Universität Innsbruck

Mixtures of Rasch Models

Hannah Frick, Friedrich Leisch, Achim Zeileis, Carolin Strobl

http://www.uibk.ac.at/statistics/

Introduction

- Rasch model for measuring latent traits
- Model assumption: Item parameters estimates do not depend on person sample
- Violated in case of differential item functioning (DIF)
- Several approaches to test for DIF:
  - LR tests, Wald tests
  - Rasch trees
  - Mixture models
- Here: Two versions of the mixture model approach

Rasch Model

Probability for person i to solve item j:

$$P(Y_{ij} = y_{ij} | \theta_i, \beta_j) = \frac{e^{y_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}}$$

- y_{ij}: Response by person i to item j
- θ_i: Ability of person i
- β_j: Difficulty of item j

ML Estimation

Factorization of the full likelihood on basis of the scores $r_i = \sum_{j=1}^{m} y_{ij}$

$$
\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\beta}) &= f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) \\
&= h(\mathbf{y}|\boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\beta}) g(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\beta}) \\
&= h(\mathbf{y}|\boldsymbol{r}, \boldsymbol{\beta}) g(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\beta})
\end{aligned}
$$

- Joint ML: Joint estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is inconsistent
- Marginal ML: Assume distribution for $\boldsymbol{\theta}$ and integrate out in $g(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\beta})$
- Conditional ML: Assume $g(\boldsymbol{r}) = g(\boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\beta})$ as given or that it does not depend on $\boldsymbol{\theta}, \boldsymbol{\beta}$ (but potentially other parameters). Hence, $g(\boldsymbol{r})$ is a nuisance term and only $h(\boldsymbol{y}|\boldsymbol{r}, \boldsymbol{\beta})$ needs to be maximized.

Mixture Models

- Mixture models are a tool to model data with unobserved heterogeneity caused by, e.g., (latent) groups
- Mixture density = $\sum$ weight $\times$ component
- Weights are a priori probabilities for the components
- Components are densities or (regression) models

Mixtures of Rasch Models

- Mixture of the full likelihoods by Rost (1990):

$$f(\boldsymbol{y}|\boldsymbol{\pi}, \boldsymbol{\psi}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \psi_{r_i,k} h(\mathbf{y}_i | r_i, \boldsymbol{\beta}_k)$$

with $\psi_{r_i,k} = g_k(r_i)$

Mixtures of Rasch Models

- Mixture of the full likelihoods by Rost (1990):

$$f(\boldsymbol{y}|\boldsymbol{\pi}, \boldsymbol{\psi}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \psi_{r_i,k} h(\mathbf{y}_i | r_i, \boldsymbol{\beta}_k)$$

with $\psi_{r_i,k} = g_k(r_i)$

- Mixture of the conditional likelihoods:

$$f(\boldsymbol{y}|\boldsymbol{\pi}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k h(\mathbf{y}_i | r_i, \boldsymbol{\beta}_k)$$

Parameter Estimation

EM algorithm by Dempster, Laird and Rubin (1977)

- Group membership is seen as a missing value
- Optimization is done iteratively by alternate estimation of group membership (E-step) and component densities (M-step)
- E-step:

$$\hat{p}_{ik} = \frac{\hat{\pi}_k h(\boldsymbol{y}_i | r_i, \hat{\boldsymbol{\beta}}_k)}{\sum_{g=1}^{K} \hat{\pi}_g h(\boldsymbol{y}_i | r_i, \hat{\boldsymbol{\beta}}_g)}$$

- M-step:
  For each component separately

$$\hat{\boldsymbol{\beta}}_k = \underset{\beta_k}{\text{argmax}} \sum_{i=1}^{n} \hat{p}_{ik} \log h(\boldsymbol{y}_i | r_i, \hat{\boldsymbol{\beta}}_k)$$

Number of Components

How can the number of components k be established?

