Longitudinal multiple-group IRT modelling: covariance pattern selection using MCMC and RJMCMC

Caio L. N. Azevedo, IMECC-Unicamp

and Jean-Paul Fox, OMD-University of Twente, Holanda

joint work with Dalton F. de Andrade, INE-UFSC

This work was financially supported by CNPq

IV CONBRATRI, December 2015

Paper

This presentation corresponds to a part of the paper entitled: **Azevedo, C. L. N. ; Fox, J.-P. ; ANDRADE, D. F. (2015). Longitudinal Multiple-Group IRT Modeling: Covariance pattern selection using MCMC and RJMCMC. International Journal of Quantitative Research in Education, speciall issue on " Bayesian Statistics in Psychometrics", 2, 3-4, p. 213-243.**

Outline of the presentation (AT THE FINAL)

- Motivation.
- Literature review.
- Main contributions of this work.
- A longitudinal IRT model with restricted covariance matrices and time-heterogenous variances.
- Simulation studies.
- Real data anaylsis.
- Final conclusions.

Motivation

- Longitudinal item response data occur when examinees are assessed at several time points.
- This kind of data consist of response patterns of different examinees responding to different tests, or instrument measurement, (in general with a structure of common items) at different measurement occasions (e.g. grades).
- This leads to a complex dependence structure that arises from the fact that measurements (i.e., the latent traits and item reponses) from the same student are typically correlated.

Motivation

- Multiple group item response data occur when examinees can be (or they are naturally) clustered according some characteristic(s) of interest as gender, grade, social level and so on.
- In general, the examinees (beloging to different groups) are submitted to different tests (or measurement instrument) with a structure of common items.
- The group heterogeneity can reflect different behaviors. Therefore, it is important to take such heterogeneity into account.

Motivation

- Also, in longitudinal IRT data, different behaviours (heterogeneity) are observed in the group among the time-points (the group changes over the time).
- When these two structures are combined we have the longitudinal multiple group data.
- Longitudinal studies in psychometric assessment are often focused on latent traits of subjects, who are clustered in different groups.

Main goals

- The original goals were:
  - To developed a longitudinal multiple group model and Bayesian tools for paramter estimation and model fit assessment (achieved).
  - To developed a RJMCMC algorithm for the longitudinal multiple group model for covariance matrix selection (achieved only for one group IRT longitudinal model and two covariance matrices). As an alternative to the methods considered by Azevedo et al 2014 and Tavares 2001.
  - Similarly to the multiple group model (MGM) we expected that the equating perfomed by the proposed model is more appropriate than the using of the posterior equating methods.

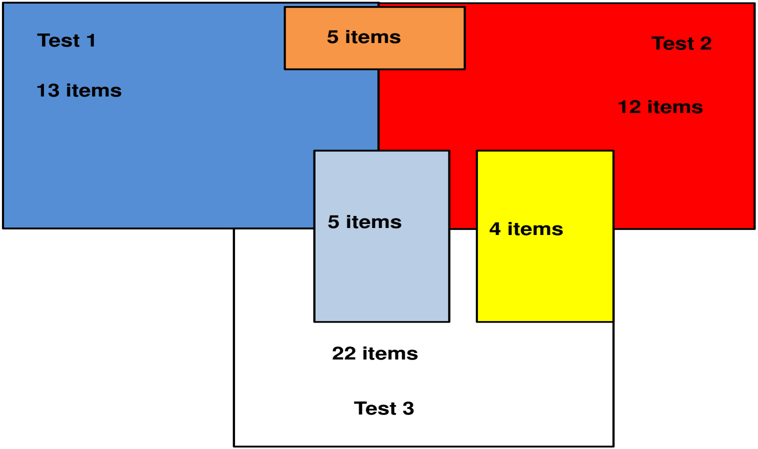
First real data set - longitudinal one group data

- The data set analyzed stems from a major study initiated by the Brazilian Federal Government known as the School Development Program.
- The aim of the program is to improve the teaching quality and the general structure (classrooms, libraries, laboratory informatics etc) in Brazilian public schools.
- A total of 400 schools in different Brazilian states joined the program. Achievements in mathematics and Portuguese language were measured over five years (from fourth to eight grade of primary school) from students of schools selected and not selected for the program.

First real data set - longitudinal single-group data

- In the present study, mathematic performances of 1,500 randomly selected students, who were assessed in the fourth, fifth, and sixth grade, were considered.
- A total of 72 test items was used, where 23, 26, and 31 items were used in the test in grade four (Test 1), grade five (Test 2), and grade six (Test 3), respectively. Five anchor items were used in all three tests.
- Another common set of five items was used in the test in grade four and five. Furthermore, four common items were used in the tests in grades five and six.

