Multilevel Item Response Theory Models: An Introduction

Caio L. N. Azevedo,

Department of Statistics, State University of Campinas, Brazil

I CONCOLTRI, Universidad Nacional de Colômbia, May 2016

Acknowledgments to Prof. Dr. Héliton Tavares, Federal University of Pará, Brazil, for providing data sets.

Main goals

- Present some multilevel Item Response Theory (IRT) models and some of their applications.
- Bayesian inference through MCMC algorithms.
- Computational implementations by using WinBUGS/R2WinBUGS.
- For a introduction about IRT we recommend the short course of Prof. Dalton Andrade: "An Introduction to Item Response Theory".

Item Response Theory (IRT)

- Psychometric theory developed to meet needs in education. It consists of sets of models that consider the so-called latent variables or latent traits (variables that can not be measured directly as income, height and gender).
- Item Response Models (IRM): represent the relationship between latent traits (knowledge in some cognitive field, depression level, genetic predisposition in manifesting some disease) of experimental units (subjects, schools, enterprises, animals, plants) and items of a measuring instrument (cognitive tests, psychiatric questionnaires, genetic studies).

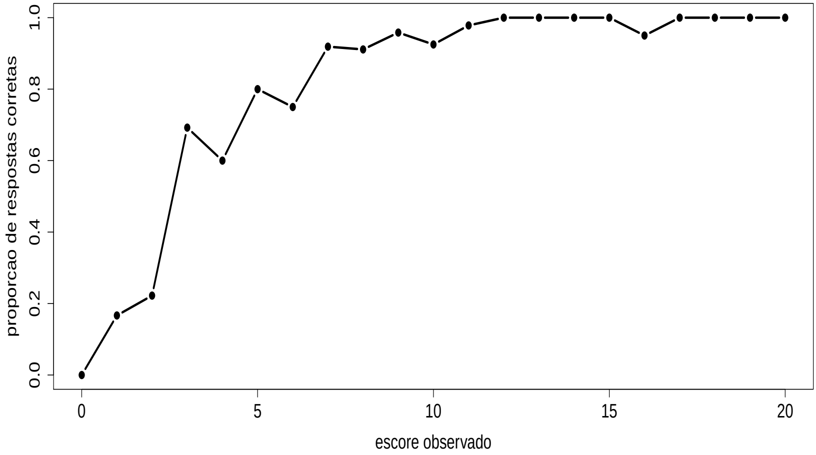
IRT: Brief review

- First models: Lord (1952), Rasch (1960) and Birnbaum (1957).
- Such modeling corresponds to/is related to the probability to get a certain score on each item.
- There are several families of IRM.

IRT models

- Type of response (is related to the link function): dichotomous, polytomous, counting process, continuous (unbounded and bounded), mixture type (continuous + dichotomous).
- Number of groups: one and multiple group.
- Number of tests (number of latent traits): univariate and multivariate.
- Latent trait (test) dimension: unidimensional and multidimensional.
- Measures over time-point (conditions): non-longitudinal (one time-point) and longitudinal.
- Nature of the latent trait : cumulative and non-cumulative (unfolding models).

Observed proportion of correct answer by score level



IRT data

- Without loss of generality, let us refer as “subjects” to the experimental units.
- A matrix of responses of the subjects to the items (binary, discrete, continuous) is available after the subjects were given to a test(s).
- Additionally, collateral information (explanatory covariables) such as gender, scholar grade, income etc could be available.

Binary IRT data

	Item			
Subject	1	2	3	4
1	0	1	0	0
2	0	0	0	0
3	0	0	1	1
4	1	0	1	0
5	0	1	0	0

Graded IRT data

	Item			
Subject	1	2	3	4
1	0	0	1	0
2	1	2	3	1
3	3	2	2	2
4	0	0	2	2
5	3	1	0	2

Three-parameter model

- Let Y_{ij} be the response of the subject j to item i (1, correct, 0, incorrect), $j = 1, 2, ..., n$, $i = 1, 2, ..., I$.

$$Y_{ij}|(\theta_j, \zeta_i) \overset{ind.}{\sim} \text{Bernoulli}(p_{ij}),$$

$$p_{ij} = c_i + (1 - c_i)F(\theta_j, \boldsymbol{\zeta}_i, \boldsymbol{\eta}_{F_i})$$

- Unidimensional, dichotomous, one group and univariate (non-longitudinal).

Three-parameter model: latent trait

- θ_j: latent trait of subject j.
- Usual assumption $\theta_j|(\mu_\theta, \psi_\theta, \boldsymbol{\eta}_\theta) \overset{i.i.d.}{\sim} D(\mu_\theta, \psi_\theta, \boldsymbol{\eta}_\theta)$, where $D(., ., .)$ stands for some distribution where $\mathcal{E}(\theta) = \mu_\theta$, $\mathcal{V}(\theta) = \psi_\theta$ (0 and 1, respectively, for model identification) and an additional vector of parameters (skewness, kurtosis) $\boldsymbol{\eta}_\theta$.

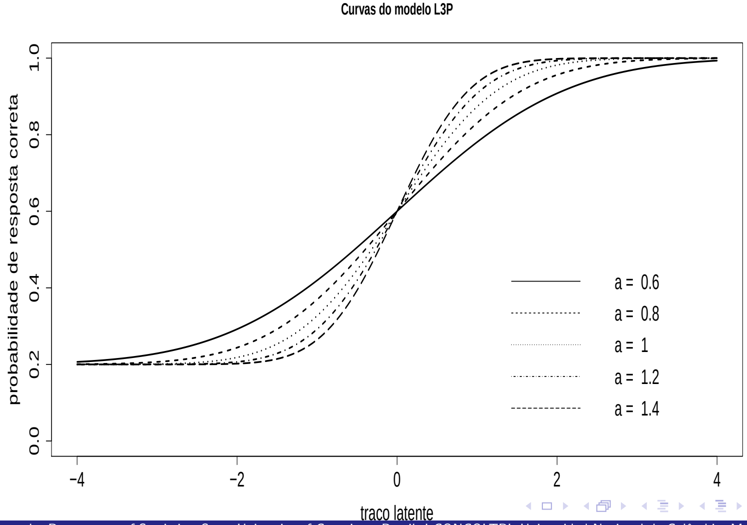
Three-parameter model (3PM): item parameters

- $\zeta_i = (a_i, b_i)'$.
- a_i: discrimination parameter (scale) of item i .
- b_i: difficulty parameter (location) of item i .
- c_i: approximate probability (low asymptote) of subjects with low level of the latent trait to get a correct response in item i (AKA guessing parameter).
- If $c_i = 0$ and $a_i = 1, c_i = 0$ we have, respectively, the two and one parameter models.

