Lesson 6: Item response theory models

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

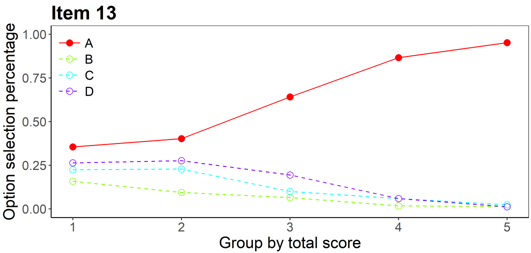
NMST570, November 13, 2018

Outline

Review: Traditional Item Analysis

Traditional item analysis describes item properties by

- percentages of correct response
- proportions of those who selected given distractor
- differences of percentages for groups by total score
- correlations of item score with total score



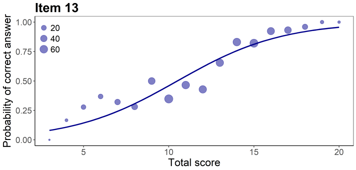
Review: Logistic Regression

Item properties described by parameters β_{0j} and β_{1j} of logistic function

$$\pi_{ij} = P(Y_{ij} = 1 | X_i, \beta_{0j}, \beta_{1j}) = \frac{\exp(\beta_{0j} + \beta_{1j} X_i)}{1 + \exp(\beta_{0j} + \beta_{1j} X_i)}$$

Also can be written as:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_{0j} + \beta_{1j} X_i$$



Notes:

- Linear model is related to response variable via a link function (GLM)
- Link functions: logit, probit (inverse of the cumulative distribution function)

Logistic Regression - IRT parametrization, Z scores

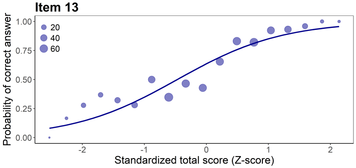
Logistic regression with IRT parametrization

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

b_j difficulty of item j
a_j discrimination of item j
Z_i standardized total score of person i

Also can be written as:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = a_j(Z_i - b_j)$$



Nonlinear Regression

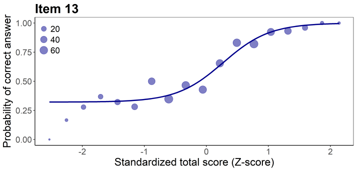
$$\pi_{ij} = P(Y_{ij} = 1 | Z_i, a_j, b_j, c_i) = (1 - c_i) \frac{\exp[a_j(Z_i - b_j)]}{1 + \exp[a_j(Z_i - b_j)]}$$

b_j difficulty of item j

a_j discrimination of item j

c_j probability of guessing of item j

Z_i standardized total score of person i



Notes:

- Not a GLM

Introduction to Item Response Theory

Framework for

- estimating *latent traits* (ability levels) θ

  by means of *manifest* (observable) variables (item responses)

  and appropriate *psychometric* (statistical) model

Notes:

- Ability θ now treated as random variable
- Items: dichotomous, polytomous, multiple-choice, ...
- IRT model: describes probability of (correct) answer as function of
  - ability level and
  - item parameters

  This function is called:
  - *Item response function (IRF)*
  - *Item characteristic curve (ICC)*

Introduction to IRT models

Aim of IRT models:

- To calibrate items (estimate difficulty, discrimination, guessing,...)
- To assess respondents' latent trait (ability, satisfaction, anxiety,...)
- To describe test properties (standard error, test information,...)

Other applications of IRT models:

- Test linking and equating
- Differential item functioning
- Computerized adaptive testing
- etc.

