Aula 3. Introduçao à TRI

Mariana Cúri - ICMC/USP

mcuri@icmc.usp.br
www.icmc.usp.br/∼mcuri

julho de 2015

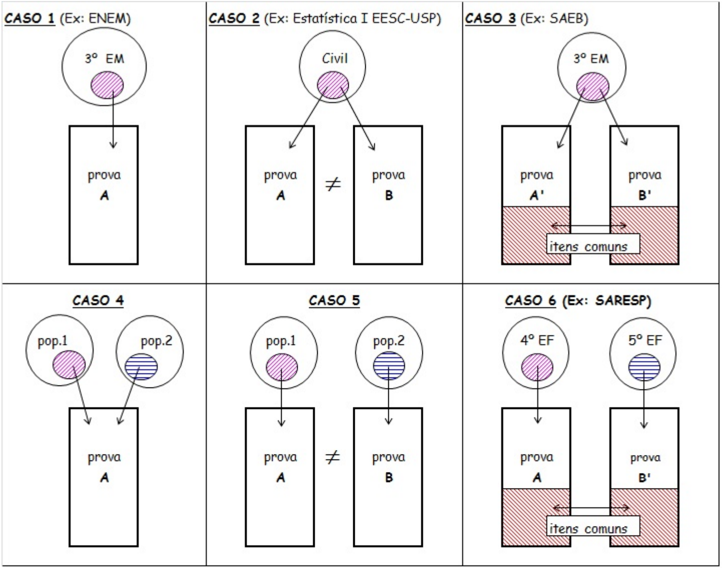
Conteúdo da Apresentação

1. Introdução: Avaliações Educacionais

- Objetivo: classificação (Vestibular), certificação (aprovação em um curso), políticas educacionais (SARESP, SAEB)
- Construto: traço(s) latente(s) não observável(eis)
    - proficiência em língua estrangeira
    - habilidade em Matemática
    - outras áreas: intensidade de depressão, nível de qualidade de vida, grau de aceitação de um novo produto no mercado, predisposiçao para desenvolver determinada doença
- Instrumento de avaliação: prova composta por itens

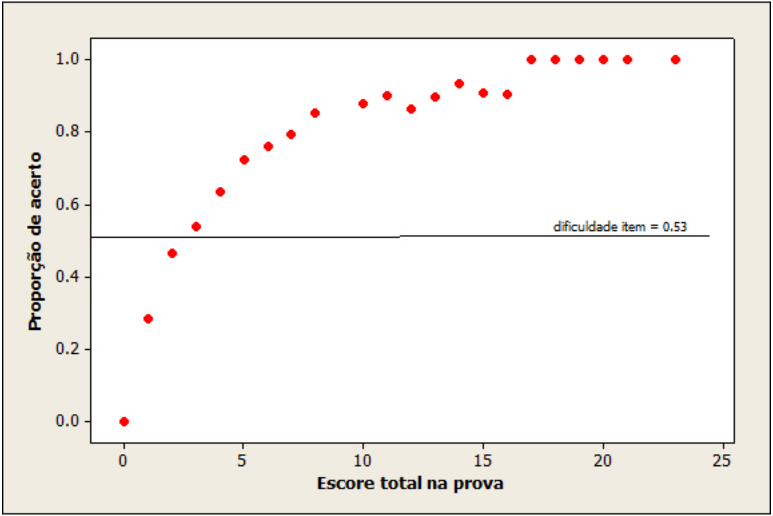
1. Introdução: Avaliações Educacionais

- Número de itens (dicotômicos, nominais, ordinais ou abertos)
- Número de categorias de resposta
- Auto-aplicativo ou entrevistador
- Número de dimensões
  (traços latentes - uni ou multidimensional)
- Grau de dificuldade dos itens/prova
- Número de provas (paralelas?)
- Indivíduos realizam a prova ao mesmo tempo?
- Número de populações
- Tipo de prova: via lápis e papel, teste informatizado, teste adaptativo


CASO 1 (Ex: ENEM)
3º EM
prova
A
CASO 2 (Ex: Estatística I EESC-USP)
Civil
prova
A
≠
prova
B
CASO 3 (Ex: SAEB)
3º EM
prova
A'
prova
B'
itens comuns
CASO 4
pop.1
pop.2
prova
A
CASO 5
pop.1
pop.2
prova
A
≠
prova
B
CASO 6 (Ex: SARESP)
4º EF
5º EF
prova
A
prova
B'
itens comuns

1. Introdução: Teoria Clássica

- atribuição de escore às alternativas de resposta dos itens:
  $\downarrow$ escore $\Leftrightarrow$ $\uparrow$ traço latente ou $\uparrow$ escore $\Leftrightarrow$ $\uparrow$ traço latente
- em testes de múltipla escolha (0=incorreta e 1=correta):
- Escore total (indivíduo): estimativa do traço latente
  número de itens corretos, varia de 0 a I
  (ou % de acerto, varia de 0 a 100%)
- Dificuldade (item): % de acertos, varia de 0 a 1 (ou 100%)
- Discriminação (item):
  % acertos grupo superior – grupo inferior, varia de -1 a 1
  Grupo superior: 27% com os escores mais altos.
  Grupo inferior 27% com os escores mais baixos.