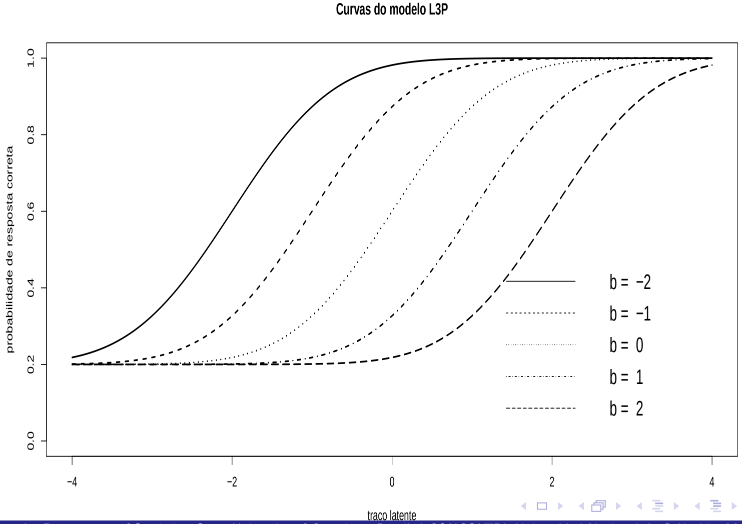
Three-parameter model: link function (item response function - IRF)

- $F(.)$ is an appropriate (in general) cumulative distribution function (cdf) related to a (continuous and real) random variable.
- $\boldsymbol{\eta}_{F_i}$ is (possibly a vector) of parameters related to the link function of item i.
- The most known choices are $F(\theta_j, \zeta_i) = \Phi(a_i(\theta_j - b_i))$ (probit) and $F(\theta_j, \zeta_i) = \frac{1}{1+e^{-a_i(\theta_j - b_i)}}$ (logit).
- Alternatives: cdf of the skew normal, skew-t, skew scale mixture, among others.

Examples of IRF for the 3PM (logistic link)



Examples of IRF for the 3PM (logistic link)



Multidimensional Compensatory three-parameter model

- Let Y_{ij} as before and:

$$Y_{ij}|(\boldsymbol{\theta}_j,\zeta_i) \overset{ind.}{\sim} \text{Bernoulli}(p_{ij}); p_{ij} = c_i + (1-c_i)F(\boldsymbol{\theta}_j,\boldsymbol{\zeta}_i,\eta_{F_i})$$

- $\boldsymbol{\theta}_j = (\theta_{j1},...,\theta_{jM})'$, θ_{jm}: latent trait of subject j related to dimension $m, m = 1,2,...,M$.

- Usual assumption $\boldsymbol{\theta}_{j.} = (\theta_{j1},\theta_{j2},....,\theta_{jm})'|(\boldsymbol{\mu_\theta},\boldsymbol{\Psi_\theta},\boldsymbol{\eta_\theta})$ $\overset{i.i.d}{\sim} D_M(\boldsymbol{\mu_\theta},\boldsymbol{\Psi_\theta},\boldsymbol{\eta_\theta})$, where $D(.,.,.)$ stands for some M-variate distribution with mean- vector $\mathcal{E}(\boldsymbol{\theta}) = \boldsymbol{\mu_\theta}$, covariance matrix $Cov(\boldsymbol{\theta}) = \boldsymbol{\Psi_\theta}$ and an additional vector of parameters (skewness, kurtosis) $\boldsymbol{\eta_\theta}$.

Population (latent traits) parameters

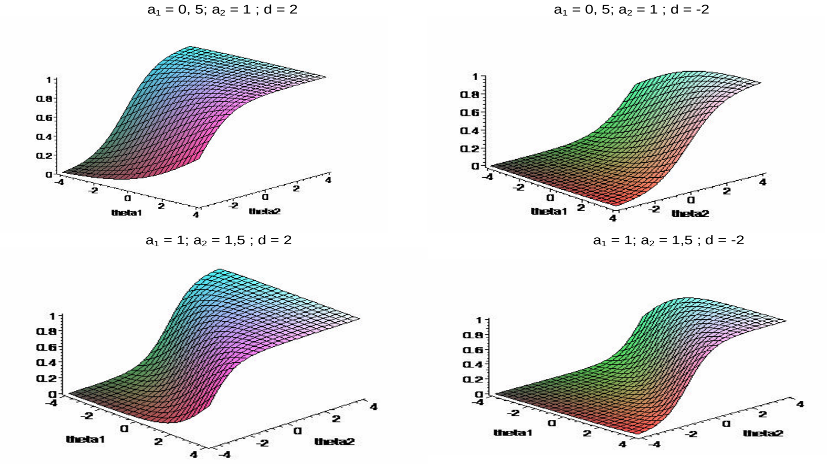
$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = \begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \vdots \\ \mu_{\boldsymbol{\theta}_M} \end{bmatrix} \text{ and } \boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \psi_{\theta_{12}} & \dots & \psi_{\theta_{1M}} \\ \psi_{\theta_{12}} & \psi_{\theta_2} & \dots & \psi_{\theta_{2M}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{1M}} & \psi_{\theta_{2M}} & \dots & \psi_{\theta_M} \end{bmatrix},$$

For model identification, $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$ and $\psi_{\theta_i} = 1, i = 1, 2, ..., M$ (that is, $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$ is a correlation matrix).

Multidimensional Compensatory three-parameter model

- $\zeta_i = (\boldsymbol{a}_i, d_i)'$.
- $\boldsymbol{a}_i = (a_{i1}, ..., a_{iM})'$, vector of parameters related to the discrimination of item i. d_i: parameter related to the difficulty of item i.
- Multidimensional difficulty: $\dfrac{-d_i}{\sqrt{\sum_{k=1}^{M} a_i^2}}$ (logistic link).
- Multidimensional discrimination: $\sqrt{\sum_{k=1}^{M} a_i^2}$ (logistic link).
- The other quantities are as defined before.

