Lesson 8: Item response theory models (part 3)

Patrícia Martinková

Department of Statistical Modelling
Institute of Computer Science, Czech Academy of Sciences

Institute for Research and Development of Education
Faculty of Education, Charles University, Prague

NMST570, November 27, 2018

Outline

Review: Dichotomous IRT models

- Dichotomous IRT models
  - Rasch model
  - 1PL IRT model
  - 2PL IRT model
  - 3PL IRT model
  - 4PL IRT model

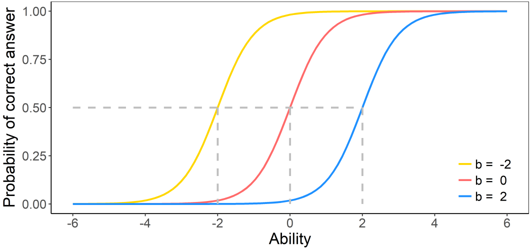
- Item Characteristic Curve (ICC)
  - Item Response Function (IRF)
- Item Information Function (IIF)
- Test Information Function (TIF)

- Likelihood function
- Parameter estimation: JML, CML, MML, Bayesian methods
- Model fit, item fit, person fit (see also Ames and Penfield (2015))

Review: 1PL IRT Model

$$\pi_{ij} = P(Y_{ij} = 1|\theta_i, a, b_j) = \frac{\exp[a(\theta_i - b_j)]}{1 + \exp[a(\theta_i - b_j)]}$$

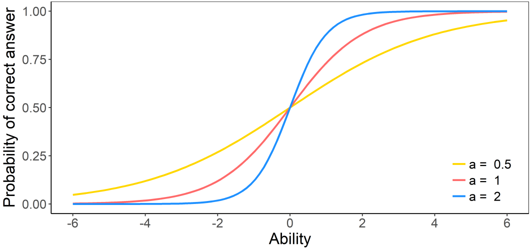
θ_i ability of person i for $i = 1, \dots, I$
b_j difficulty of item j (location of inflection point) for $j = 1, \dots, J$
a discrimination common for all items (slope at inflection point)



Review: 2PL IRT Model

$$\pi_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j) = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]}$$

θ_i ability of person i for $i = 1, \ldots, I$
b_j difficulty of item j (location of inflection point)
a_j discrimination of item j (slope at inflection point) for $j = 1, \ldots, J$



Polytomous Items

Example 1: Quality of Life Questionnaire

"I am satisfied with my life."

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

- Ordered item (Likert-type question)

Polytomous Items

Example 2: Reasoning Ability Instrument

"In what way are orange and a banana alike?"

2pts Provides pertinent general categorization (e.g. "Both are fruit.")

1pts Provides one or more common properties (e.g. "Both are food.")

0pts Provides specific properties of each member of pair, or wrong answer (e.g. "Both are round.")

Polytomous Items

Example 3: Math Misconceptions

$-6 - (-10) = ?$

a -16
b -4
c 4

For this item:

- Alternatives (distractors) can provide useful information for the diagnosis of mathematical misconceptions.
- Dichotomizing would lead to discarding this information.

Nominal and Ordinal Models - Categorization

Difference models

- Setting mathematical form to cumulative probabilities

Examples:

- Graded Response Model (GRM; Samejima, 1970)
- Modified Graded Response Model (MGRM; Muraki, 1990)

Divide-by-total models

- Response category probabilities are defined as the ratio between category-related functions and their sum

Examples:

- Partial Credit Model (PCM; Masters, 1982)
- Generalized Partial Credit Model (GPCM; Muraki, 1992)
- Rating Scale Model (RSM; Andrich, 1978)
- Nominal Response Model (NRM; Bock, 1972)

Nominal and Ordinal Models - Categorization

Cumulative logit models

- Assumes linear form of cumulative logits
  - Graded Response Model (GRM; Samejima, 1970)
  - Modified Graded Response Model (MGRM; Muraki, 1990)

Adjacent-categories logits models

- Assumes linear form of adjacent logits
  - Generalized Partial Credit Model (GPCM; Muraki, 1992)
  - Partial Credit Model (PCM; Masters, 1982)
  - Rating Scale Model (RSM; Andrich, 1978)

Baseline-category logit models

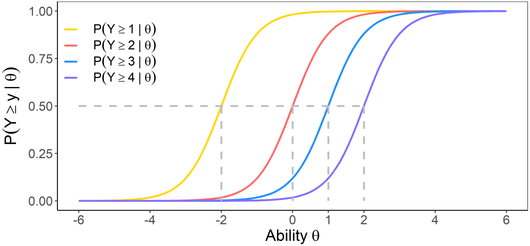
- Assumes linear form of baseline-category logits
  - Nominal Response Model (NRM; Bock, 1972)

Graded Response Model (GRM)

GRM models *cumulative probabilities* (Samejima, 1970):

$$\pi^*_{ijk} = P(Y_{ij} \geq k | \theta_i, a_j, b_{jk}) = \frac{\exp[a_j(\theta_i - b_{jk})]}{1 + \exp[a_j(\theta_i - b_{jk})]}$$

b_{jk} locations of item j (location of inflection points of cumulative functions)

