação do algoritmo EM
  Seja $L(\mathbf{b}|\mathbf{Y}_{.}.., \boldsymbol{\theta})$ a densidade conjunta do dados completos (verossimilhança) .
  Se $\widehat{\mathbf{b}}^{(t)}$ é uma estimativa de $\mathbf{b}$ na iteração t, então os passos EM para obtenção de $\widehat{\mathbf{b}}^{(t+1)}$ são

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

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

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 $\pi = (\pi_{\mathbf{1}}, \ldots, \pi_{\mathbf{q}})'$.
- Denote por $\mathbf{f_i}$=$(f_{i1}, \cdots, f_{iq})'$ a quantidade de indivíduos em cada nível de habilidade e $\mathbf{r_i}$ =$(r_{i1}, \cdots, r_{iq})'$ a quantidade daqueles que respondem corretamente ao item i com nível de habilidade *l*, ambos observados na amostra. Além disso $\mathbf{r} = (\mathbf{r_1}, \cdots, \mathbf{r_I})'$.
- A probabilidade conjunta que os f_{il} indivíduos tenham habilidades $\overline{\theta}_l$, $l = 1, \cdots, q$, é dada pela distribuição multinomial:

$$P\left(\mathbf{F_i} = \mathbf{f_i}|\pi\right) \equiv P(\mathbf{f_i}|\pi) = \frac{\mathbf{n_{(i)}!}}{\prod_{\mathbf{l=1}}^{\mathbf{q}} \mathbf{f_{il}!}} \prod_{\mathbf{l=1}}^{\mathbf{q}} \pi_{\mathbf{j}}^{\mathbf{f_{il}}}, \quad \mathbf{i = 1, \cdots, I},$$

- 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

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

Desenvolvimento da verossimilhança - Algoritmo EM

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

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

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

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

Desenvolvimento da verossimilhança - Algoritmo EM

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

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

em que

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

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

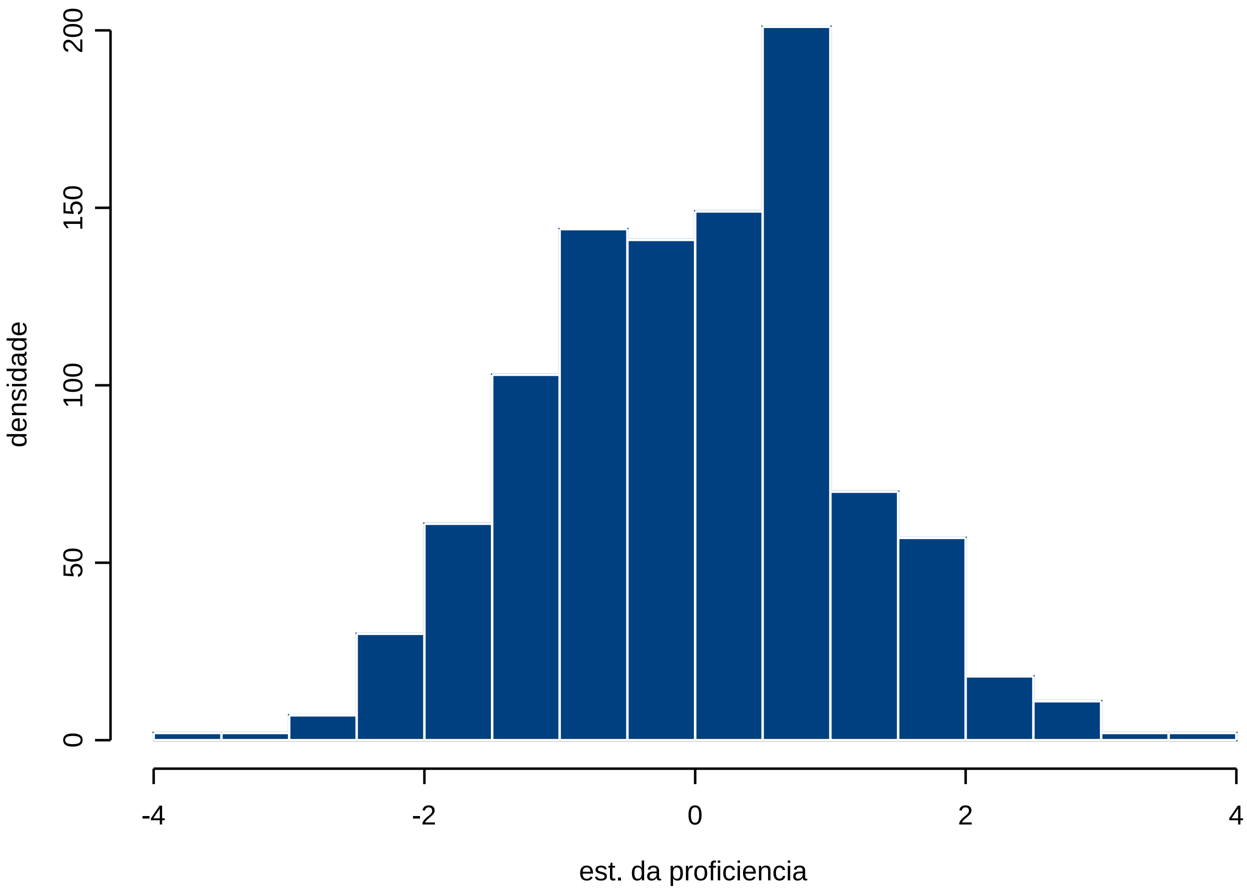
Estimação das habilidades - Máxima Verossimilhança Perfilada