Proporção de acerto
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Escore total na prova
dificuldade item = 0.53

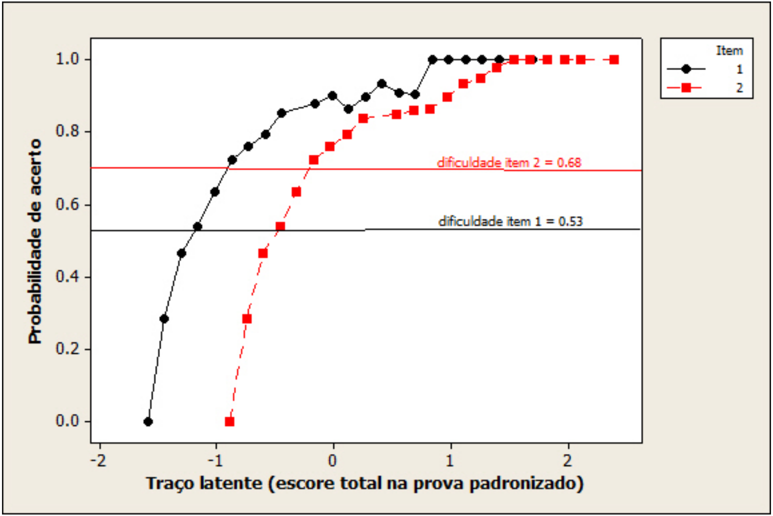
1. Introdução: Teoria Clássica

- Resultados dependem do particular conjunto de itens da prova (Prova - elemento central)
- Resultados dependentes do grupo de respondentes
- Compração entre indivíduos: somente com mesma prova ou provas paralelas
- Comparação proporção acertos entre séries?

1. Introdução: TRI

- Surgiu formalmente a partir dos trabalhos de Lord (1952) e Rasch (1960)
- Item - elemento central
- Permite a comparação entre indivíduos, mesmo submetidos a provas diferentes
- Analisa itens com diferentes escores para as categorias sem desbalancear a estimativa do traço latente
- 2 tipos de parâmetros: de itens e individuais (traços latentes)
- Modelos: probabilidade de determinada resposta ao item = f(parâmetros do item, traço latente)

1. Introdução: TRI



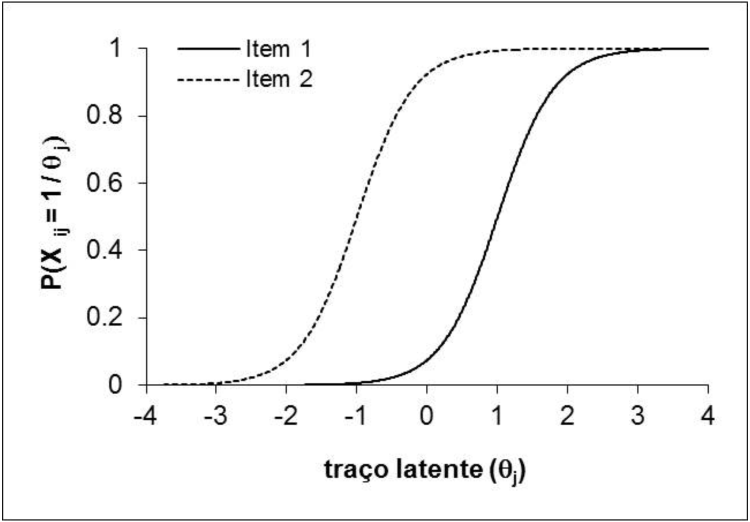
$X_i = 0$ ou 1: resposta do indivíduo ao item i (incorreta ou correta)
$X_i \sim$ Bernoulli (P_i)
$P_i = P(X_i = 1) = f(\theta, b_i)$,
sendo b_i a dificuldade do item i e θ, o traço latente do indivíduo.

Definições comuns de $f(\theta, b_i)$ na literatura:

$\Phi(.)$: fda da $N(0,1)$ - Modelos de ogiva normal
$\frac{1}{1+e^{-(\theta-b_i)}}$: função logística - Modelo (logístico) de Rasch

$$log \frac{P_i}{1-P_i} = -(\theta - b_i)$$

1. Introdução: TRI



Os modelos propostos dependem:

1. da natureza do item: dicotômicos, ordinais ou nominais
2. do número de populações envolvidas: apenas uma ou mais de uma população
3. da quantidade de traços latentes considerados: apenas um ou mais de um
4. Mais usual: Modelos logísticos unidimensionais para itens dicotômicos
   Se diferenciam pelo número de parâmetros utilizados para descrever o item:
   - 1 parâmetro = somente a dificuldade do item (modelo de Rasch);
   - 2 parâmetros = a dificuldade e a discriminação;
   - 3 parâmetros = a dificuldade, a discriminação e a probabilidade de acerto por indivíduos de baixo traço latente ("chute").