Common Assumptions of IRT models

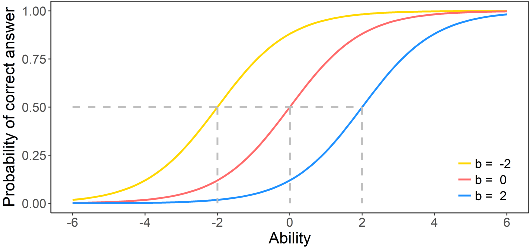
1. Unidimensionality of latent variable
   - all items measure only one construct
   - can be tested
   - examples when violated?
2. Local independence
   - also called conditional independence
   - given latent ability, the responses to items are independent
   - examples when violated?
3. Monotonicity
   - the ICC is monotonically increasing or decreasing with the ability level
4. Invariance of parameters
   - Estimates of item parameters are the same over samples of examinees
   - Estimates of ability parameters are the same over samples of items
   - examples when violated?
5. Independence of respondents
   - examples when violated?

Rasch Model

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

θ_i ability of person i
b_j difficulty of item j (location of inflection point)

Item Characteristic Curve (ICC)
also called Item Response Function (IRF)

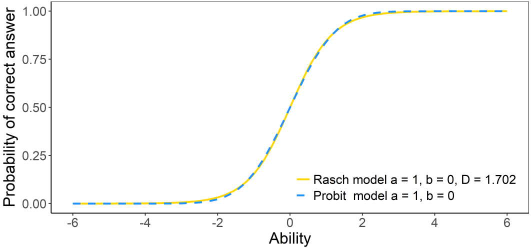


Logistic vs. Probit model (Note on Scaling parameter D)

Rasch model is sometimes defined as:

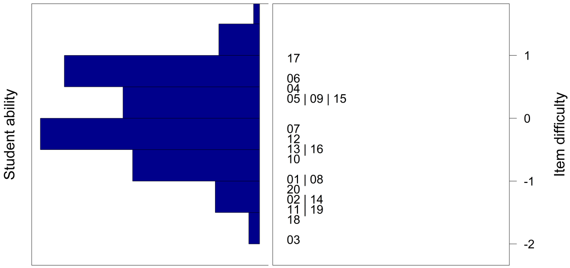
$$\pi_{ij} = P(Y_{ij} = 1|\theta_i, b_j) = \frac{\exp(D[\theta_i - b_j])}{1 + \exp(D[\theta_i - b_j])}$$

$D = 1.702$ is scaling parameter introduced in order to match logistic and probit metrics very closely (Lord and Novick, 1968)



Item-Person Map (Wright Map)

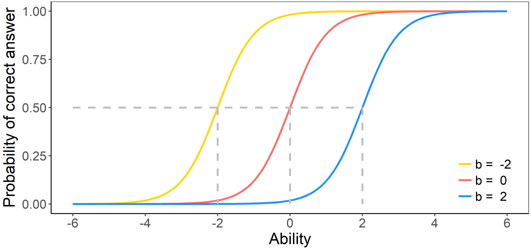
IRT models allow us to put *items* and *persons* on the same scale



1PL IRT Model

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

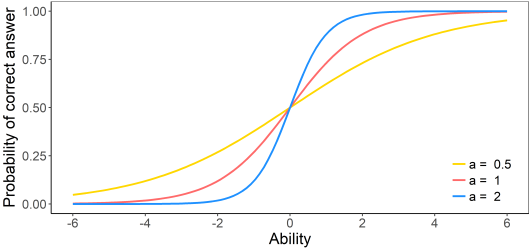
θ_i ability of person i
b_j difficulty of item j (location of inflection point)
a discrimination common for all items (slope at inflection point)



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
b_j difficulty of item j (location of inflection point)
a_j discrimination of item j (slope at inflection point)



3PL IRT Model

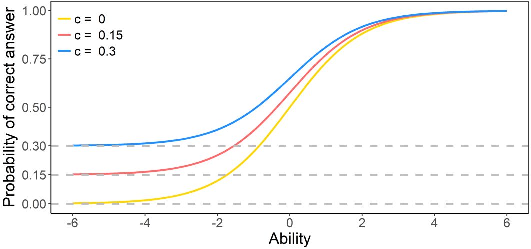
$$\pi_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]}$$