Item response surfaces (IRS)-logistic link (IRF) and $c_i = 0$



$a_1 = 0,5$; $a_2 = 1$; $d = 2$

$a_1 = 0,5$; $a_2 = 1$; $d = -2$

$a_1 = 1$; $a_2 = 1,5$; $d = 2$

$a_1 = 1$; $a_2 = 1,5$; $d = -2$

Three-parameter multiple group model

- Let Y_{ijk} be the response of the subject j, from group k to item i (1, correct, 0, incorrect), $j = 1, 2, ..., n_k$, $i = 1, ..., I$ and $k = 1, 2, ..., K$.

$$Y_{ijk}|(\theta_{jk}, \zeta_i) \overset{ind.}{\sim} \text{Bernoulli}(p_{ijk}), p_{ijk} = c_i + (1 - c_i)F(\theta_{jk}, \boldsymbol{\zeta}_i, \boldsymbol{\eta}_{F_i})$$

- θ_{jk}: latent trait of subject j from group k.
- Usual assumption $\theta_{jk}|(\mu_{\theta_k}, \psi_{\theta_k}, \boldsymbol{\eta}_{\theta_k}) \overset{i.i.d}{\sim} D(\mu_{\theta_k}, \psi_{\theta_k}, \boldsymbol{\eta}_{\theta_k})$, where $D(.,.,.)$ stands for some distribution $\mathcal{E}(\theta) = \mu_{\theta_k}$, $\mathcal{V}(\theta_k) = \psi_{\theta_k}$ (0 and 1, for the reference group, respectively, for model identification) and an additional vector of parameters (skewness, kurtosis) $\boldsymbol{\eta}_{\theta_k}$.

Three-parameter multiple group model

- In general we expect to observe a large number of subjects in each group and a small number of groups. The groups are independent in the sense that we have the each subject belongs to one and only group.
- All the other quantities remain the same.

Three-parameter longitudinal model

- Let Y_{ijt} be the response of the subject j, in time-point t to item i (1, correct, 0, incorrect), $j = 1, 2, ..., n$, $i = 1, ..., I$ and $t = 1, 2, ..., T$.

$$Y_{ijt}|(\theta_{jt}, \zeta_i) \overset{ind.}{\sim} \text{Bernoulli}(p_{ijt}), p_{ijt} = c_i + (1 - c_i)F(\theta_{jt}, \zeta_i, \boldsymbol{\eta}_{F_i})$$

- θ_{jt}: latent trait of subject j in time-point t.
- Usual assumption $\boldsymbol{\theta}_{j.} = (\theta_{j1}, \theta_{j2}, ..., \theta_{jT})'|(\boldsymbol{\mu_\theta}, \boldsymbol{\Psi_\theta}, \boldsymbol{\eta_\theta})$ $\overset{i.i.d.}{\sim} D_T(\boldsymbol{\mu_\theta}, \boldsymbol{\Psi_\theta}, \boldsymbol{\eta_\theta})$, where $D(.,.,.)$ stands for some T-variate distribution with mean- vector $\mathcal{E}(\boldsymbol{\theta}) = \boldsymbol{\mu_\theta}$, covariance matrix $Cov(\boldsymbol{\theta}) = \boldsymbol{\Psi_\theta}$ and an additional vector of parameters (skewness, kurtosis) $\boldsymbol{\eta_\theta}$.

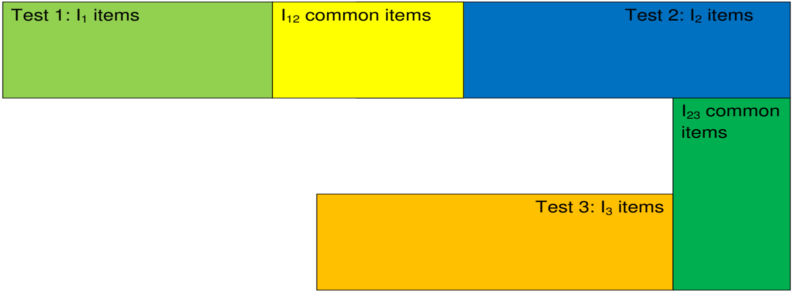
Population (latent traits) parameters

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = \begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \vdots \\ \mu_{\boldsymbol{\theta}_T} \end{bmatrix} \text{ and } \boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \psi_{\theta_{12}} & \dots & \psi_{\theta_{1T}} \\ \psi_{\theta_{12}} & \psi_{\theta_2} & \dots & \psi_{\theta_{2T}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{1T}} & \psi_{\theta_{2T}} & \dots & \psi_{\theta_T} \end{bmatrix},$$

For model identification, $\mu_{\theta_1} = 1$ and $\psi_{\theta_1} = 1$.

For multiple group and/or longitudinal framework, one or more different tests are administered by the examinees of each group/ in each time point. The tests have common items and the structure can be recognized as an incomplete block design.

Example of tests design



Test 1: I_1 items

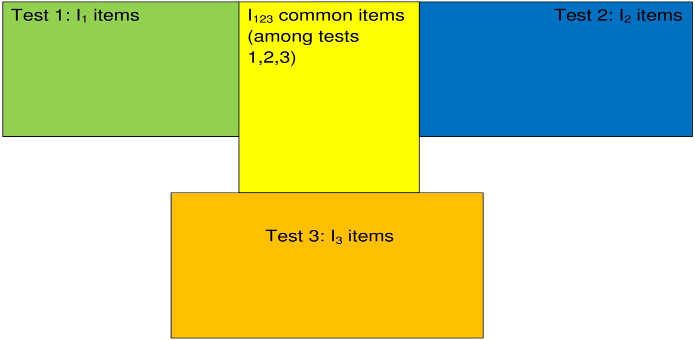
I_{12} common items

Test 2: I_2 items

I_{23} common items

Test 3: I_3 items

Example of tests design



Test 1: I_1 items

I_{123} common items (among tests 1,2,3)

Test 2: I_2 items

Test 3: I_3 items

Bayesian inference

- Let us consider the three-parameter one group model.
- Original likelihood (under the conditional independence assumptions)

$$L(\boldsymbol{\theta},\zeta) = \prod_{i=1}^{I}\prod_{j=1}^{n} p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}$$

$\boldsymbol{\theta} = (\theta_1, ..., \theta_n)'$ e $\zeta = (\zeta_1, ..., \zeta_I)'$.

Bayesian inference

- Prior distribution

$$p(\boldsymbol{\theta},\zeta) \quad = \quad \prod_{i=1}^{n} p(\theta_j) \prod_{i=1}^{I} p(\zeta_i) = \prod_{i=1}^{n} p(\theta_j) \prod_{i=1}^{I} p(\boldsymbol{a}_i)p(\boldsymbol{b}_i)p(\boldsymbol{c}_i)$$

where $\boldsymbol{a} = (a_1, ..., a_I)'$, $\boldsymbol{b} = (b_1, ..., b_I)'$ e $\boldsymbol{c} = (c_1, ..., c_I)'$.

Bayesian inference

- Joint posterior distribution:

$$p(\boldsymbol{\theta}, \zeta | \boldsymbol{y}) = \prod_{i=1}^{I} \prod_{j=1}^{n} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}} \prod_{i=1}^{n} p(\theta_j) \prod_{i=1}^{I} p(\boldsymbol{a}_i) p(\boldsymbol{b}_i) p(\boldsymbol{c}_i)$$

- It is intractable but the so-called full conditional distributions are either known (and easy to sample from) or they can be sampled by using some (auxiliary) algorithm such as the Metropolis-Hastings, slice sampling, adaptive rejection sampling.