Avaliações Educacionais que usam a TRI (nacionais e internacionais)

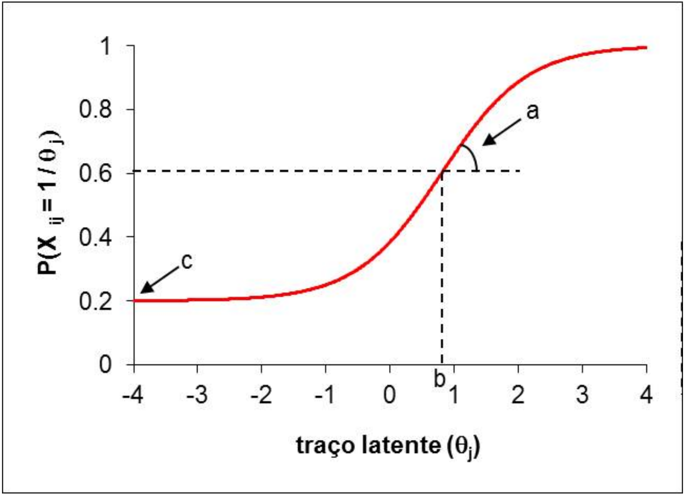
- ENEM
- SAEB
- ENCCEJA
- SARESP
- TOEFL
- GRE
- PISA

2. Modelos da TRI: ML3

$$P(X_{ij} = 1 \mid \theta_j, a_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + e^{-a_i(\theta_j - b_i)}},$$

i=1, ..., I (itens)
j=1, ..., n (indivíduos)
X_{ij}=1, se indiv j acerta o item i, e X_{ij}=0, c.c.
θ_j é o nível do traço latente do indiv j
a_i parâmetro de discriminação do item i,
derivada no ponto de inflexão
b_i parâmetro de dificuldade do item i,
se $b_i = \theta_j$, $P(X_{ij} = 1 \mid \theta_j, a_i, b_i, c_i) = (1 + c_i)/2$
c_i parâmetro de acerto ao acaso ("chute") do item i

Curva Caractrística de Item (CCI)



Função de Informaçao do Item

Pelas c.r. (devido à família exponencial):

$$I_i(\theta) = \frac{\left(\dfrac{\partial P_i(\theta)}{\partial \theta}\right)^2}{P_i(\theta)(1 - P_i(\theta))}$$

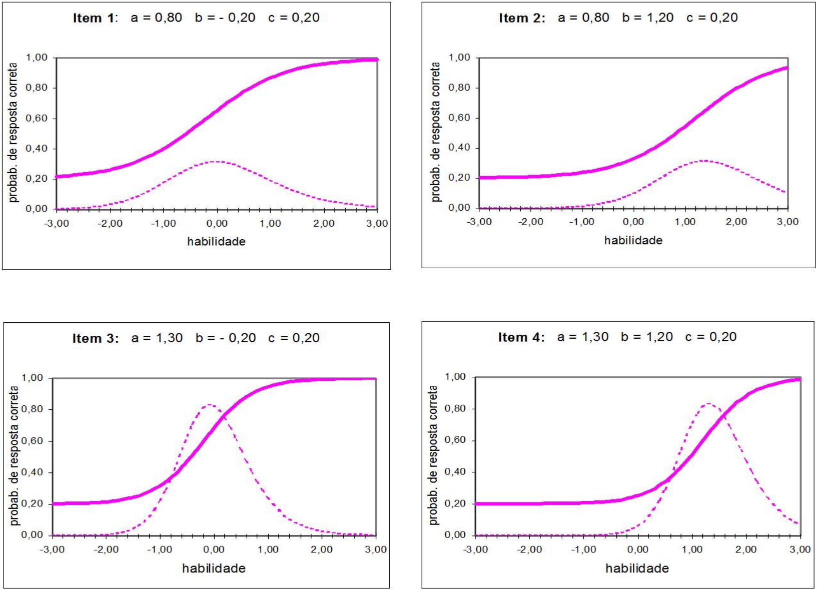
e

$$I(\theta) = \sum_{i=1}^{I} I_i(\theta),$$

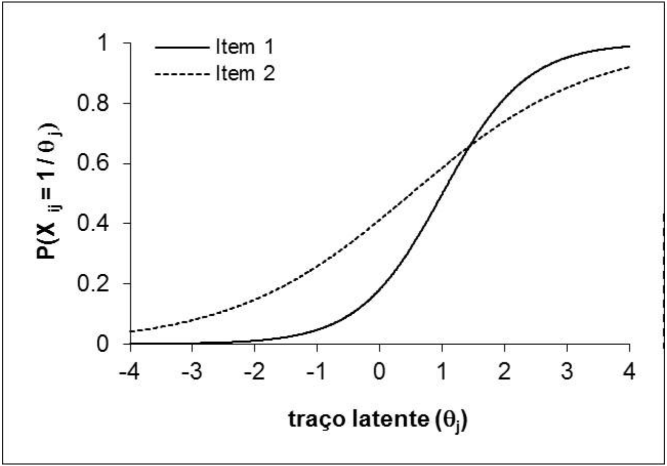
em que $P_i(\theta) = P(X_{ij} = 1 \mid \theta_j, a_i, b_i, c_i)$, para $\theta_j = \theta$.

Em particular, para o ML3: $I_i(\theta) = a_i^2 \frac{(1-P_i(\theta))}{P_i(\theta)} \left[\frac{P_i(\theta)-c_i}{1-c_i}\right]^2$.