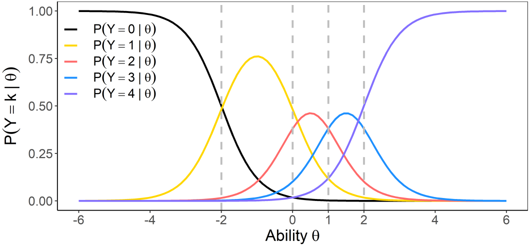


$$a = 2, b_1 = -2, b_2 = 0, b_3 = 1, b_4 = 2$$

Graded Response Model (GRM)

Category probabilities:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i, a_j, b_{jk}) = \pi^*_{ijk} - \pi^*_{ij(k+1)}$$



$$a = 2, b_1 = -2, b_2 = 0, b_3 = 1, b_4 = 2$$

Partial Credit Model (PCM)

PCM (Masters, 1982) models *adjacent-categories* with 1PL model:

$$\log\left(\frac{P(Y_j = 1|\theta)}{P(Y_j = 0|\theta)}\right) = \log\left(\frac{\pi_{j1}}{\pi_{j0}}\right) = \theta - \delta_{j1}$$

$$\log\left(\frac{\pi_{j2}}{\pi_{j1}}\right) = \theta - \delta_{j2}$$

$$\log\left(\frac{\pi_{j3}}{\pi_{j2}}\right) = \theta - \delta_{j3}$$

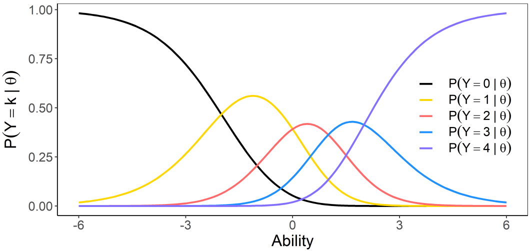
δ_{jt} are threshold parameters

- correspond to ability levels for which the response categories intersect
- constrained to sum to zero (or denominator od π_{ijk} is constrained)

Partial Credit Model (PCM)

Category probabilities:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i, \delta_{jk}) = \frac{\exp \sum_{t=0}^{k}(\theta_i - \delta_{jt})}{\sum_{r=0}^{K_j} \exp \sum_{t=0}^{r}(\theta_i - \delta_{jt})}$$



$$d_1 = -2, d_2 = 0, d_3 = 1, d_4 = 2$$

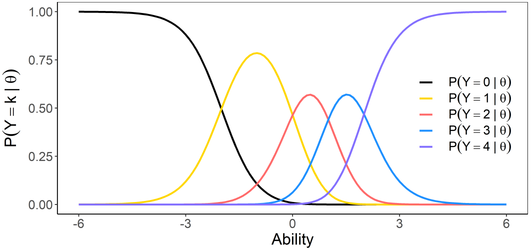
Generalized Partial Credit Model (GPCM)

GPCM (Muraki, 1992) models *adjacent-categories* with 2PL model:

$$\log\left(\frac{\pi_{jk}}{\pi_{j(k-1)}}\right) = \alpha_j(\theta - \delta_{jk}), k = 1, ..., K_j$$

Category probabilities:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i, \alpha_j, \delta_{jk}) = \frac{\exp\sum_{t=0}^{k}\alpha_i(\theta_i - \delta_{jt})}{\sum_{r=0}^{K_j}\exp\sum_{t=0}^{r}\alpha_i(\theta_i - \delta_{jt})}$$



$$a = 2, d_1 = -2, d_2 = 0, d_3 = 1, d_4 = 2$$

Rating Scale Model (RSM)

- RSM (Andrich, 1978) is restricted GPCM
- all items share the same rating scale structure
- all test items have exactly K categories
- the threshold parameters can be split into item-specific location and response threshold: $\delta_{it} = \delta_i + \lambda_t$

- *Category probabilities:*

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i, \delta_j, \lambda_t) = \frac{\exp \sum_{t=0}^{k} (\theta_i - [\delta_i + \lambda_t])}{\sum_{r=0}^{K_j} \exp \sum_{t=0}^{r} (\theta_i - [\delta_i + \lambda_t])}$$

Nominal Response Model (NRM)

NRM (Bock, 1972) models *baseline*-categories logits:

$$\log\left(\frac{\pi_{jk}}{\pi_{j0}}\right) = \alpha_{jk}\theta + c_{jk}$$

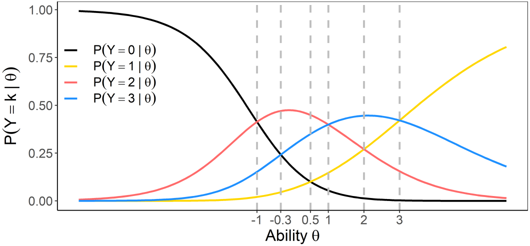
- *baseline* chosen arbitrary, e.g. first alternative or correct answer
- item/category-specific intercepts and slopes
- traditional (slope/intercept) parametrization

- category boundaries can be retrieved from αs and cs
- setting differences of αs equal to 1 leads to Generalized Partial Credit Model