Augmented data scheme (probit link)

- It facilitates the implementation of MCMC (and of the CADEM) algorithms.
- Depending of the augmented data structure it facilitates the implementation of the model in WinBUGS/OpenBUGS/JAGS/Stan.
- Useful to define the so-called (latent/augmented) residuals (model checking).
- Useful to define more general IRT models.

Augmented data scheme (probit link)

- Let us consider $a_i\theta_j - d_i$, where $d_i = a_i b_i$ e $a_i(\theta_j - b_i)$.
- If $c_i = 0, \forall i$ (two-parameter) model (Albert (1992)):

$$Z_{ij}|(\theta_j, \zeta_i, y_{ij}) \sim N(a_i\theta_j - d_i, 1),$$

where y_{ij} is the indicator of Z_{ij} being greater than zero and $d_i = a_i b_i$.

- For other link functions (IRF) it is possible to define other augmented data schemes.

Augmented data scheme (probit link)

- For the three-parameter model we have two options.
- Beguin and Glas' scheme (2001).

$$Z_{ij}|(\theta_j,\zeta_i,W_{ij}) \sim N(a_i\theta_j - d_i, 1),$$

where W_{ij} is the indicator of Z_{ij} being greater than zero, and

$$\begin{aligned}
P(W_{ij}=1|Y_{ij}=1,\theta_j,\zeta_i) &\propto \Phi(a_i\theta_j - d_i)\\
P(W_{ij}=0|Y_{ij}=1,\theta_j,\zeta_i) &\propto c_i(1-\Phi(a_i\theta_j - d_i))\\
P(W_{ij}=1|Y_{ij}=0,\theta_j,\zeta_i) &= 0\\
P(W_{ij}=0|Y_{ij}=0,\theta_j,\zeta_i) &= 1
\end{aligned}$$

Augmented data scheme (probit link)

- Sahu's scheme (2002).

Z_{ij}	w_{ij}	y_{ij}
$Z_{ij} \| (\theta_j, \zeta_i, w_{ij}, y_{ij}) \overset{ind.}{\sim} N(a_i\theta_j - d_i)\mathbb{1}_{(-\infty,0)}(z_{ij})$	0	0
$Z_{ij} \| (\theta_j, \zeta_i, w_{ijk}, y_{ij}) \overset{ind.}{\sim} N(a_i\theta_j - d_i)\mathbb{1}_{(0,\infty)}(z_{ij})$	0	1
$Z_{ij} \| (\theta_j, \zeta_i, w_{ij}, y_{ij}) \overset{ind.}{\sim} N(a_i\theta_j - d_i)$	1	1

W_{ij}	z_{ij}
1	negative
bernoulli(c_i)	positive

Bayesian modeling

- Hierarchical representation (based on the augmented data scheme) of the two-parameter probit model.

$$
\begin{aligned}
Z_{ij}|(\theta_j, \zeta_i, y_{ij}) &\overset{ind}{\sim} N(a_i\theta_j - d_i, 1) \\
\theta_i &\overset{i.i.d.}{\sim} N(0, 1) \\
a_i &\overset{i.i.d.}{\sim} N(\mu_a, \psi_a)\mathbb{1}(a_i)_{(0,\infty)} \\
d_i &\overset{i.i.d.}{\sim} N(\mu_d, \psi_d)
\end{aligned}
$$

Bayesian modeling

- Augmented data likelihood (under the conditional independence assumptions)

$$p(\boldsymbol{z}, \boldsymbol{w}|\boldsymbol{y}) = \prod_{j=1}^{n}\prod_{i=1}^{I} p(z_{ij}, w_{ij}|y_{ij})$$

- Joint (augmented) posterior

$$p(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\theta}, \zeta|\boldsymbol{y}) = \prod_{j=1}^{n}\prod_{i=1}^{I} p(z_{ij}, w_{ij})p(y_{ij})\prod_{i=1}^{n} p(\theta_j)\prod_{i=1}^{I} p(\boldsymbol{a}_i)p(\boldsymbol{b}_i)p(\boldsymbol{c}_i)$$

MCMC algorithm for the three-parameter probit model: original likelihood

Let (.) denote the set of all necessary parameters, then:

1. Start the algorithm by choosing suitable initial values.

   Repeat steps 2–3.

2. Simulate θ_j from $\theta_{j.} \mid (.), j = 1, ..., n$.

3. Simulate (a_i, b_i) from $(a_i, b_i) \mid (.)$, i =1,...,I. (may be done separately for each parameter)

4. Simulate c_i from $c_i \mid (.)$, i =1,...,I.

MCMC algorithm for the three-parameter probit model: augmented likelihood

Let (.) denote the set of all necessary parameters, then:

1. Start the algorithm by choosing suitable initial values.

   Repeat steps 2–3.

2. Simulate Z_{ij} from $Z_{ij} \mid (.), i = 1, ..., I, \ j = 1, ..., n$.
3. Simulate W_{ij} from $W_{ij} \mid (.), i = 1, ..., I, \ j = 1, ..., n$.
4. Simulate θ_j from $\theta_{j.} \mid (.), j = 1, ..., n$.
5. Simulate (a_i, b_i) from $(a_i, b_i) \mid (.)$, i =1,...,I. (may be done separately for each parameter)
6. Simulate c_i from $c_i \mid (.)$, i =1,...,I.

Bayesian modeling

- The above hierarchical representation can be easily implemented in the WinBUGS (OpenBugs,JAGS, Stan) packages.
- The two augmented data schemes for the three-parameter are not easily (impossilble?) implemenented in those packages. However it can be implemented by using the original likelihood (depending on the IRF).
- Also, usual MCMC algorithms can be implemented in R programs using the so-called full conditional distributions (easy to obtain and, generally, easy to sample from).

WinBUGS code: parameter probit model

Show the probit2P.r file.

Model validation and comparison

- Posterior predictive checking (plots, measures of goodness of fit, Bayesian p-value).
- Residual analysis.
- Statistics of model comparison (AIC, BIC, E(AIC), E(BIC), DIC, E(DIC), LPLM).
- Statistics of goodness of fit.

Multilevel Data

- Multilevel data are characterized when sample (experimental) units are nested in other ones.
- We can have two or more levels.
- For example:
  - Students (level 1) nested within schools (level 2) (two-level structure).
  - Students (level 1) nested within classrooms (level 2) nested within schools (level 3) (three-level structure).
  - Longitudinal data: measurement occasions (level 1) nested within subjects (level 2) (two-level structure).