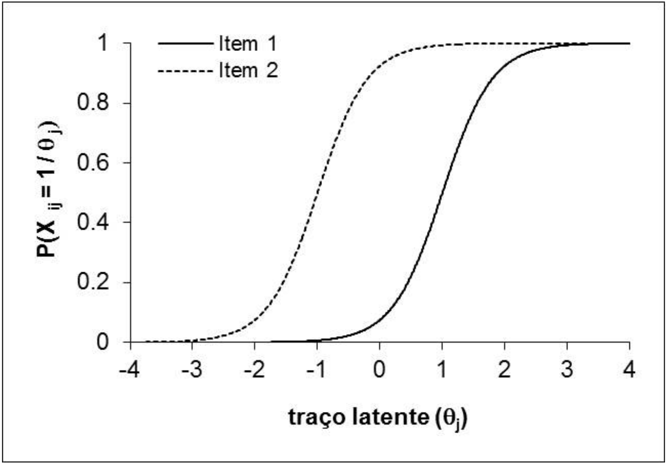
2. Modelos da TRI: Função de Informação do ML3



$$P(X_{ij} = 1 \mid \theta_j, a_i, b_i) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$



$$P(X_{ij} = 1 \mid \theta_j, b_i) = \frac{1}{1 + e^{-a(\theta_j - b_i)}}$$



2. Modelos da TRI: modelos de ogiva normal

$$P(X_{ij} = 1 \mid \eta_{ij}) = \int_{-\infty}^{\eta_{ij}} \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-t^2}{2}\right)} dt.$$

equivale ao modelo logístico

$$P(X_{ij} = 1 \mid \eta_{ij}) = \frac{1}{1 + e^{(-\eta_{ij})}}.$$

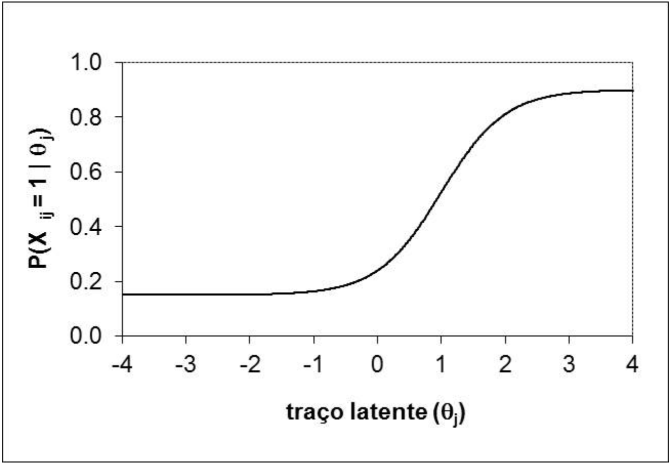
com
$\eta_{ij} = \theta_j - b_i$ no modelo de Rasch,
$\eta_{ij} = a_i(\theta_j - b_i)$ nos modelos de 2 e 3 parâmetros e

$$P(X_{ij} = 1 \mid \theta_j, a_i, b_i, c_i) = \int_{-\infty}^{a_i(\theta_j - b_i)} c_i + (1 - c_i) \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-t^2}{2}\right)} dt,$$

no ML3

2. Modelos da TRI: ML4

$$P(X_{ij} = 1 \mid \theta_j, a_i, b_i, c_i, \gamma_i) = c_i + \frac{(\gamma_i - c_i)}{1 + e^{-Da_i(\theta_j - b_i)}}$$



2. Modelos da TRI: Samejima - modelo de resposta gradual

$$P^{+}_{ik}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{ik})}}$$

k = 0,1, ..., m_i

m_i+1: n^o categorias do item i

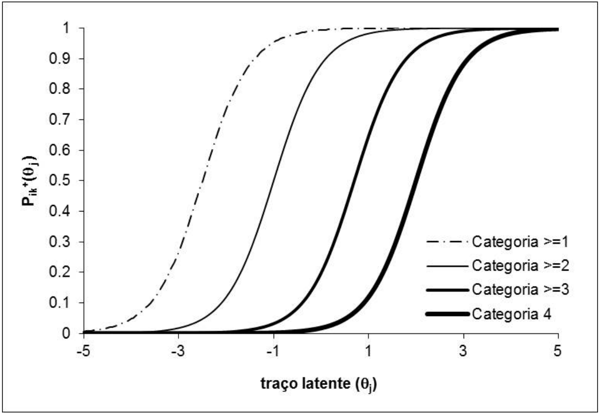
$P^{+}_{ik}(\theta_j)$: prob. de um indivíduo com traço latente θ_j escolher a categoria de resposta k ou qualquer outra de ordem acima de k no item i

a_i: parâmetro de discriminação comum a todas as categorias do item i

b_{ik}: parâmetro de gravidade que representa o nível latente necessário para a escolha da categoria de resposta acima de k com probabilidade igual a 0.50 $(b_{i1} \leq b_{i2} \leq ... \leq b_{im_i})$

$P^{*}_{i0}(\theta_j) = 1$ e $P^{*}_{ik+1}(\theta_j) = 0$.

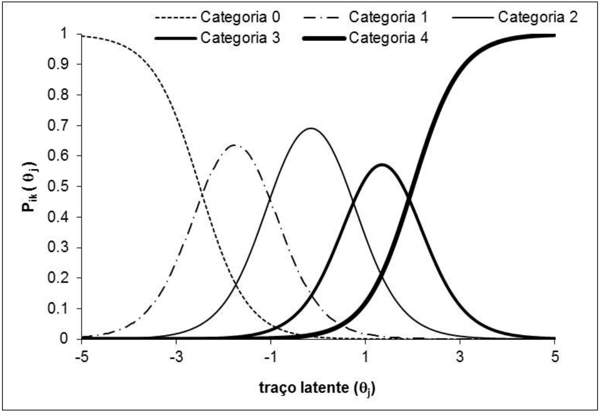
2. Modelos da TRI: Samejima - modelo de resposta gradual



$$P_{ik}(\theta_j) = P_{ik}^{+}(\theta_j) - P_{ik+1}^{+}(\theta_j)$$

2. Modelos da TRI: Samejima - modelo de resposta gradual

$$P_{ik}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{ik})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{ik+1})}}$$

