Mixtures of Rasch Models with R Package psychomix

Hannah Frick, Carolin Strobl, Friedrich Leisch, Achim Zeileis

http://eeecon.uibk.ac.at/~frick/

Outline

- Rasch model
- Mixture models
- Rasch mixture models
- Illustration: Simulated data
- Application: Verbal aggression

Introduction

- Latent traits measured through probabilistic models for item response data.
- Here, Rasch model for binary items.
- Crucial assumption of measurement invariance: All items measure the latent trait in the same way for all subjects.
- Check for heterogeneity in (groups of) subjects, either based on observed covariates or unobserved latent classes.
- Mixtures of Rasch models to address heterogeneity in latent classes.

Rasch Model

Probability for person *i* to solve item *j*:

$$P(Y_{ij} = y_{ij} | \theta_i, \beta_j) \; = \; \frac{\exp\{y_{ij}(\theta_i - \beta_j)\}}{1 + \exp\{\theta_i - \beta_j\}}.$$

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

By construction:

- No covariates, all information is captured by ability and difficulty.
- Both parameters θ and β are on the same scale: If $\beta_1 > \beta_2$, then item 1 is more difficult than item 2 for *all* subjects.

Central assumption of measurement invariance needs to be checked for both manifest and latent subject groups.

Rasch Model: Estimation

- Joint estimation of θ and β is inconsistent.
- Conditional ML estimation: Use factorization of the full likelihood on basis of the scores $r_i = \sum_{j=1}^{m} y_{ij}$:

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

  Estimate β from maximization of $h(y|r, \beta)$.
- Also maximizes $L(\theta, \beta)$ if $g(r|\cdot)$ is assumed to be independent of θ and β; but potentially depending on auxiliary parameters δ: $g(r|\delta)$.

Mixture Models

- Assumption: Data stems from different classes but class membership is unknown.
- Modeling tool: Mixture models.
- Mixture model = $\sum$ weight $\times$ component.
- Components represent the latent classes. They are densities or (regression) models.
- Weights are a priori probabilities for the components/classes, treated either as parameters or modeled through concomitant variables.

Rasch Mixture Models: Framework

Full mixture:

- Weights: Either (non-parametric) prior probabilities π_k or weights $\pi(k|x, \alpha)$ based on concomitant variables x, e.g., a multinomial logit model.
- Components: Conditional likelihood for item parameters and specification of score probabilities

$$f(y|\pi, \alpha, \beta, \delta) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi(k|x_i, \alpha)\ h(y_i|r_i, \beta_k)\ g(r_i|\delta_k).$$

- Estimation of all parameters via ML through the EM algorithm.

Rasch Mixture Models: Score Probabilities

- Original proposition by Rost (1990): Discrete distribution with parameters (probabilities) $g(r) = \Psi_r$.
- Number of parameters necessary is potentially very high: (number of items - 1) $\times$ (number of components).
- More parsimonious: Assume parametric model on score probabilities, e.g., using mean and variance parameters.
- General approach: Conditional logit model encompassing the original saturated parameterization and a mean/variance parameterization (with only two parameters per component) as special cases

$$g(r|\delta) \; = \; \frac{\exp\{z_r^\top \delta\}}{\sum_{j=1}^{m-1} \exp\{z_j^\top \delta\}}.$$

Rasch Mixture Models: Score Probabilities

Saturated score model:

- Pro: Can capture all score distributions, i.e., never misspecified.
- Con: Needs many (nuisance) parameters, i.e., challenging in model estimation/selection.

Mean-variance score model:

- Pro: Parsimonious, i.e., convenient for model estimation/selection.
- Con: Potentially misspecified, e.g., for multi-modal distributions.

Illustration: Moderate DIF

Data generating process:

- $m = 20$ items, $n = 500$ subjects.
- 2 items with DIF: $\beta_5 = \pm 1.7$ and $\beta_{15} = \mp 1.7$ in groups 1 and 2, all other items with $\beta_j = 0$.
- All abilities $\theta_i = 0$.