Test design - first data set



First real data set - longitudinal single-group data

- In an exploratory analysis, the Multiple Group Model (MGM), described in Azevedo et al. (2012), was used to estimate the latent student achievements given the response data.
- The MGM for cross-sectional data assumes that students are nested in groups and latent traits are assumed to be independent given the mean level of the group.
- Pearson's correlations, variances, and covariances were estimated among the vectors of estimated latent traits corresponding to grade four to six. The estimates are represented in the next slide.

Within-student correlation structure of the latent traits estimated by the MGM

	Grade four	**Grade five**	**Grade six**
Grade four	1.000	**.723**	**.629**
Grade five	-	1.152	**.681**
Grade six	-	-	1.071

Bayesian estimates (posterior expectation) of the variances and and correlations - variances (obtained directly from the MGM) in the main diagonal and correlations (estimated using the estimated latent traits) in the upper triangle, respectively.

Second real data set - longitudinal two group data

- Data were analysed from the AGHLS, which is a multidisciplinary longitudinal cohort study, originally set up to examine growth and health among teenagers.
- The AGHLS is focused on research questions related to relationships between anthropometry, physical activity, cardiovascular disease risk, lifestyle, musculoskeletal health, psychological health and wellbeing.
- The presented sample consists of 443 participants who were followed over the period 19932006 with a maximum of three measurement points for each individual. A subscale of the STAI-DY questionnaire was used to measure the latent variable state anxiety, using a total of thirteen items.

Second real data set - longitudinal two group data

- Data from three years were used in the analysis; 1993, 2000 and 2006, referred to, respectively, as years 1, 2 and 3.
- Two groups were considered, male students (group 1) and female students (group 2). Therefore, two groups were assessed at three occasions (similar to the design of the first simulation study).
- A total of 59 male students and 72 female students were assessed at each measurement occasion, providing 393 responses per item.

Second real data set - longitudinal two group data

- In an explanatory analysis, the MGM (Azevedo et al., 2012), was used to estimate the latent traits given the response data according to a cross-sectional design.
- A total of six groups were considered (the male and the female groups measured at each of the three occasions). The MGM for cross-sectional data assumed that students were nested in groups and latent traits were assumed to be independently distributed over occasions given the mean level of the group.
- Subsequently, Pearsons correlations were estimated for the pairs of estimated latent traits corresponding to years one to three.

Within-student correlation structure of the latent traits estimated by the MGM

	Male			Female		
	Year 1	**Year 2**	**Year 3**	**Year 1**	**Year 2**	**Year 3**
Year 1	1.000	**0.616**	**0.499**	0.832	**0.596**	**0.635**
Year 2	-	0.783	**0.709**	-	1.073	**0.748**
Year 3	-	-	0.846	-	-	0.879

Estimated posterior variances, covariances, and correlations among estimated latent traits are given in the diagonal, lower and upper triangle, respectively.

Literature review

- Conoway (1990): longitudinal Rasch model (one-parameter model):
  - Uniform covariance matrix.
  - Complete test design.
- Dunson (2003): a general modeling framework that allows mixtures of count, categorical and continuous response:
  - Complete test design.
  - Did not explore sepecific covariance structures.

Literature review

- Andrade and Tavares (2005): longitudinal three-parameter model (population latent traits parameters estimation.
- Tavares and Andrade (2006): longitudinal three-parameter model (item and population latent traits parameters estimation):
  - Three-parameter model. time-homogeneous covariance matrices.
  - Numerical problems in handling many time-points.
  - All covariance matrices have closed expressions for their inverses and determinants.

Literature review

- Tavares (2001) developed a longitudinal multipel group model under a frequentist framework but no numerical results were provided (same issues presented in previuos slide).
- Usual statistics of model comparison (for covariance matrix selection) were considered by Tavares (2001) (frequentist, AIC, BIC) and Azevedo et al (2014) (Bayesian, E(AIC), E(BIC), E(DIC)). While in the former no simulation studies were performed, the results provided by the latter indicates that these statistics did not work well.

Main contributions of this work

- General modeling of the time-heterogenous dependency between student achievements (some restricted covariance matrices never used for longitudinal IRT data) and between-grupos heterogeneity in a single modeling framework to analyze longitudinal multiple group IRT data.
- Bayesian inference which handles identification rules, restricted parametric covariance structures and situations with several time-points (there is no limit, numerically speaking).
- Simulation study: parameter recovery and model selection.