θ_i ability of person i

b_j difficulty of item j (location of inflection point)

a_j discrimination of item j (slope at inflection point)

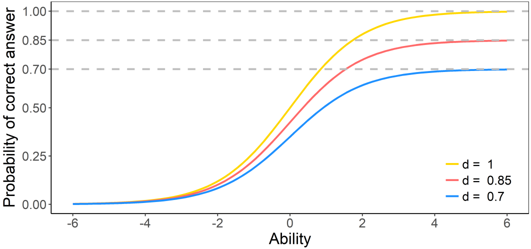
c_j pseudo-guessing parameter of item j (lower/left asymptote)



4PL IRT Model

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

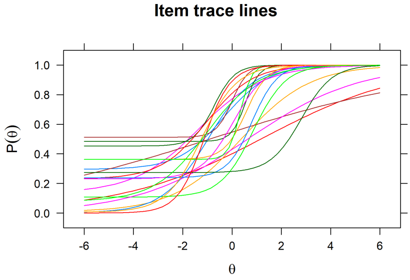
θ_i ability of person i
b_j difficulty of item j (location of inflection point)
a_j discrimination of item j (slope at inflection point)
c_j pseudo-guessing parameter of item j (lower/left asymptote)
d_j innatention parameter of item j (upper/right asymptote)

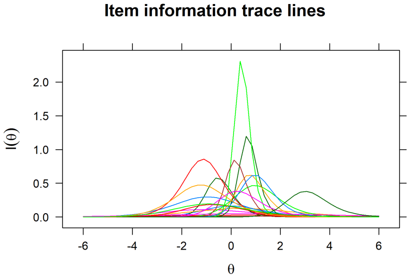


Information Function

$$P(\theta, a_j, b_j, c_j, d_j) = c_j + (d_j - c_j)\frac{\exp[a_j(\theta - b_j)]}{1 + \exp[a_j(\theta - b_j)]},$$

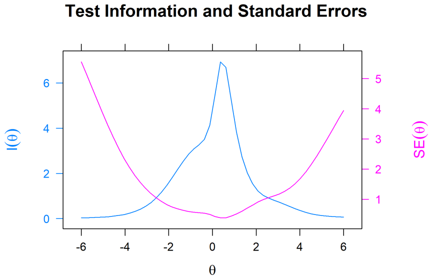
$$\mathrm{I}_j(\theta, a_j, b_j, c_j, d_j) = \frac{\delta P}{\delta \theta} = a_j(d_j - c_j)\frac{\exp[a_j(\theta - b_j)]}{\{1 + \exp[a_j(\theta - b_j)]\}^2}$$

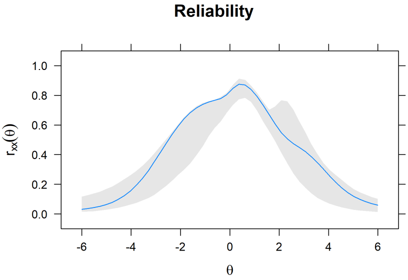




Test Information and Reliability

$$\mathrm{I}(\theta) = \sum_j \mathrm{I}_j(\theta, a_j, b_j, c_j, d_j)$$





$$SEM = \sigma\sqrt{(1 - r_{xx})}$$

Further Topics

Further issues

- Estimation of item parameters
- Estimation of student abilities
- Item and Person Fit Assessment, etc.

Further models

- Polytomous IRT models (ordinal/nominal)
- Multidimensional IRT models
- Hierarchical IRT models, etc.
- Accounting for Differential item functioning, etc.

Applications

- Test equating
- Computerized adaptive testing, etc.

Vocabulary

- Item Characteristic Curve (ICC)
- Item Response Fuction (IRF)
- Item Information Function (IIF)
- Test Information Function (TIF)
- Likelihood function
- Rasch model, 1PL, 2PL, 3PL, 4PL IRT models