Nominal Response Model (NRM)

Category probabilities:

$$\pi_{ijk} = P(Y_{ij} = k|\theta_i, \alpha_{j0}, \ldots, \alpha_{jK_j}, c_{j0}, \ldots, c_{jK_k}) = \frac{\exp(\alpha_{jk}\theta_i + c_{jk})}{\sum_{r=0}^{K_j} \exp(\alpha_{jr}\theta_i + c_{jr})}$$

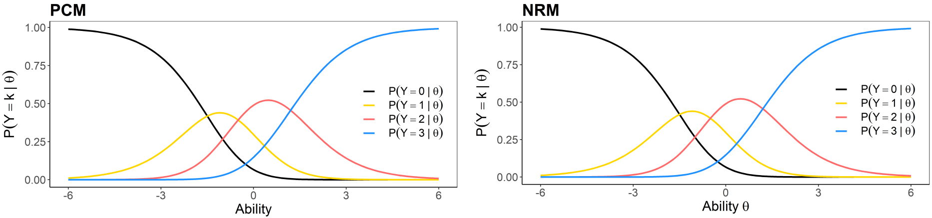


$$a_1 = 2, d_1 = -1, a_2 = 1, d_2 = 1, a_3 = 1.5, d_3 = 0.5$$

Comparing PCM and NRM

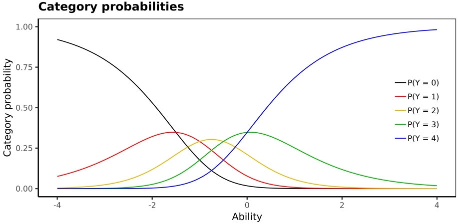
PCM with $\delta_1 = -1.5, \delta_2 = -0.5$ and $\delta_3 = 1.2$
NRM with $\alpha_1 = 1, \delta_1 = 1.5, \alpha_2 = 2, \delta_2 = 2,$ and $\alpha_3 = 3, \delta_3 = 0.8$

Choice	PCM $\sum_{t=0}^{k}(\theta - \delta_t)$	NRM $\alpha_k\theta + c_k$
$k = 0$	$\theta - \delta_0 = 0$	$\alpha_0\theta + c_0 = 0$
$k = 1$	$\theta - \delta_0 + \theta - \delta_1 = \theta + 1.5$	$\alpha_1\theta + c_1 = 1\theta + 1.5$
$k = 2$	$\theta - \delta_0 + \theta - \delta_1 + \theta - \delta_2 = 2\theta + 2$	$\alpha_2\theta + c_2 = 2\theta + 2$
$k = 3$	$\theta - \delta_0 + \theta - \delta_1 + \theta - \delta_2 + \theta - \delta_3 = 3\theta + 0.8$	$\alpha_3\theta + c_3 = 3\theta + 0.8$

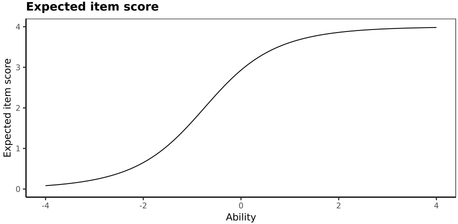


Expected item score

Category response function:



Expected item score:



Ordinal/Nominal Models - Other Topics

Analogously as for dichotomous models, also for ordinal/nominal models:

- Parameter estimation
- Abilities estimation
- Item information function
- Test information function
- Reliability

Applications:

- IRT Linking and Equating
- Differential Item Functioning
- Computerized Adaptive testing

Other Models

- Multidimensional models
- Testlet models
- Models for response times
- Nonparametric models
- Models for nonmonotone items
- Hierarchical response models
- Generalized modelling approaches

Vocabulary

- Item Characteristic Curve (ICC)
- Item Response Fuction (IRF)
- Item Information Function (IIF)
- Test Information Function (TIF)
- Likelihood function
- Parameter estimation: JML, CML, MML, Bayesian approaches
- Model fit, item fit, person fit

- 1PL, 2PL, 3PL, 4PL IRT models
- Graded Response Model (GRM)
- Partial Credit Model (PCM)
- Generalized Partial Credit Model (GPCM)
- Rating Scale Model (RSM)
- Nominal Response Model (NRM)

Thank you for your attention!

www.cs.cas.cz/martinkova

References

A. J. Ames and R. D. Penfield. An ncme instructional module on item-fit statistics for item response theory models. *Educational Measurement: Issues and Practice*, 34(3):39–48, 2015.

D. Andrich. A rating formulation for ordered response categories. *Psychometrika*, 43(4):561–573, 1978.

R. D. Bock. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1):29–51, 1972.

G. N. Masters. A rasch model for partial credit scoring. *Psychometrika*, 47(2): 149–174, 1982.

E. Muraki. Fitting a polytomous item response model to likert-type data. *Applied Psychological Measurement*, 14(1):59–71, 1990.

E. Muraki. A generalized partial credit model: Application of an em algorithm. *ETS Research Report Series*, 1992(1), 1992.

F. Samejima. Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 35(1):139–139, 1970.