Main contributions of this work

- Bayesian model fit assessment and model comparison tools.
- All developments were made considering the two-parameter model but can be straightforwardly extended to the three-parameter, polytomous or continuous IRT models.
- A reversible-jump MCMC (RJMCMC) algorithm for joint Bayesian parameter estimation and covariance matrix selection for a single-group longitudinal IRT model.
- A longitudinal multiple group model to handle the scaling process simultaneously with the estimation of latent traits, item and population parameters.

Framework

- One or more tests are administered to subjects clustered into different groups, which are followed along several time-points.
- The subjects are randomly selected at the first time-point from their respective groups.
- At each time-point t, $t = 1, ..., T$, a test of $\mathcal{I}_{kt}$ items, from a total of $I \leq \sum_{k=1}^{K} \sum_{t=1}^{T} \mathcal{I}_{kt}$ items, is administered to each group k $(k = 1, ..., K)$ of n_{kt} subjects.
- Here, a complete data case is assumed, $n_{kt} = n_k, \forall k$. Across measurement occasions common items are used, which defines an incomplete test design.

A longitudinal multiple-group IRT Model

$$
\begin{aligned}
Y_{ijtk} \mid (\theta_{jkt}, \zeta_i) &\sim \text{Bernoulli}(P_{ijkt}) \\
P_{ijkt} = P(Y_{ijkt} = 1 \mid \theta_{jkt}, \zeta_i) &= \Phi(a_i \theta_{jkt} - b_i) \\
\boldsymbol{\theta}_{jk.} | \boldsymbol{\eta}_{\boldsymbol{\theta}_k} &\overset{i.i.d.}{\sim} N_T(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}),
\end{aligned}
$$

- Y_{ijkt} : is the answer of subject (student) j, of group k, to item i, in time-point t. It is equal to 1 if the subject answers the item correctly and 0 otherwise.
- $\boldsymbol{\theta}_{jk.} = (\theta_{jk1}, ..., \theta_{jkT})'$: is the vector of the latent traits of the subject j of group k.
- θ_{jkt}: is the latent trait of the subject j, of group k in time-point t.

Population (latent traits) parameters

$$\boldsymbol{\mu}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \mu_{\theta_{k1}} \\ \mu_{\theta_{k2}} \\ \vdots \\ \mu_{\theta_{kT}} \end{bmatrix}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \psi_{\theta_{k12}} & \dots & \psi_{\theta_{k1T}} \\ \psi_{\theta_{k12}} & \psi_{\theta_{k2}} & \dots & \psi_{\theta_{k2T}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{k1T}} & \psi_{\theta_{k2T}} & \dots & \psi_{\theta_{kT}} \end{bmatrix}, \quad (1)$$

$\boldsymbol{\eta}_{\boldsymbol{\theta}_k}$ is a vector consisting of $\boldsymbol{\mu}_{\boldsymbol{\theta}_k}$ and the non-repeated elements of $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}$.

Modeling the covariance matrix

Heteroscedastic uniform model - HU

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_2}}\rho_\theta & \ldots & \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_T}}\rho_\theta \\ \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_2}}\rho_\theta & \psi_{\theta_2} & \ldots & \sqrt{\psi_{\theta_2}}\sqrt{\psi_{\theta_T}}\rho_\theta \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_T}}\rho_\theta & \sqrt{\psi_{\theta_2}}\sqrt{\psi_{\theta_T}}\rho_\theta & \ldots & \psi_{\theta_T} \end{bmatrix},$$

Heteroscedastic Toeplitz model - HT

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_2}}\rho_\theta & 0 & \ldots & 0 \\ \sqrt{\psi_{\theta_1}}\sqrt{\psi_{\theta_2}}\rho_\theta & \psi_{\theta_2} & \sqrt{\psi_{\theta_2}}\sqrt{\psi_{\theta_3}}\rho_\theta & \ldots & 0 \\ 0 & \sqrt{\psi_{\theta_2}}\sqrt{\psi_{\theta_3}}\rho_\theta & \psi_{\theta_3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \psi_{\theta_T} \end{bmatrix}.$$

Modeling the covariance matrix

Heteroscedastic covariance model - HC

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \rho_\theta & \dots & \rho_\theta \\ \rho_\theta & \psi_{\theta_2} & \dots & \rho_\theta \\ \vdots & \vdots & \ddots & \vdots \\ \rho_\theta & \rho_\theta & \dots & \psi_{\theta_T} \end{bmatrix},$$

First-order autoregressive moving-average model - ARMAH

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\gamma_\theta & \dots & \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\gamma_\theta\rho_\theta^{T-2} \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\gamma_\theta & \psi_{\theta_2} & \dots & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\gamma_\theta\rho_\theta^{T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\gamma_\theta\rho_\theta^{T-2} & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\gamma_\theta\rho_\theta^{T-3} & \dots & \psi_{\theta_T} \end{bmatrix}.$$