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

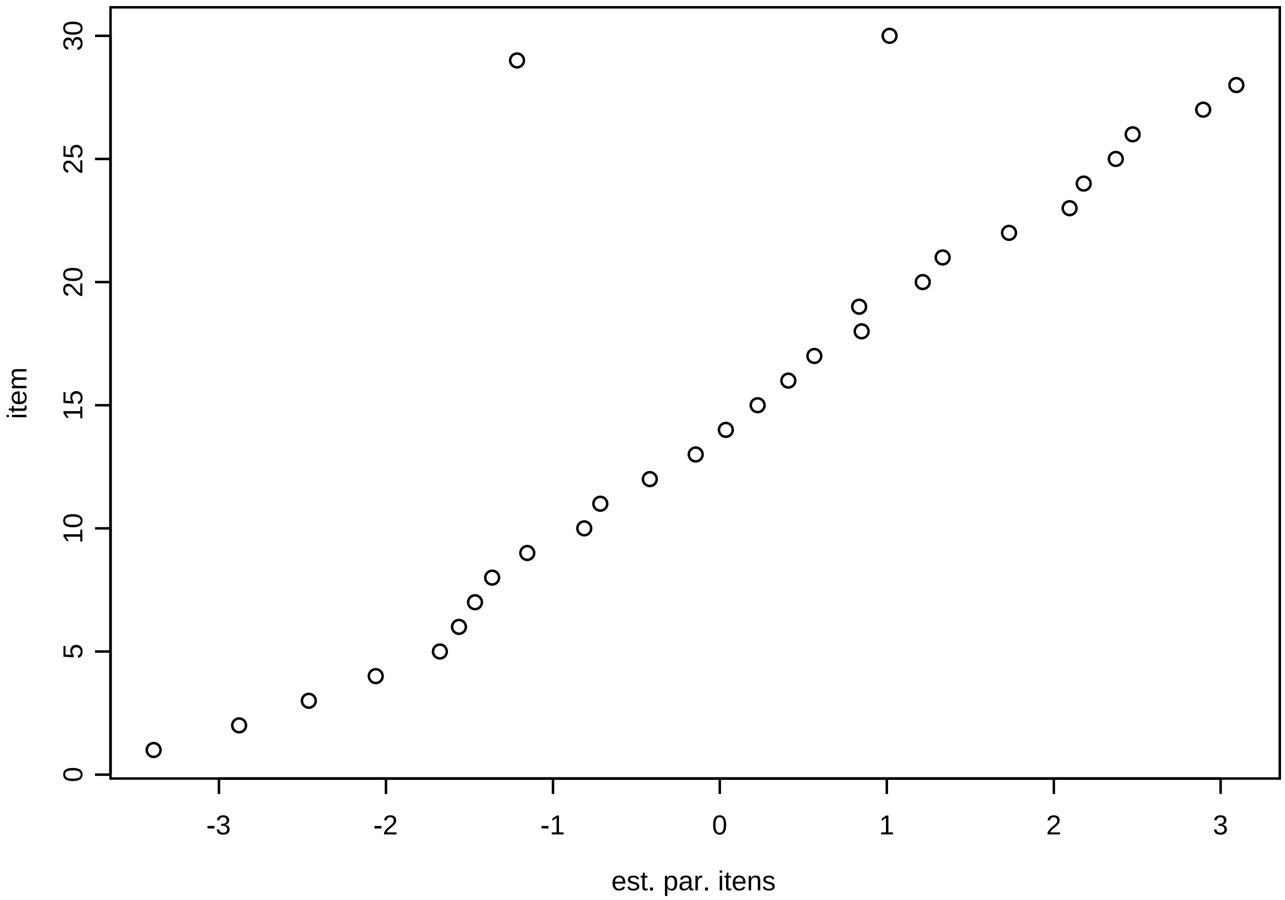
$$L(\boldsymbol{\theta}, \widehat{\mathbf{b}}) = \prod_{i=1}^{I} \prod_{j=1}^{n} \widehat{P}_{ij}^{y_{ij}} \widehat{Q}_{ij}^{1-y_{ij}} ,$$