Multilevel models

- In general we expect to observe some dependence (correlation) among the sample units (observations) that are nested (within groups).
- Usefull (generally in a very easy way):
  - To model and measure separately effects of interest in different levels.
  - To account for different sources of variability.
  - To accommodate dependency structures.
- AKA hierarchical models (american nomenclature) whereas multilevel (european nomenclature).
- Closely related to the so-called mixed models (repeated measurement data).

A very simple two - level multilevel linear model

- Let us suppose that estimates of the latent traits θ_{jk}, $j = 1, ..., n_k; k = 1, ..., K$ from subjects (in any number) belonging to different groups (several of them), for example schools.
- We suspect that the subjects that belong to the same group are more similar among them when we compared with those from other schools.
- We want to consider this nested structure through a linear multilevel model.

A very simple two - level multilevel linear model

$$
\begin{aligned}
\theta_{jk} &= \beta_{0k} + \xi_{jk} \text{ (level 1)} \\
\beta_{0k} &= \gamma + u_k \text{ (level 2)} \\
\xi_{jk} &\overset{i.i.d}{\sim} N(0,\psi), u_k \overset{i.i.d}{\sim} N(0,\tau), \xi_{jk} \perp u_k \ \forall j,k
\end{aligned}
$$

It incorporates dependence among the subjects belonging to the same group (they will be more similar to each other, than to those subjects belonging to other groups).

A Two - level multilevel linear model with covariates

- Let us suppose that estimates of the latent traits θ_{jk}, $j = 1, ..., n_k; k = 1, ..., K$ from subjects (in any number) belonging to different groups (a large number) and some collateral information (covariables), says $X_{rjk}, r = 1, ..., p$ are available.
- We want to measure the impact of those covariables on the latent traits, to use this information to improve the late traits estimates and also to consider the nested structure through a linear multilevel model.

A Two - level multilevel linear model with covariables

$$
\begin{aligned}
\theta_{jk} &= \beta_{0k} + \sum_{r=1}^{p} \beta_r X_{rjk} + \xi_{jk} \text{ (level 1)} \\
\beta_{0k} &= \gamma + u_k \text{ (level 2)} \\
\xi_{jk} &\overset{i.i.d}{\sim} N(0,\psi), u_k \overset{i.i.d}{\sim} N(0,\tau), \xi_{jk} \perp u_k \ \forall j,k
\end{aligned}
$$

Besides to incorporate dependence among the subjects belonging to the same group, this model considers additional information to estimate the latent traits and allows to measure the impact of those information in the latent traits.

A General Two - level multilevel linear model

$$
\begin{aligned}
\theta_{jk} &= \beta_{0k} + \sum_{r=1}^{p} \beta_r X_{rjk} + \xi_{jk} = \boldsymbol{X}_{jk}\boldsymbol{\beta}_k + \xi_{jk} \text{ (level 1)} \\
\boldsymbol{\beta}_k &= \boldsymbol{W}_k \gamma + \boldsymbol{u}_k \text{(level 2)} \\
\xi_{jk} &\overset{i.i.d}{\sim} N(0,\psi), \boldsymbol{u}_k \overset{i.i.d}{\sim} N_{(p+1)}(\mathbf{0},\boldsymbol{\Omega}), \xi_{jk} \perp \boldsymbol{u}_k \ \forall j,k
\end{aligned}
$$

Besides to incorporate dependence among the subjects belonging to the same group, this model considers additional information to estimate the latent traits and allows to measure the impact of those information in the latent traits.

Some applications of multilevel modeling in IRT

- Data with natural hierarchical (nested) structures: students nested in classrooms (and/or schools).
- Longitudinal data: students followed at the final of each scholar grade.
- DIF (Differential Item Functioning): a nested structure in the item parameters.
- Lack of local independence: correlation among the responses that can not be accounted by multidimensional models.

Bayesian inference for multilevel IRT model

- The joint posterior distribution will depend on the multilevel structure adopted along with the prior distribution required by the new parameters.
- The original/augmented likelihood can be modified as well as the prior distributions for the latent traits and item parameters, depending on the multilevel structure adopted.

A multilevel IRT model

- Let us consider the three parameter model (for multiple group) and the three multilevel models presented before.

$$
\begin{aligned}
Y_{ijk}|(\theta_{jk},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijk}) \\
p_{ijk} &= c_i + (1-c_i)F(\theta_{jk},\boldsymbol{\zeta}_i,\boldsymbol{\eta}_{F_i}) \\
\theta_{jk} &= \beta_{0k} + \xi_{ijk} \text{ (level 1)} \\
\beta_{0k} &= \gamma + u_k \text{ (level 2)} \\
\xi_{jk} &\overset{i.i.d}{\sim} N(0,\psi), u_k \overset{i.i.d}{\sim} N(0,\tau), \xi_{jk} \perp u_k \ \forall j,k
\end{aligned}
$$

A multilevel IRT model

- Besides to incorporate dependence among the subjects belonging to the same group, this model considers additional information to estimate the latent traits and allows to measure the impact of those information in the latent traits.

A multilevel IRT model with covariables

$$
\begin{array}{rcl}
Y_{ijk}|(\theta_{jk},\zeta_i) & \overset{ind.}{\sim} & \text{Bernoulli}(p_{ijk}) \\
p_{ijk} & = & c_i + (1 - c_i)F(\theta_{jk},\boldsymbol{\zeta}_i,\boldsymbol{\eta}_{F_i}) \\
\theta_{jk} & = & \beta_{0k} + \sum_{r=1}^{p}\beta_r X_{rjk} + \xi_{ijk} \text{ (level 1)} \\
\beta_{0k} & = & \gamma + u_k \text{ (level 2)} \\
\xi_{jk} & \overset{i.i.d}{\sim} & N(0,\psi), u_k \overset{i.i.d}{\sim} N(0,\tau), \xi_{jk} \perp u_k \ \forall j,k
\end{array}
$$

Besides to incorporate dependence among the subjects belonging to the same group, this model considers additional information to estimate the latent traits and allows to measure the impact of those information in the latent traits.