Modeling the covariance matrix

Ante-dependence model - AD

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta_1} & \cdots & \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\prod_{t=1}^{T-1}\rho_{\theta_t} \\ \sqrt{\psi_{\theta_1}\psi_{\theta_2}}\rho_{\theta_1} & \psi_{\theta_2} & \cdots & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\prod_{t=2}^{T-1}\rho_{\theta_t} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\psi_{\theta_1}\psi_{\theta_T}}\prod_{t=1}^{T-1}\rho_{\theta_t} & \sqrt{\psi_{\theta_2}\psi_{\theta_T}}\prod_{t=2}^{T-1}\rho_{\theta_t} & \cdots & \psi_{\theta_T} \end{bmatrix},$$

Development of the MCMC and RJMCMC algorithms

- The model identification problem can be handled by fixing the mean and the variance of the latent trait distribution (of the first time-point, for example) in the longitudinal single-group IRT model. Similarly, for the longitudinal multiple-group model it suffices to fix the mean and the variance of the latent trait distribution (of the first group at the first time-point, for example).
- An agumented data scheme as in Albert (1992) was considered.
- All technical details can be foun in Azevedo et al (2015).

Convergence and autocorrelation assessment and parameter recovery

- The Geweke diagnostic, based on a burn-in period of 16,000 iterations, indicated convergence of the chains of all model parameters.
- Furthermore, the Gelman-Rubin diagnostic were close to one, for all parameters. Convergence was established easily without requiring informative initial parameter values or long burn-in periods.
- Therefore, the burn-in was set to be 16,000, and a total of 46,000 values were simulated, and samples were collected at a spacing of 30 iterations producing a valid sample with 1,000 values.

Parameter recovery and model selection assessment

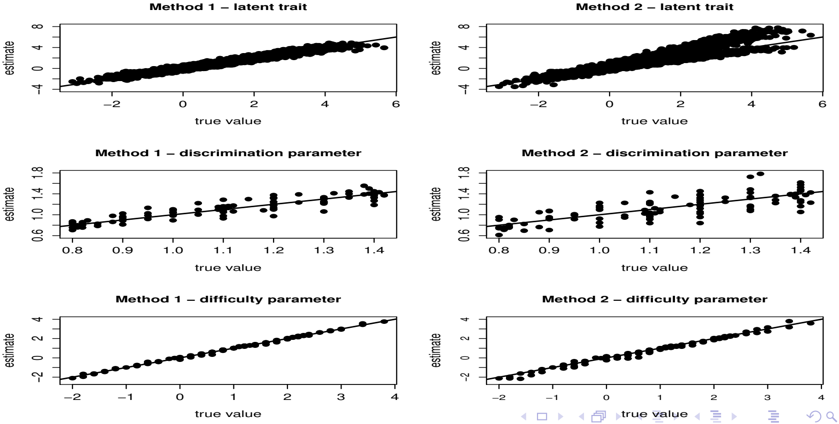
- The results of simulation studies indicated :
  - That all parameters were properly recovered by the MCMC and RJMCMC algorithms.
  - The RJMCMC algorithm identifies the true underlying covariance matrix with great accuracy (100% of the times).
  - For the longitudinal multiple-group data the item parameters and the latent traits are more accurately estimated when we perform the scaling process simultaneously with the estimation of the parameters (method 1) compared with the posterior equating (method 2).

Simulation study: covariance selection for the longitudinal single group study

Longitudinal single group study: averaged proportion of visits for each model

Scenario	True model	Model	
		ARH	ARMAH
1	ARH	.977	.023
2	ARH	.980	.020
3	ARMAH	.261	.739

Simulation study: Implicit scaling (LMGIRT model) compared to posterior equating (LIRT model)



Simulation study: Implicit scaling (LMGIRT model) compared to posterior equating (LIRT model)