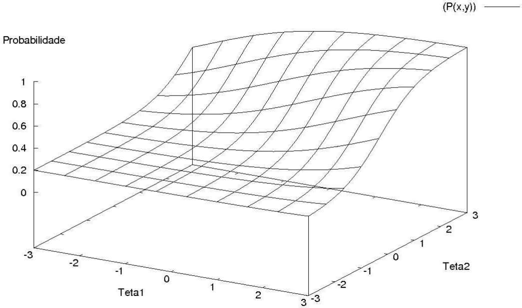


2. Modelos da TRI: Multidimensional (compensatório)

$$P(X_i = 1|\boldsymbol{\theta}, \mathbf{a}_i, b_i, c_i) = c_i + (1 - c_i)\frac{1}{1 + \exp\left[-\sum_{k=1}^{p} a_{ki}\theta_k + b_i\right]},$$

com $\mathbf{a}_i = (a_{1i}, \ldots, a_{pi})$, p: número de traços latentes e $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.

Para $a_1 = 0,8$, $a_2 = 1,4$, $b = -2,0$ e $c = 0,2$



2. Modelos da TRI: Outros modelos

Ordinais

- Modelo de Escala Gradual
- Modelo de Crédito parcial
- Modelo de Crédito Parcial Generalizado
- Modelo Nominal

Referência: Andrade, Tavares e Cunha (2000)

Multidimensionais

- logísticos
- ogiva
- não compensatórios
- bifatorial

Referência: Reckase (1997), Li and Lissitz (2000), Rost and Carstensen (2002) e Gardner et al (2002)

2. Modelos da TRI: Múltiplos grupos

$$P(X_{ijk} = 1 \mid \theta_{jk}, a_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + e^{-a_i(\theta_{jk} - b_i)}},$$

i=1, ..., I (itens)
j=1, ..., n_k (indivíduos no grupo k)
k=1, ..., g (grupos)

Referência: Bock, R.D., Zimowski, M.F. (1997). *Multiple group IRT. In Handbook of Modern Item Response Theory*. W.J. van der Linden and R.K. Hambleton Eds. New York: Springer-Verlag

3. Estimação

Tipos de parâmetro $\begin{cases} \text{indivíduos: } \theta_j \\ \text{itens: } \zeta_i = (\ a_i, \quad b_i, \quad c_i\ )^t, \text{no ML3, por exemplo} \end{cases}$

Suposições:

- indep entre respostas de $\neq$ indiv
- indep entre respostas de $\neq$ itens condicionada a θ_j
- mesma probabilidade de seleção amostral
- dados omissos são não informativos

	θ_j	ζ_i
MV	X	conhecido
MV	conhecido	X
MV conjunta	X	X
MV marginal		X
MCMC	X	X

MV ou Bayesiano EAP ou MAP

Função de verossimilhança dos modelos unidimensionais dicotômicos:

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

- Parâmetros de itens conhecidos
  $L(\boldsymbol{\theta}) : \frac{\partial log L(\boldsymbol{\theta})}{\partial \theta_j} = 0$
- Traços latentes conhecidos
  $L(\zeta) : \frac{\partial log L(\zeta)}{\partial \zeta_i} = \mathbf{0}$

Há necessidade de uso de processo iterativo
Não está definido para alguns padrões de resposta

3. Estimação: Definição da escala de medida

Falta de identificabilidade: θ, ζ desconhecidos

Exemplo : indivíduo com θ = 1,20 na escala (0 ; 1) .
Qual sua habilidade na escala (200 ; 40) ?

$$a(\theta - b) = (a / 40) [(40 \times \theta + 200) - (40 \times b + 200)] = a^* (\theta^* - b^*)$$

Resposta : θ^* = 248

1. $\theta^* = 40 \times \theta + 200$
2. $b^* = 40 \times b + 200$
3. $a^* = a / 40$
4. $P(Xi=1 | \theta) = P(Xi=1 | \theta^*)$

Solução: fdp para θ: $g(\theta/\tau)$

Etapa 1: Tornar a verossimilhança independente de θ_j e estimar ζ_i
Etapa 2: Estimar θ_j, considerando-se ζ_i conhecidos

População de indivíduos $\rightarrow$ seleção aleatória: $\theta_j \sim g(\theta \mid \boldsymbol{\eta})$
$\theta_j \sim N(0,1), \boldsymbol{\eta} = (\mu = 0, \sigma^2 = 1)$: define a métrica

Função de verossimilhança marginal:

$$L(\zeta, \boldsymbol{\eta}) = \prod_{j=1}^{n} \int_{\mathbb{R}} \prod_{i=1}^{I} P(X_{ij} = x_{ij} \mid \theta, \zeta)\, g(\theta \mid \boldsymbol{\eta})\, d\theta$$

$$\frac{\partial log L(\zeta, \boldsymbol{\eta})}{\partial \zeta_i} = 0$$

Proposta Bock & Aitkin: estimar itens individualmente
Reestruturação EE + Hermite-Gauss: nós $\bar{\theta}_k$, k=1, ..., q.