A general multilevel IRT model

$$\begin{aligned}
Y_{ijk}|(\theta_{jk},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijk})\\
p_{ijk} &= c_i + (1-c_i)F(\theta_{jk},\zeta_i,\boldsymbol{\eta}_{F_i})\\
\theta_{jk} &= \beta_{0k} + \sum_{r=1}^{p}\beta_r X_{rjk} + \xi_{jk} = \boldsymbol{X}_{jk}\boldsymbol{\beta}_k + \xi_{jk} \text{ (level 1)}\\
\boldsymbol{\beta}_k &= \boldsymbol{W}_k\gamma + u_k \text{ (level 2)}\\
\xi_{jk} &\overset{i.i.d}{\sim} N(0,\psi), \boldsymbol{u}_k \overset{i.i.d}{\sim} N_{(p+1)}(\mathbf{0},\boldsymbol{\Omega}), \xi_{jk}\perp\boldsymbol{u}_k \;\forall j,k
\end{aligned}$$

It incorporates dependence among the subjects belonging to the same group (they will be more similar to each other, than to those subjects belonging to other groups).

A longitudinal multilevel IRT model - Uniform covariance matrix

$$
\begin{aligned}
Y_{ijt}|(\theta_{jt},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijt}) \\
p_{ijt} &= c_i + (1-c_i)F(\theta_{jt},\boldsymbol{\zeta}_i,\boldsymbol{\eta}_{F_i}) \\
\theta_{jt} &= \mu_{\theta_t} + \sqrt{\psi_{\theta_t}}\tau_j + \xi_{jt} \text{ (level 1)} \\
\tau_j &\overset{i.i.d.}{\sim} N(0,\sigma^2) \text{ (level 2)} \\
\xi_{jt} \overset{i.i.d.}{\sim} N(0,\psi_t), \xi_{jt} &\perp \tau_j, \forall j, t
\end{aligned}
$$

It incorporates dependence among the latent traits within subjects.

Implied covariance matrix

Heteroscedastic uniform model - HU

$$
\boldsymbol{\Psi_\theta} = \begin{bmatrix}
\psi^*_{\theta_1} & \sqrt{\psi^*_{\theta_1}}\sqrt{\psi^*_{\theta_2}}\rho^*_\theta & \cdots & \sqrt{\psi^*_{\theta_1}}\sqrt{\psi^*_{\theta_T}}\rho^*_\theta \\
\sqrt{\psi^*_{\theta_1}}\sqrt{\psi^*_{\theta_2}}\rho^*_\theta & \psi^*_{\theta_2} & \cdots & \sqrt{\psi^*_{\theta_2}}\sqrt{\psi^*_{\theta_T}}\rho^*_\theta \\
\vdots & \vdots & \ddots & \vdots \\
\sqrt{\psi^*_{\theta_1}}\sqrt{\psi^*_{\theta_T}}\rho^*_\theta & \sqrt{\psi^*_{\theta_2}}\sqrt{\psi^*_{\theta_T}}\rho^*_\theta & \cdots & \psi^*_{\theta_T}
\end{bmatrix},
$$

where $\psi^*_{\theta_t} = \psi_{\theta_t}(1+\sigma^2)$, and Corre$(\theta_t, \theta_{t'}) = \rho^*_\theta = \dfrac{\sigma^2}{1+\sigma^2}, t \neq t'$ t =1,2,...,T.

A longitudinal multilevel IRT model - Hankel (Heteroscedastic) covariance matrix

$$
\begin{aligned}
Y_{ijt}|(\theta_{jt},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijt}) \\
p_{ijt} &= c_i + (1-c_i)F(\theta_{jt},\zeta_i,\boldsymbol{\eta}_{F_i}) \\
\theta_{jt} &= \mu_{\theta_t} + \tau_j + \xi_{jt} \text{ (level 1)} \\
\tau_j &\overset{i.i.d.}{\sim} N(0,\sigma^2) \text{ (level 2)} \\
\xi_{jt} \overset{i.i.d.}{\sim} N(0,\psi_t), \xi_{jt} &\perp \tau_j, \forall j, t
\end{aligned}
$$

It incorporates dependence among the latent traits related to the same subjects.

Implied covariance matrix

Heteroscedastic covariance model - HC

$$\mathbf{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi^*_{\theta_1} & \sigma^2 & \dots & \sigma^2 \\ \sigma^2 & \psi_{\theta_2} & \dots & \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \dots & \psi_{\theta_T} \end{bmatrix},$$

where $\psi^*_{\theta_t} = \psi_{\theta_t} + \sigma^2$, and Corre$(\theta_t, \theta_{t'}) = \dfrac{\sigma^2}{\sigma^2 + \psi_{\theta_t}}$, $t \neq t'$, t,t' =1,2,...,T

A multilevel IRT model for DIF (item level)

$$
\begin{aligned}
Y_{ijk}|(\theta_{jk},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijk}) \\
p_{ijk} &= c_i + (1-c_i)\Phi(a_{ik}\theta_{jk} - d_{ik}) \\
d_{ik} &= \underbrace{d_i}_{\text{item parameter}} + \underbrace{\beta_{ik}}_{\text{groups nested within items}} \\
a_{ik} &= \underbrace{a_i}_{\text{item parameter}} + \underbrace{e^{\alpha_{ik}}}_{\text{groups nested within items}} \\
\theta_{jk} &\overset{ind.}{\sim} N(\mu_{\theta_k}, \psi_{\theta_k}) \\
\beta_{ik} \overset{i.i.d.}{\sim} N(0,\sigma^2_{\beta_i}) &\perp \alpha_{ik} \overset{i.i.d.}{\sim} N(0,\sigma^2_{\alpha_i}), \forall i, k
\end{aligned}
$$

Application 1 : Multiple group IRT data

- Originally, it is a longitudinal (with dropouts) with 4 time-points.
- 568 first-grade students were selected from eight public primary schools (at the first-time point). Along the subsequent grades, some students dropped out from the study for different reasons. The present data set consists of the following number of students, from the first up to the fourth grade: 556, 556, 401 and 295.
- The students are nested in classes and classes are nested in schools.

Application 1 : Multiple group IRT data

- Available information: dissertative items correct as right/wrong, age of student, classroom, gender, school, teacher.
- Analysis presented in Azevedo et al (2012) revelead that posterior correlations among the latent traits are not significative and, therefore, a multiple group IRT model can be considered.
- Four tests (corresponding to each grade): Teste 1 - 20 items. For grade two till four, the responses to the 20 new items and the preceding 20 test items are considered, which leads to 40 items for each grade and a total of 80 different test items.