- A priori known number of groups in the data
- LR test: Regularity conditions are not fulfilled

  $\rightarrow$ Distribution under H_0 unknown

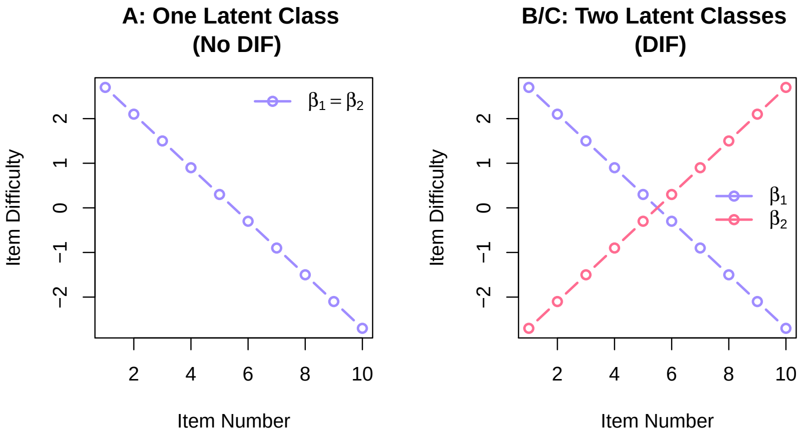
  $\rightarrow$ Bootstrap necessary
- Information criteria: AIC, BIC, ICL

Simulation Design

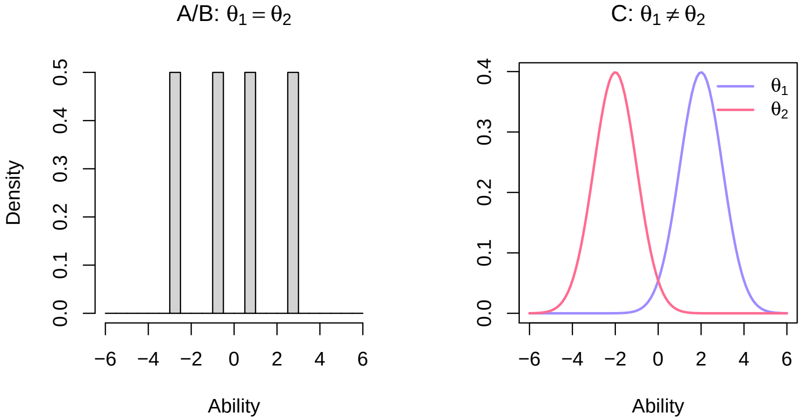
- 10 items, 1800 people, equal group sizes
- Latent groups in item and/or person parameters:

	$\beta_1 = \beta_2$	$\beta_1 \neq \beta_2$
$\theta_1 = \theta_2$	A	B
$\theta_1 \neq \theta_2$		C

Item Parameters



Person Parameters



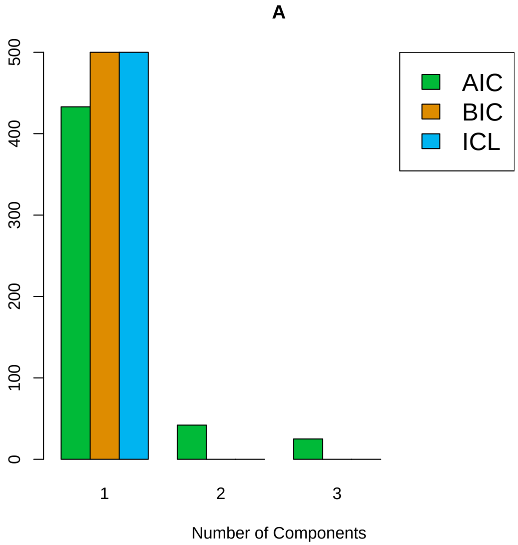
Criteria for Goodness of Fit

- Number of components
- Rand index:

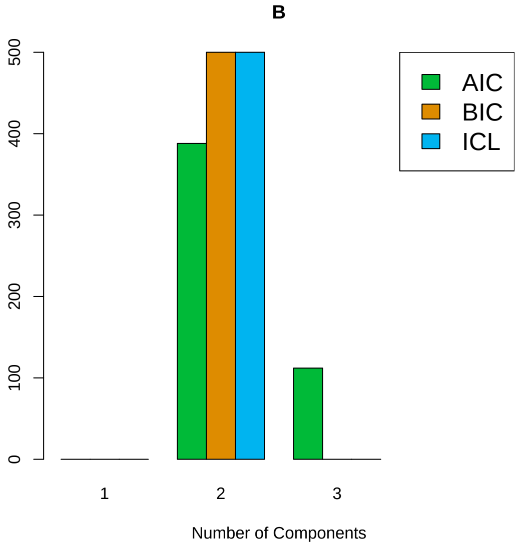
  Agreement between true and estimated partition
- Mean residual sum of squares:

  Agreement between true and estimated (item) parameter vector

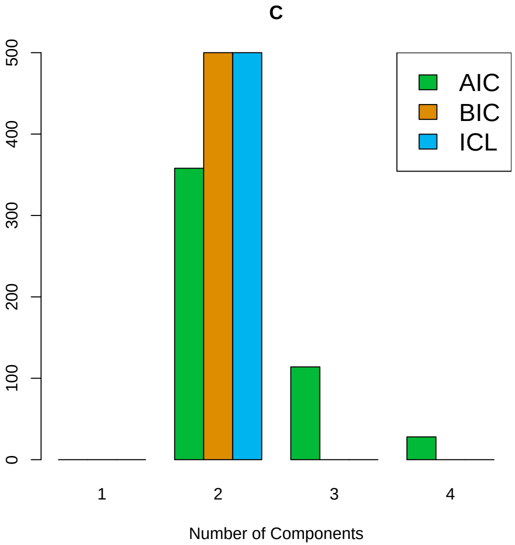
No Latent Classes (No DIF)



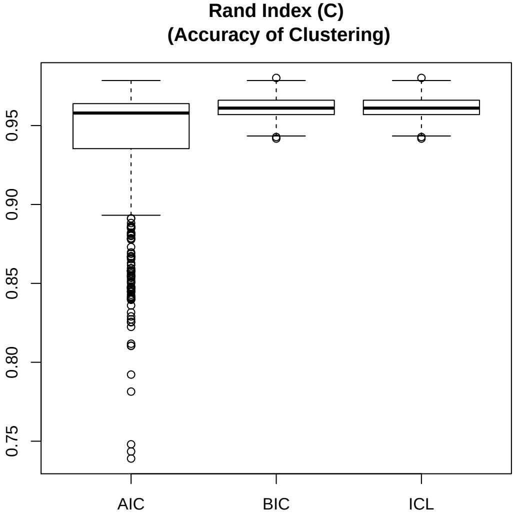
Two Latent Classes (DIF)



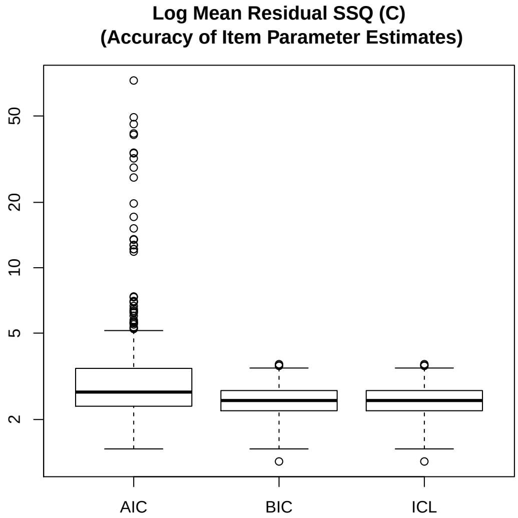
Latent Structure in Item and Person Parameters (DIF + Ability Differences)



Latent Structure in Item and Person Parameters (DIF + Ability Differences)



Latent Structure in Item and Person Parameters (DIF + Ability Differences)



Summary and Outlook

- Model suitable for detecting latent classes with DIF
- Model also suitable when a latent structure in the person parameters is present
- AIC tends to overestimate the correct number of classes, BIC and ICL work well
- Clustering of the observations works well
- Estimation of the item parameters in the components works reasonably well
- Comparison with Rost’s MRM to follow

Literature

- Arthur Dempster, Nan Laird, and Donald Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B*, 39(1): 1–38, 1977.
- Bettina Grün and Friedrich Leisch. Flexmix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. *Journal of Statistical Software*, 28(4): 1–35, 2008.
- Georg Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests.* The University of Chicago Press, 1960.
- Jürgen Rost. Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis. *Applied Psychological Measurement*, 14(3): 271–282, 1990.
- Carolin Strobl. *Das Rasch-Modell - Eine verständliche Einführung für Studium und Praxis*. Rainer Hampp Verlag, 2010.