Derivação das fórmulas para o modelo ML1:
Slides Prof. Caio Lucidius Naberezny Azevedo - UNICAMP
http://www.ime.unicamp.br/~cnaber/Material_TRI.htm

Estimação Frequentista - pag 43 à 47
Estimação Bayesiana - pag 1 à 7

3. Estimaçao: MCMC

Função de verossimilhança:

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

Função de distribuição *a posteriori*:

$$f(\boldsymbol{\theta}, \zeta) \propto L(\boldsymbol{\theta}, \zeta) g(\boldsymbol{\theta} \mid \boldsymbol{\eta}) h(\zeta \mid \boldsymbol{\tau})$$

$\uparrow$

distribuição estacionária de uma cadeia de Markov

com $g(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ e $h(\zeta \mid \boldsymbol{\tau})$ distribuições *a priori*

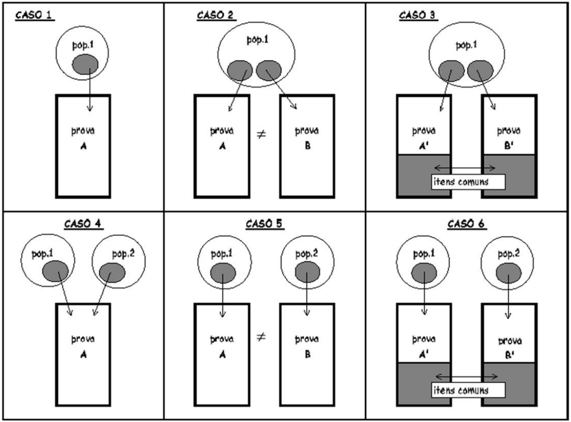
- Diferentes grupos: séries, turnos, países
- Grupos definidos previamente

$$P(X_{ijk} = 1 \mid \theta_{jk}, a_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + e^{-a_i(\theta_{jk} - b_i)}},$$

i=1, ..., I (itens)
j=1, ..., n_k (indivíduos no grupo k)
k=1, ..., g (grupos)

- Estimação: $\theta_{jk} \mid \boldsymbol{\eta}_k \sim N(\mu_k, \psi_k)$
- Identificabilidade: $\mu_1 = 0, \psi_1 = 1$
- Estimam-se (μ_k, ψ_k), para $k = 2, ..., g$

- Colocar itens de provas distintas ou habilidades de populações diferentes numa mesma escala, podendo ser comparados



Provas: - apenas itens novos
- apenas itens já calibrados
- itens novos e já calibrados

4. Equalização: itens novos

- via TRI: estimação de todos os dados conjuntamente equaliza em todos os casos MENOS no 5
- caso 5: resultados em métricas diferentes; sem comparações
- casos 4 e 6: modelos para múltiplos grupos
- caso 6: representa o melhor exemplo do uso da equalização e o maior avanço da TRI sobre a Teoria Clássica

4. Equalização: caso 6

- Quantos itens comuns?
- Depende do tipo de equalização e da qualidade dos itens comuns
- Quanto maior o parâmetro de discriminação dos itens e quanto mais próximos estiverem os parâmetros de dificuldade dos itens da média da população avaliada, menor o número de itens comuns para uma boa equalização
- Ex: 2 provas de 30 itens - pelo menos 6 em comum

4. Equalização: SAEB

SAEB: Sistema Nacional de Avaliação da Educação Básica

- bienal desde 1995
- séries: 4a. e 8a. do EF e 3a. do EM
- uma análise para cada disciplina
- itens de múltipla escolha (95: itens 0,1,2)
- um grande número de itens para cobrir a grade curricular
- provas diferentes para uma mesma série/disciplina (BIB)
- aluno faz somente uma das provas de uma das disciplinas
- http://www.inep.gov.br/basica/saeb/

4. Equalização: SAEB

- O número de itens requerido pelos especialistas, para cada série e disciplina, é maior do que um estudante pode responder em 2 horas.
- Equalização: obter resultados comparáveis (mesma escala) para as 4a., 8a. and 3a. séries e também ao longo do tempo.
- Matemática, 3a. série: 169 itens.
  - 13 blocos com 13 itens cada (169=132)
  - Provas: 26 cadernos de provas com 3 blocos
    total de 39=3x13 itens
  - 130 itens "não apresentados" a cada um dos alunos
  - Cadernos possuem itens comuns com alguns dos outros
  - Blocos comuns e/ou itens já aplicados em anos anteriores
  - Blocos da 4a. série na 8a. série
  - Blocos da 8a. série na 3a. série do ensino médio

Blocos Incompletos Balanceados - BIB

Cadernos de provas	Conjuntos de itens			Cadernos de provas	Conjuntos de itens		
1	1	2	5	**14**	1	3	8
2	2	3	6	**15**	2	4	9
3	3	4	7	**16**	3	5	10
4	4	5	8	**17**	4	6	11
5	5	6	9	**18**	5	7	12
6	6	7	10	**19**	6	8	13
7	7	8	11	**20**	7	9	1
8	8	9	12	**21**	8	10	2
9	9	10	13	**22**	9	11	3
10	10	11	1	**23**	10	12	4
11	11	12	2	**24**	11	13	5
12	12	13	3	**25**	12	1	6
13	13	1	4	**26**	13	2	7

> Cada conjunto de 13 itens aparece em 6 cadernos de provas

> Cada conjunto de itens aparece duas vezes em cada uma das 3 posições nos cadernos de provas