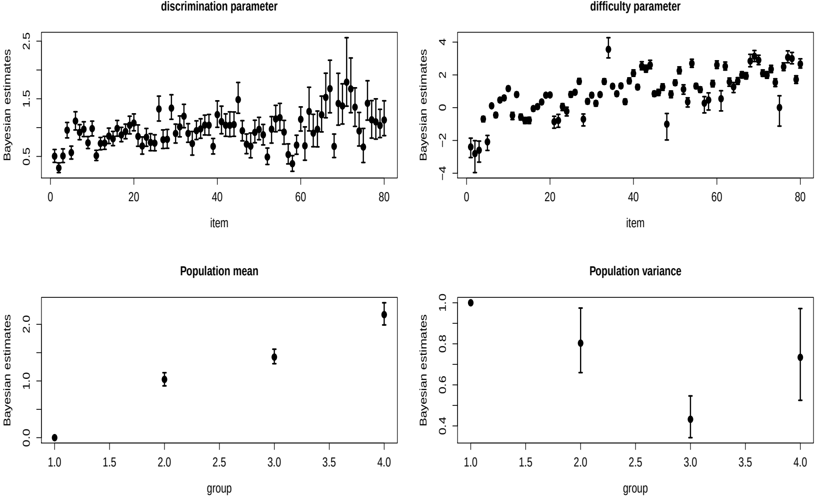
Tests design

	Item			
Test	1 - 20	21 - 40	41 - 60	61 - 80
1	■			
2	■	■		
3		■	■	
4			■	■

Inference

- Since the item were dissertative we fitted two pameter multiple group model ($c_i = 0$), that is $Y_{ijk}|(\theta_{jk}, \zeta_i) \overset{ind.}{\sim} \text{Bernoulli}(p_{ijk}), p_{ijk} = \Phi(a_i\theta_{jk} - d_i)$, remembering that $b_i = d_i/a_i$ (difficulty parameter).
- Priors: $\theta_{jk} \overset{ind.}{\sim} N(\mu_{\theta_k}, \psi_{\theta_k})$, $\mu_{\theta_k} = 0, \psi_{\theta_k} = 1$ (for model identification, reference group: 1), $a_i \overset{i.i.d.}{\sim} \text{lognormal}(0, 0.25)$, $d_i \overset{i.i.d}{\sim} N(0, 4)$, $\mu_{\theta_k} \overset{ind.}{\sim} N(0, 10)$ and $\psi_{\theta_k} \overset{ind.}{\sim} \text{Ga}(0.1, 0.1)$ where $X \sim \text{Ga}(r, s)$ implies that $\mathcal{E}(X) = rs$.
- Show Bugs code (probit2PnormMGM.r file).

Results



discrimination parameter
Bayesian estimates
0.5
1.5
2.5
0
20
40
60
80
item
difficulty parameter
Bayesian estimates
−4
−2
0
2
4
0
20
40
60
80
item
Population mean
Bayesian estimates
0.0
1.0
2.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
group
Population variance
Bayesian estimates
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
group

Application 2 : Multilevel IRT data

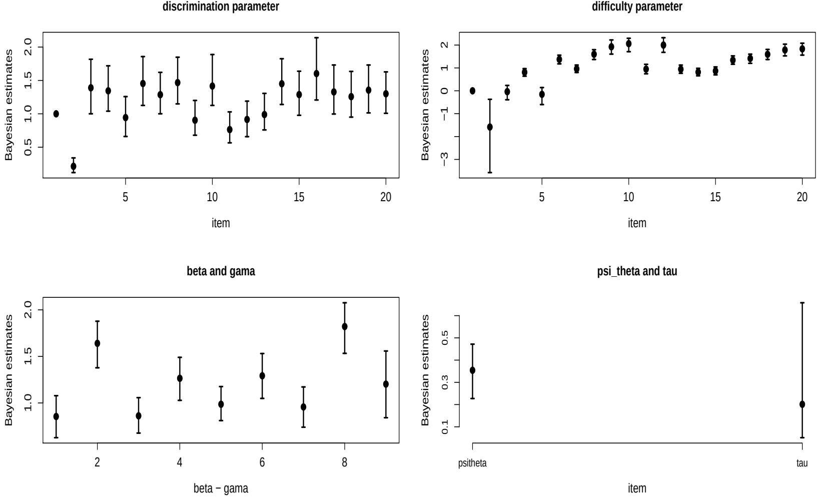
- Corresponds to the the data of Application 1, taking only the first time-point (grade 1), considering a two level nested structure, that is, students within schools (8 groups), without covariables.

$$
\begin{array}{rcl}
Y_{ijk}|(\theta_{jk},\zeta_i) & \overset{ind.}{\sim} & \text{Bernoulli}(p_{ijk}) \\
p_{ijk} & = & \Phi(a_i\theta_{jk} - d_i) \\
\theta_{jk} & = & \beta_{0k} + \xi_{ijk} \text{ (level 1)} \\
\beta_{0k} & = & \gamma + u_k \text{ (level 2)} \\
\xi_{jk} & \overset{i.i.d}{\sim} & N(0,\psi), u_k \overset{i.i.d}{\sim} N(0,\tau), \xi_{jk} \perp u_k \ \forall j,k
\end{array}
$$

Inference

- In this case, it is more suitable than the multiple group model, since we have 8 groups. Instead of estimate 14 population parameters (mean and variances) we estimate two variance components (ψ, τ), eight random effects β_{0k} and one location parameter γ (11).
- Priors: Defined in the previews slide for θ_{jk} and β_{0k}, $a_i \overset{i.i.d.}{\sim}$ lognormal$(0, 0.25)$ and $d_i \overset{i.i.d}{\sim} N(0,4)$, $\tau \sim N(0,100)$, $\psi \sim \text{Ga}(0.1, 0.1)$ and $\tau \sim \text{Ga}(0.1, 0.1)$.
- Model identification $a_1 = 1, b_1 = 0$.
- Show Bugs code (probit2PnormMult.r file).

Results



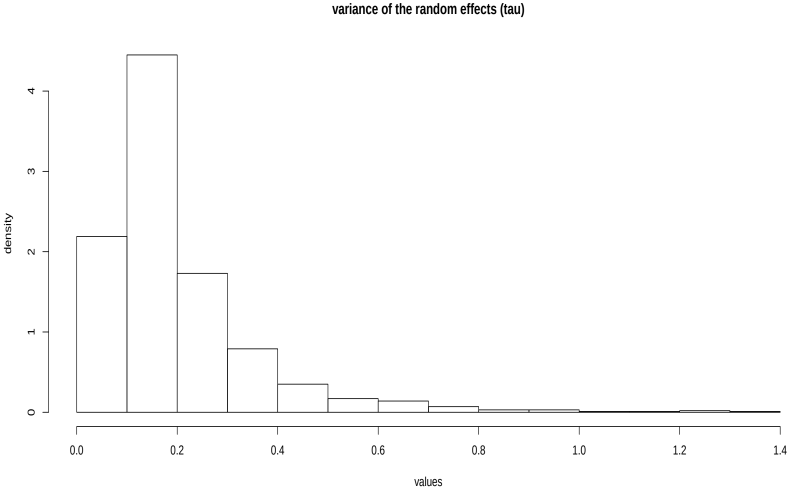
discrimination parameter

difficulty parameter

beta and gama

psi_theta and tau

Results


variance of the random effects (tau)
density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
values