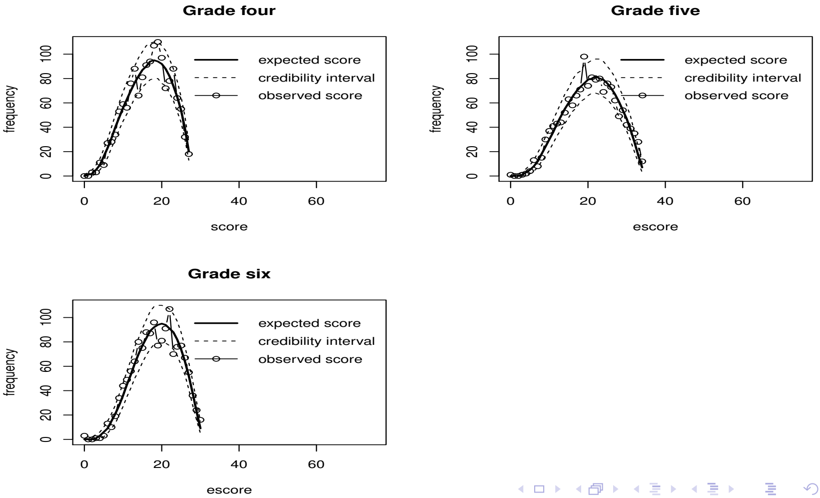
Parameter	Abias		Rabias	
	Method 1	**Method 2**	**Method 1**	**Method 2**
discrimination parameter	6.62	12.10	5.97	10.85
difficulty parameter	5.50	13.38	42.14	75.66
latent trait	1289.87	3982.28	5605.28	10695.73

Abias: absolute bias ; Rabias: relative absolute bias.

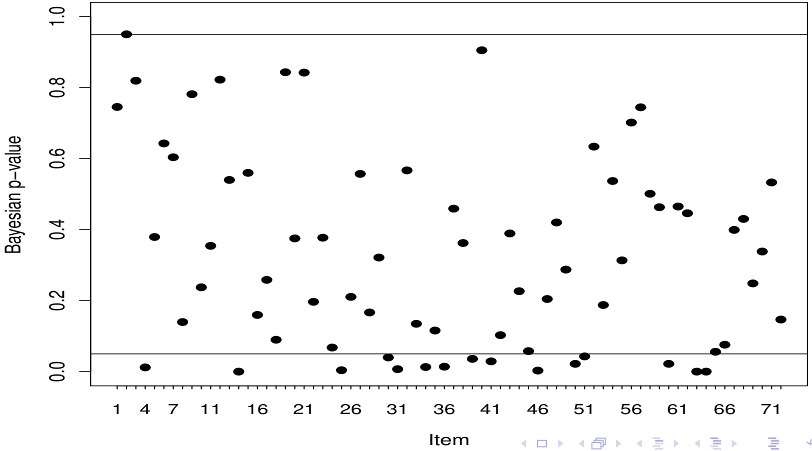
Real data anaylsis: first data set

- The ARMAH covariance structure was visited 80.1% of the MCMC iterations.
- This covariance matrix was selected.
- In Azevedo et al (2014) the unstructure covariance matrix was selected (using E(AIC), E(BIC), E(DIC)).
- A more parsimonious covariance structure model was selected by the RJMCMC algorithm.