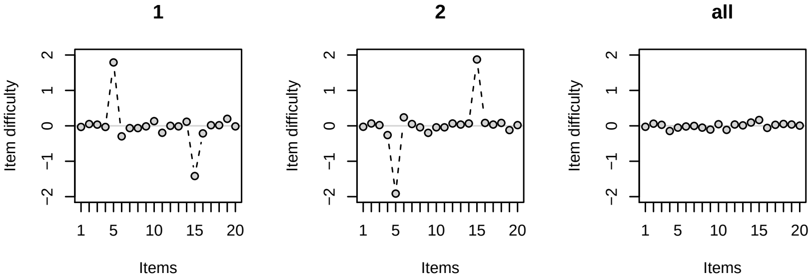


Illustration: Moderate DIF

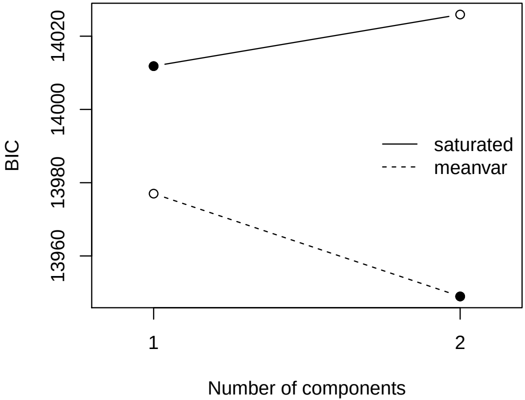


Illustration: No DIF

Data generating process:

- $m = 20$ items, $n = 100$ subjects.
- No DIF: all item difficulties $\beta = 0$.
- Differences in scores: $\theta_i = 1.5$ in group 1, $\theta_i = -1.5$ in group 2.

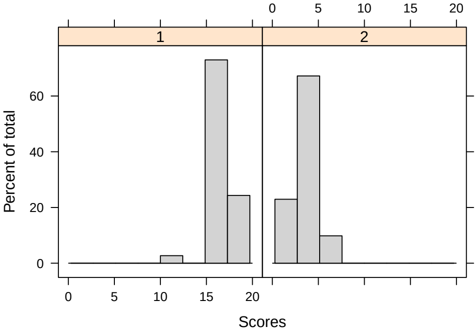


Illustration: No DIF

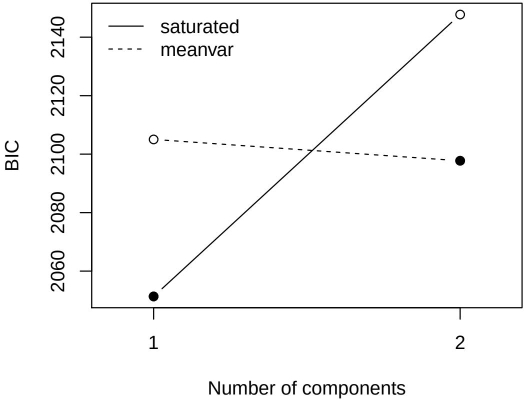


Illustration: No DIF

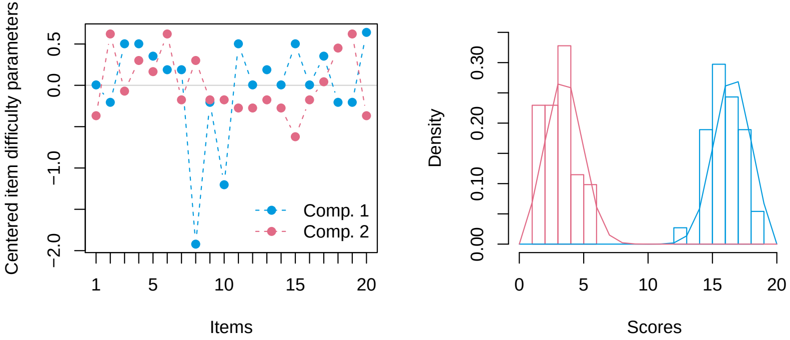


Figure: Estimated item parameters (left) and score probabilities with empirical score distribution (right) of the 2-class Rasch mixture model with a meanvar score specification.

Rasch Mixture Models: Score Probabilities

Goal: Robust specification.

Motivation: When checking for measurement invariance, items are of interest, not the scores.

Idea: Use

$$g(r) = constant$$

Equivalent to: Score distribution is the same over all components.

Interpretation: Score distribution can be multi-modal (if the same over all components).

Open question: What happens if the assumption of equivalence across components is violated? Currently under investigation.

Software

- Available in R in package **psychomix** at `http://CRAN.R-project.org/package=psychomix`
- Based on package **flexmix** (Grün and Leisch, 2008) for flexible estimation of mixture models.
- Based on package **psychotools** for estimation of Rasch models.
- Frick et al. (2012), provides implementation details and hands-on practical guidance. See also `vignette("raschmix", package = "psychomix")`.

Application: Verbal Aggression Data

- Behavioral study of psychology students: 243 women and 73 men.
- Description of frustrating situations:
  - S1: A bus fails to stop for me.
  - S2: I miss a train because a clerk gave me faulty information.
- Behavioral mode: Want or do.
- Verbally aggressive response: Curse, scold, or shout.
- 12 resulting items: S1WantCurse, S1DoCurse, S1WantScold, . . . , S2WantShout, S2DoShout
- Covariates: Gender and an anger score.