- Assim, utiliza-se a verossimilhança acima, tal como no caso em que os parâmetros dos itens eram conhecidos.
- Considere um teste com n=1000 indivíduos e I=30 itens, com todos os parâmetros desconhecidos.

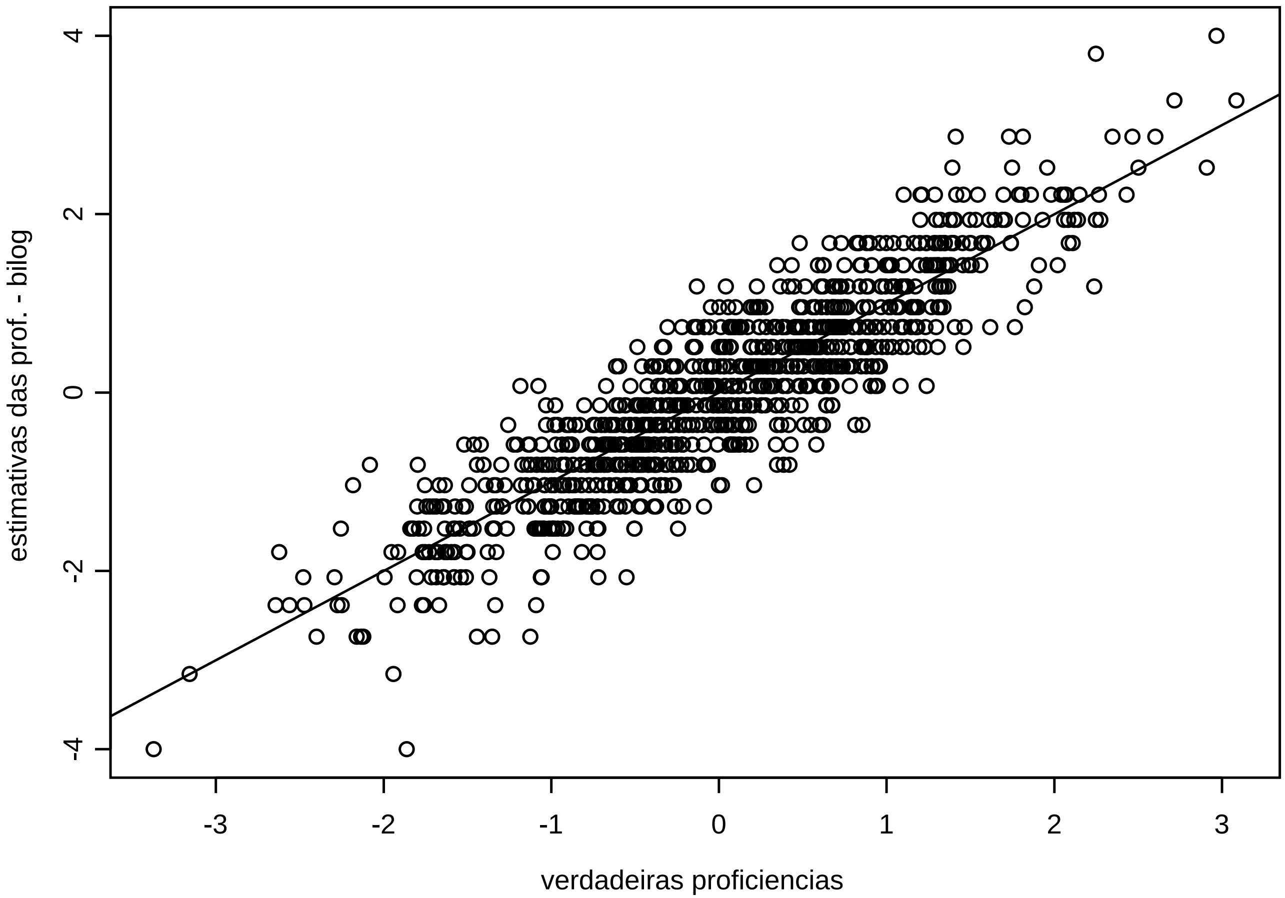