> Um par de conjuntos de itens aparece somente uma vez em um caderno de provas

4. Equalização: itens já calibrados

- Desejamos apenas estimar as habilidades dos indivíduos
- Situação comum devido à criação de Bancos de Itens
- conjunto de itens que já foram testados e calibrados a partir de um número significativo de sujeitos de uma dada população
- parâmetros "conhecidos"
- As habilidades estimadas a partir de itens do banco estarão na mesma métrica do grupo de indivíduos utilizados na calibração inicial

4. Equalização: itens novos + já calibrados

- Situação comum devido à ampliação de Bancos de Itens
- continuamente em formação / itens saem e itens entram no banco
- Problema:
  itens novos devem ser calibrados na mesma métrica de itens do banco: programas computacionais específicos
- Objetivos :
  criar e testar itens novos
  comparar o desempenho da rede pública estadual de São Paulo com o desempenho nacional, por ex

4. Equalização: *a posteriori*

- Pode ser feita quando há itens comuns entre 2 populações
- Calibra-se separadamente 2 conjuntos de itens, que foram submetidos a 2 populações de interesse
- Para os itens comuns, teremos 2 conjuntos de estimativas, cada uma na métrica de suas respectivas populações

4. Equalização: *a posteriori*

- Estabelece-se algum tipo de relação (preferencialmente simétrica) que permita colocarmos os parâmetros de um dos conjuntos de itens na escala do outro
- Utiliza-se essa relação para transformar os parâmetros de todos os itens (comuns e não comuns) de um conjunto na escala do outro
- Com todos os itens na mesma métrica, pode-se estimar as habilidades de todos os respondentes, que também estarão na mesma escala

- Pela propriedade de invariância, temos:
  $b_1 = \alpha * b_2 + \beta$ e $a_1 = \frac{1}{\alpha} * a_2$
- Uma vez determinados os coeficientes α e β, as estimativas dos parâmetros dos itens do grupo 2 podem facilmente ser colocados na mesma escala das estimativas do grupo 1

É um método simétrico

$\alpha = \frac{S_1}{S_2}$ e $\beta = \bar{X}_1 - \bar{X}_2$, em que

S_1 e S_2 são os desvios padrão e
$\bar{X}_1$ $\bar{X}_2$ são as médias amostrais

das estimativas dos parâmetros de dificuldade dos itens comuns nos grupos 1 e 2, respectivamente.

Para a equalização das habilidades:

$$\theta_1 = \alpha * \theta_2 + \beta$$

em que θ_i é a habilidade na escala do grupo i.

Perguntas frequentes:

1. Quem acerta mais itens tem sua estimativa de habilidade maior?
2. Como a presença do parâmetro de "acerto ao acaso" (c) influencia na estimativa da habilidade?
3. Responder "fora do padrão esperado" (acertar as difíceis e erras as fáceis) diminui a estimativa da habilidade?
4. Duas estimativas de habilidade de um mesmo indivíduo feitas a partir de respostas a provas diferentes geram valores equivalentes?

Samejima, F. (2000). Logistic positive exponent family of models: virtue of asymmetric item characteristic curves. **Psychometrika, 65**. 319-335

- EE para θ: $\sum_{i=1}^{I} a_i X_i = \sum_{i=1}^{I} a_i P_i(\theta)$
- $\hat{\theta} \uparrow$ com $\sum_{i=1}^{I} a_i$
- 5 itens: $\mathbf{a}$ =(1 1 1 1 1), $\mathbf{b}$ =(-3 -1,5 0 1,5 3)
- No modelo de ogiva normal:
  - para apenas 1 item correto: $\hat{\theta} \uparrow$ quando b item acertado $\uparrow$
  - para apenas 1 item incorreto: $\hat{\theta} \uparrow$ quando b item errado $\uparrow$
  - não há uma regra simples determinante da posição relativa de duas estimativas de habilidades para diferentes padrões de respostas
- No modelo logístico:
  - essa contradiçao não ocorre
  - quanto maior for a do item correto, maior será $\hat{\theta}$ - para itens com mesma discriminação, $\hat{\theta} \uparrow$ com o número de acertos
  - porém, a dificuldade do item não é levada em consideração para estimar θ

	Response Pattern	Normal Ogv.	Logistic
1	00000	neg. infinity	neg. infinity
2	10000	−2.28385	−2.28753
3	01000	−2.27016	−2.28753
4	00100	−1.84831	−2.28753
5	00010	−1.34811	−2.28753
6	01100	−1.15759	−0.75260
7	00001	−0.86577	−2.28753
8	11000	−0.75034	−0.75260
9	10100	−0.75021	−0.75260
10	01010	−0.75013	−0.75260
11	00110	−0.75011	−0.75260
12	00101	−0.36062	−0.75260
13	10010	−0.34310	−0.75260
14	01001	−0.27309	−0.75260
15	00011	−0.19116	−0.75260
16	01110	−0.15292	0.75260
17	10001	0.15292	−0.75260
18	00111	0.19116	0.75260
19	01101	0.27309	0.75260
20	10110	0.34310	0.75260
21	01011	0.36062	0.75260
22	10011	0.75011	0.75260
23	10101	0.75013	0.75260
24	11010	0.75021	0.75260
25	11100	0.75034	0.75260
26	01111	0.86577	2.28753
27	11001	1.15759	0.75260
28	10111	1.34811	2.28753
29	11011	1.84831	2.28753
30	11101	2.27016	2.28753
31	11110	2.28385	2.28753
32	11111	pos. infinity	pos. infinity

Minhas simulações: Resultados Simulações no Excel
Artigo Caio

Perguntas frequentes:

1. Quem acerta mais itens tem sua estimativa de habilidade maior? SIM
2. Como a presença do parâmetro de "acerto ao acaso" (c) influencia na estimativa da habilidade? $c \uparrow$ implica $\hat{\theta} \downarrow$
3. Responder "fora do padrão esperado" (acertar as difíceis e erras as fáceis) diminui a estimativa da habilidade? depende do modelo: no ML2 NÃO, mas no ML3 SIM
4. Duas estimativas de habilidade de um mesmo indivíduo feitas a partir de respostas a provas diferentes geram valores equivalentes? SIM

6. Interpretação da Escala

- métrica arbitrária para parâmetros dos itens e habilidades
- define a ordem, mas não o significado prático
- ex: na escala (0,1), qual a interpretação de $\theta = -0,8$ versus $\theta = 1,5$

Para interpretação:

- criação de escalas de conhecimento que tornam possível a interpretação pedagógica dos resultados
- definição de níveis âncora e itens âncora

Níveis âncora: pontos selecionados na escala da habilidade para serem interpretados pedagogicamente

Item âncora

Considere 2 níveis âncora consecutivos θ_1 e θ_2, com $\theta 1 < \theta 2$
Um item i é âncora no nível θ_2 se, e somente se:

- $P(X_i = 1 \mid \theta = \theta_2) \geq 0,65$
- $P(X_i = 1 \mid \theta = \theta_1) < 0,50$
- $P(X_i = 1 \mid \theta = \theta_2) - P(X_i = 1 \mid \theta = \theta_1) \geq 0,30$

	Níveis Âncora									
Item	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
1							sim			
2					sim					
3									sim	
4						sim				
5		sim							sim	
6				sim						
7						sim				
8								sim		
9							sim			
10				sim						

6. Interpretação da Escala: um exemplo

Slides Raquel da Cunha Valle
Fundação Carlos Chagas

7. Aplicação: *Softwares Computacionais*

- SAS, SPSS, Stata? Não
- Programas individuais em Splus, R e SAS
- No R: ltm, mirt
- Testfact
- Bilog-MG
- Xcalibre
- Multilog
- Parscale
- Noharm
- WinBUGS
  Bayesian Modeling - Jorge L. Bazàn
  (http://argos.pucp.edu.pe/~jlbazan/software.html)

7. Aplicação: PISA

- Programa Internacional de Avaliação dos Estudantes (PISA) é aplicado a alunos na faixa dos 15 anos, idade na qual a maioria dos estudantes finalizam a escolaridade básica obrigatória
- realizado a cada 3 anos desde 2000
- disciplinas: Leitura, Matemática e Ciências
- planejamento BIB
- amostragem complexa: estrato (UF) e 3 subestratos (pública/privada, rural/urbana e IDH)
- modelo de Rasch

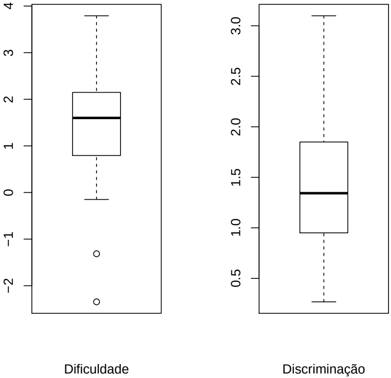
7. Aplicação: PISA

- PISA 2009 - Brasil
- 20127 estudantes brasileiros participaram (destes, 4000 prova informatizada)
- Apenas as questões de matemática (35 questões)
- 6 provas diferentes (B8=B12 e B10=B27)

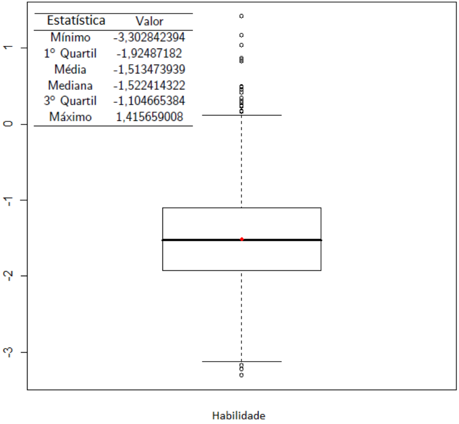
Tabela: Estimativas dos parâmetros de itens - PISA 2012 Brasil

Item	b	Erro Padrão	a	Erro Padrão
MAT01	-0,150	0,0304	1,020	0,0456
MAT02	1,773	0,0533	1,633	0,0754
MAT03	0,791	0,0279	1,707	0,0662
MAT04	0,772	0,0266	1,830	0,0713
MAT05	2,281	0,0809	1,437	0,0749
MAT06	1,097	0,0439	1,136	0,0488
...	...	...	...	...
MAT35	1,793	0,0479	2,095	0,1034
Média	1,469		1,434	

7. Aplicação: PISA



7. Aplicação: PISA



Estatística	Valor
Mínimo	-3,302842394
1º Quartil	-1,92487182
Média	-1,513473939
Mediana	-1,522414322
3º Quartil	-1,104665384
Máximo	1,415659008

Tabela: Itens Âncora

Item	Níveis Âncora (θ)						
	-3	-2	-1	0	1	2	3
MA1							
MA2						sim	
MA3					sim		
MA4							sim
MA5						sim	
MA6							sim
...							
MA35						sim	

- PorSimples
- BDI
- CAT - BDI

FIM