Verbal Aggression: Analysis

Fit model:

```
R> set.seed(1)
R> mix <- raschmix(resp2 ~ 1, data = va12, k = 1:4,
+    scores = "meanvar", nrep = 5)
R> mixC <- raschmix(resp2 ~ gender + anger, data = va12,
+    k = 1:4, scores = "meanvar", nrep = 5)
```

Select model:

```
R> rbind(mix = BIC(mix), mixC = BIC(mixC))

            1        2        3        4
mix  3874.632 3857.549 3854.367 3887.003
mixC 3874.632 3863.068 3854.820 3880.484

R> va12_mix <- getModel(mixC, which = "3")
```

Plot item profiles and effects of concomitant variables:

```
R> xyplot(va12_mix)
R> effectsplot(va12_mix)
```

Verbal Aggression: Item Profiles

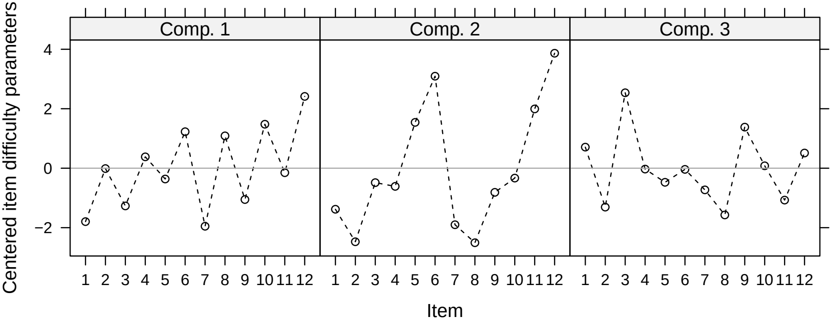
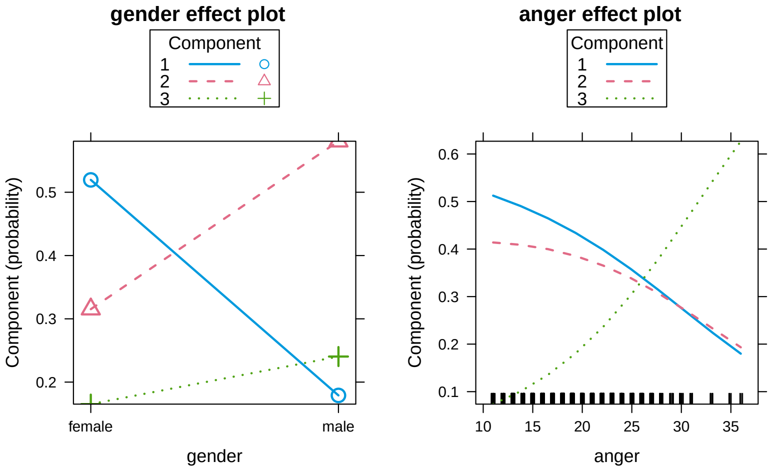


Figure: Item difficulty profiles for the 3-component Rasch mixture model. Items 1–6: Situation S1 (bus). Items 7–12: Situation S2 (train). Order: want/do curse, want/do scold, want/do shout.

Verbal Aggression: Effects Displays



Verbal Aggression: Summary

- Number of components: 3 different sets of item parameters necessary.
- Relationship between items differs between the latent classes.
- For shouting: Want is less extreme than do. For cursing and scolding, this depends on the latent class.
- One class does not differentiate much between the items, for the two other classes, cursing/scolding/shouting is increasingly extreme.
- Some dependence on covariates gender and anger score (albeit slightly poorer BIC).

Summary

- Mixture Rasch models are a flexible means to check for measurement invariance.
- General framework incorporates concomitant variable models for mixture weights along with various score models.
- Pros and cons of the different score distributions require further investigation.
- Implementation of all flavors in R package **psychomix**.

References

Frick H, Strobl C, Leisch F, Zeileis A (2012). “Flexible Rasch Mixture Models with Package psychomix.” *Journal of Statistical Software*, 48(7), 1–25. `http://www.jstatsoft.org/v48/i07/`

Fischer GH, Molenaar IW (eds.) (1995). *Rasch Models: Foundations, Recent Developments, and Applications*. Springer-Verlag, New York.

Grün B, Leisch F (2008). "FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters." *Journal of Statistical Software*, 28(4), 1–35. `http://www.jstatsoft.org/v28/i04/`

Rost J (1990). “Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis.” *Applied Psychological Measurement*, 14(3), 271–282.