Resultado do processo de estimação



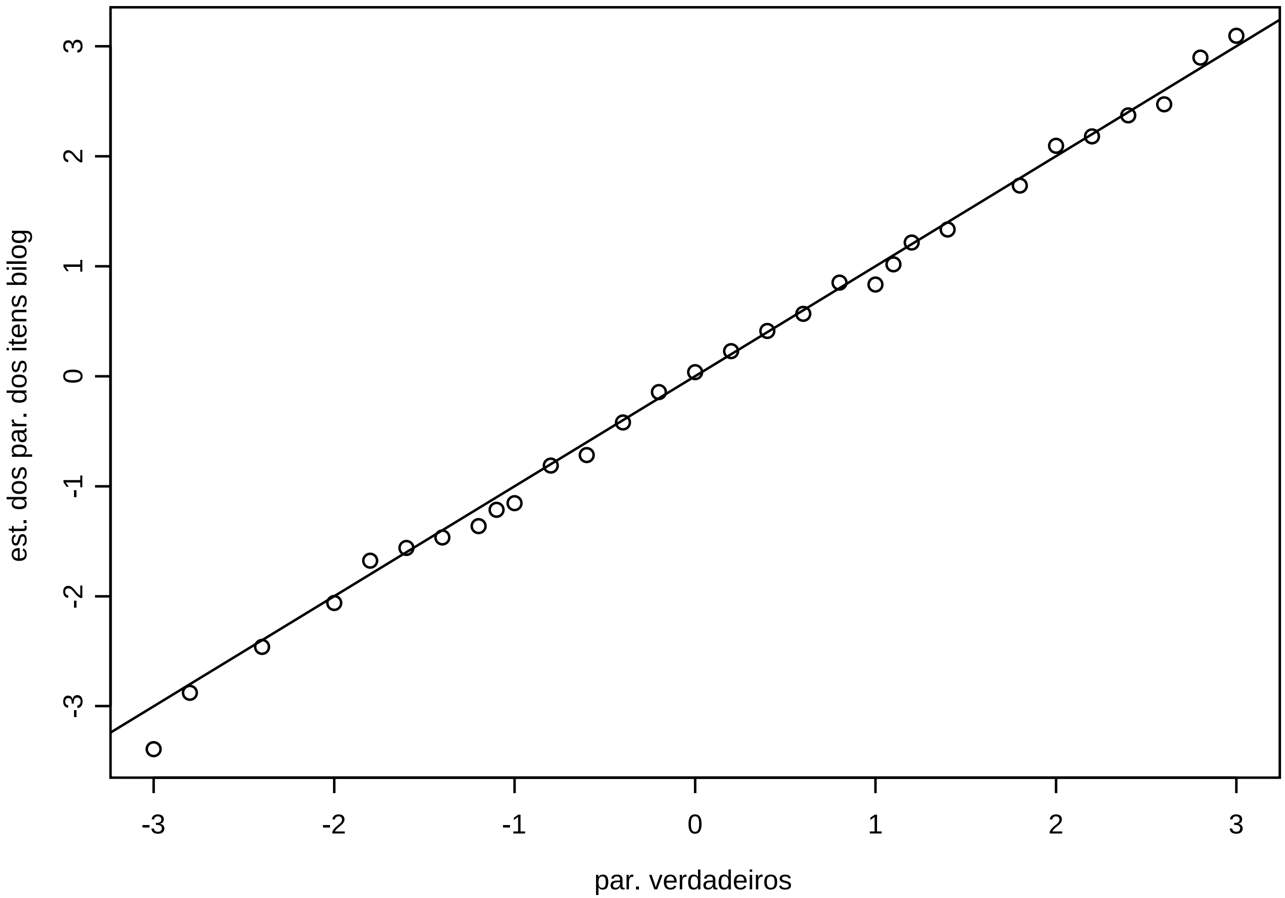
Resultado do processo de estimação



Resultado do processo de estimação - comp. com os verd. valores



Resultado do processo de estimação - comp. com os verd. valores



Comparação das verossimilhanças