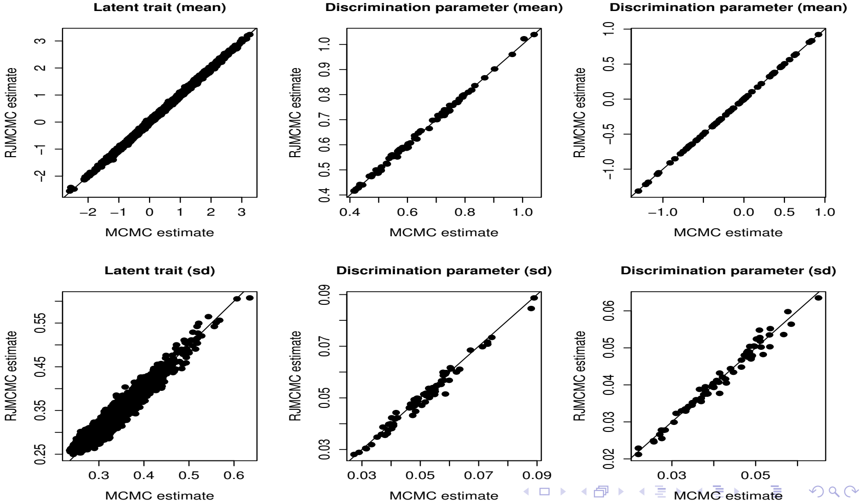
Real data anaylsis: first data set



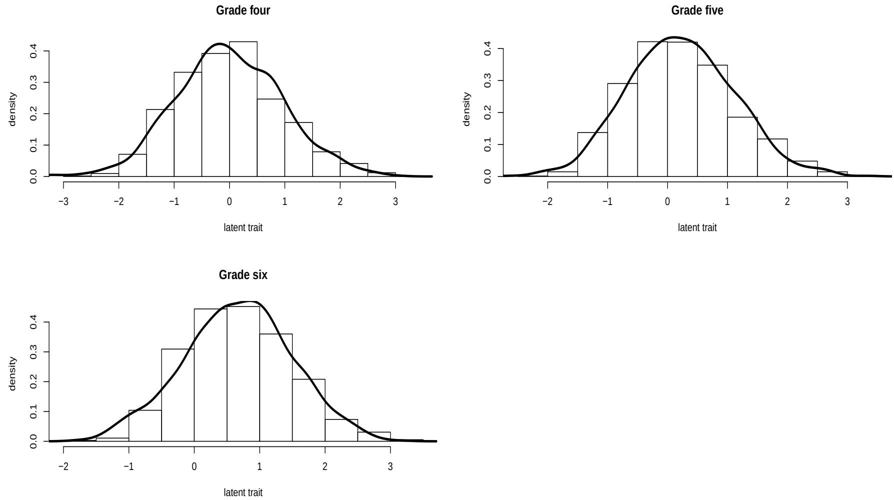
Real data anaylsis: first data set



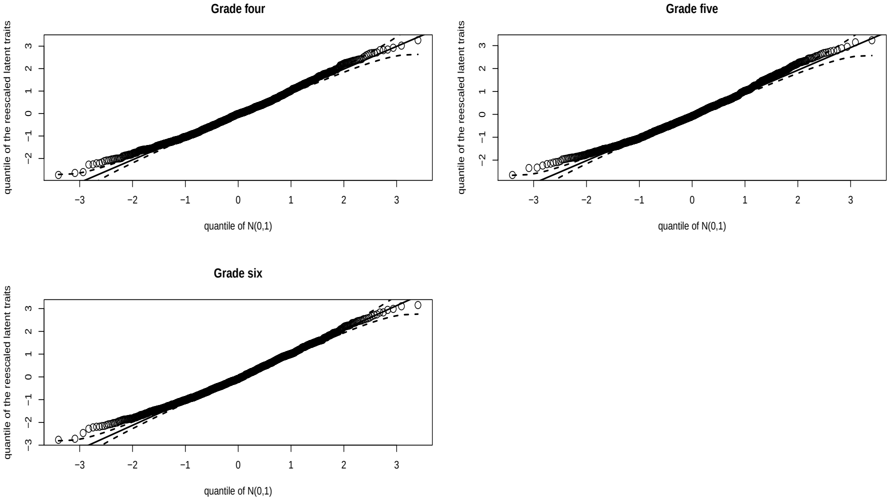
Real data anaylsis: first data set



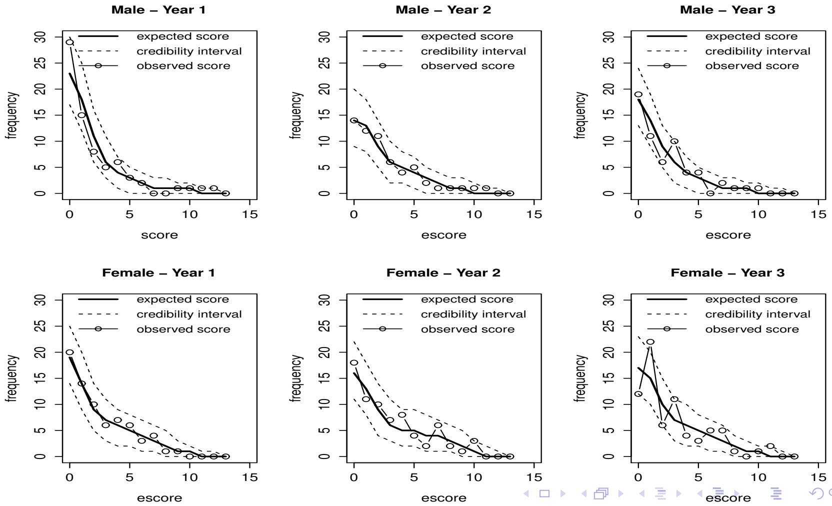
Real data anaylsis: first data set



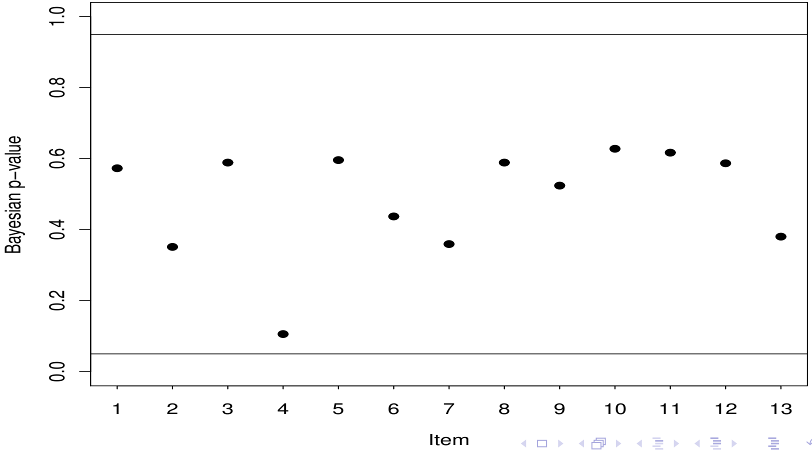
Real data anaylsis: first data set



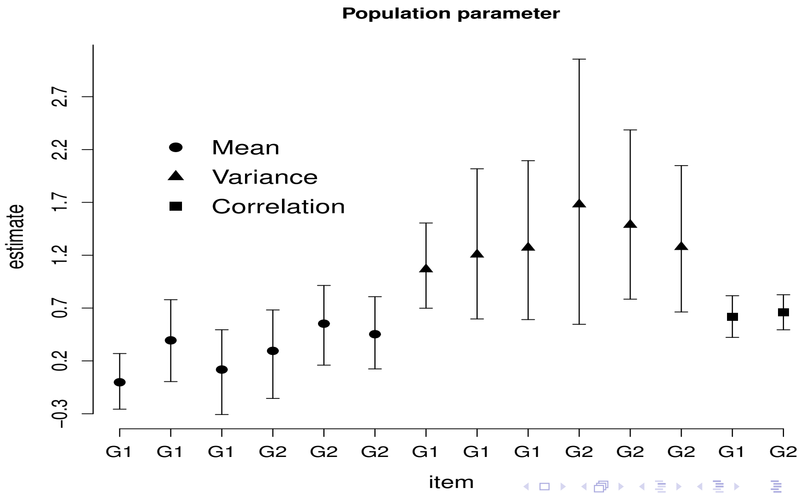
Real data anaylsis: second data set



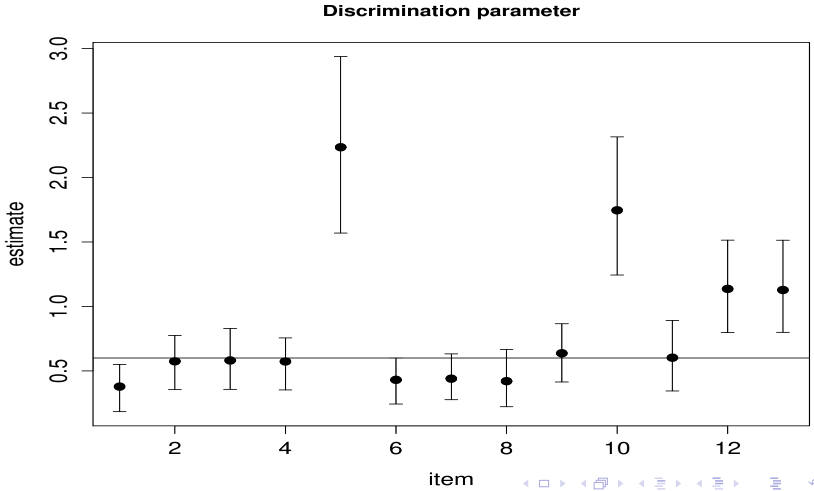
Real data anaylsis: second data set



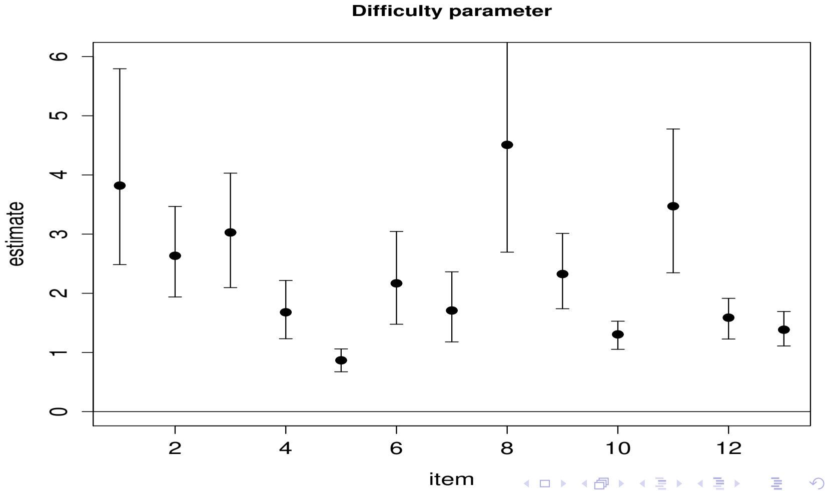
Real data anaylsis: second data set



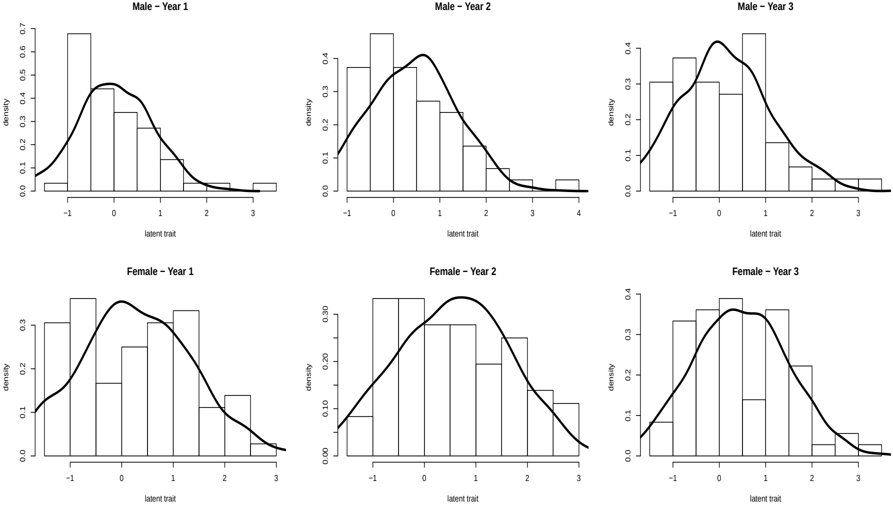
Real data anaylsis: second data set



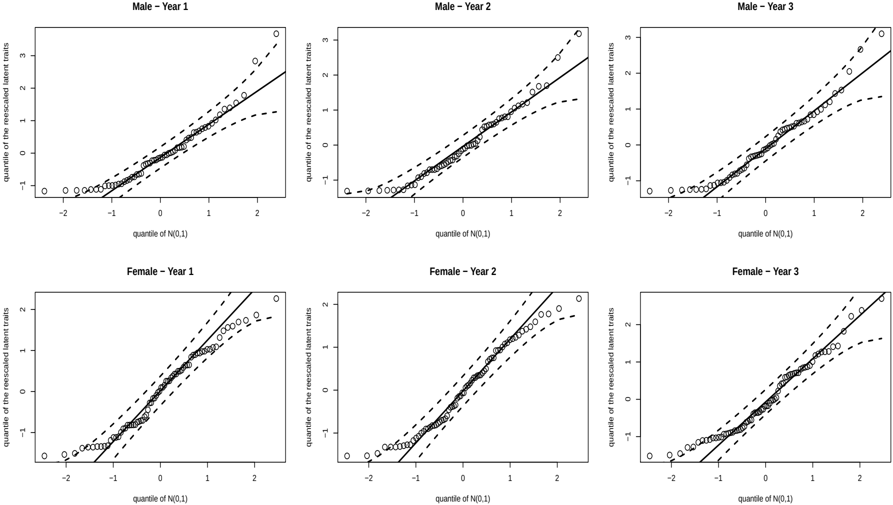
Real data anaylsis: second data set



Real data anaylsis: second data set



Real data anaylsis: second data set



Final conclusions

- Bayesian model fit assessment tools shown to be very useful mechanisms for model validation.
- Extensions in order to consider other types of item reponse, latent traits distributions, multilevel structures, among other possibilities, can be straightforwardly developed.

Some references

- Azevedo, C. L. N., Fox,J.-P. and Andrade, D. F. (2014), Bayesian longitudinal item response modeling with restricted covariance pattern structures, *Statistics & Computing, DOI: 10.1007/s11222-014-9518-5*.
- Tavares, H. R. (2001). Item response theory for longitudinal data (In Portuguese), PhD Thesis, Department of Statistics, University of São Paulo.
- Azevedo, C. L. N. (2008). Longitudinal multilevel multiple group in item response theory: estimation methods and structural selection under a Bayesian perspective (In Portuguese), PhD Thesis, Department of Statistics, University of São Paulo.

Some references

- Andrade, D.F. and Tavares, H.R. (2005), Item response theory for longitudinal data: population parameter estimation. J. Multivar. Anal. **95**, 1–22.
- Tavares, H.R. and Andrade, D.F. (2006), Item response theory for longitudinal data: item and population ability parameters estimation. Test **15**, 97–123.
- Azevedo, C.L.N., Andrade, D.F. and Fox, J.-P. (2012), A bayesian generalized multiple group IRT model with model-fit assessment tools. Comput. Stat. Data Anal. **56**, 4399-4412.

Thank you very much!

contact: cnaber@ime.unicamp.br