Application 3 : Multilevel IRT data with covariables

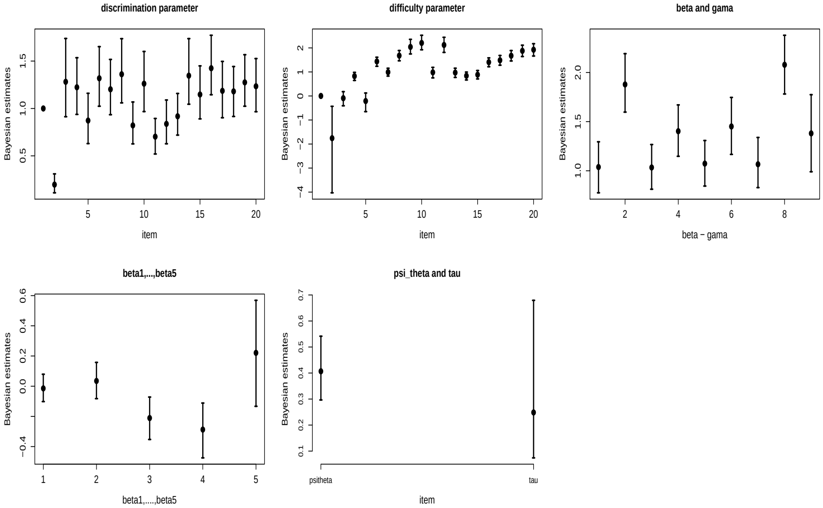
- Corresponds to the the data of Application 2 considering three - covariables: age, gender (0: male, 1: male), classroom (factor with four levels, classroom - 1A; 1B; 1C; 1D, reference level 1A)

$$
\begin{aligned}
Y_{ijk}|(\theta_{jk},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijk}) \\
p_{ijk} &= \Phi(a_i\theta_{jk} - d_i) \\
\theta_{jk} &= \beta_{0k} + \beta_1(age_{jk} - 7) + \beta_2 X_{1jk} + \beta_3 X_{2jk} + \beta_4 X_{3jk} \\
&+ \beta_5 X_{4jk} + \xi_{ijk} \text{ (level 1)} \\
\beta_{0k} &= \gamma + u_k \text{ (level 2)} \\
\xi_{jk} &\overset{i.i.d}{\sim} N(0,\psi), u_k \overset{i.i.d}{\sim} N(0,\tau), \xi_{jk} \perp u_k \ \forall j,k
\end{aligned}
$$

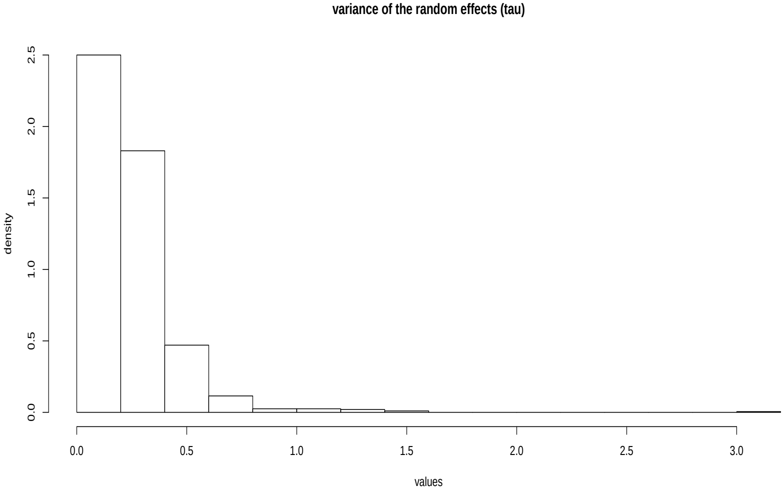
Inference

- (continuation of model) where $X_{1jk}, X_{2jk}, X_{3jk}, X_{4jk}$ where dummy variables indicating the female gender (the first) and the classroom $(X_{2jk}, X_{3jk}, X_{4jk})$, respectively.
- Priors: as presented in Application 2, additionally, $\beta_i \overset{i.i.d.}{\sim} N(0, 100), i = 1, 2, ..., 5$
- Model identification $a_1 = 1, b_1 = 0$.
- Show Bugs code (probit2PnormMultCov.r file).

Results


discrimination parameter
Bayesian estimates
0.5
1.0
1.5
5
10
15
20
item
difficulty parameter
Bayesian estimates
−4
−3
−2
−1
0
1
2
5
10
15
20
item
beta and gama
Bayesian estimates
1.0
1.5
2.0
2
4
6
8
beta − gama
beta1,...,beta5
Bayesian estimates
−0.4
0.0
0.2
0.4
0.6
1
2
3
4
5
beta1,....,beta5
psi_theta and tau
Bayesian estimates
0.1
0.2
0.3
0.4
0.5
0.6
0.7
psitheta
tau
item

Results


variance of the random effects (tau)
density
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
values

Application 4 : Multilevel IRT data with DIF

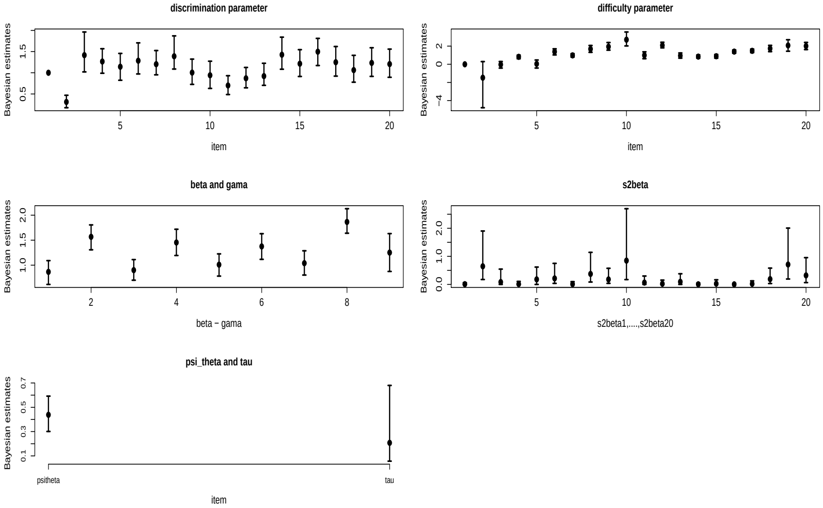
- Corresponds to the the data of Application 2 considering a possible effect of DIF on the difficulty parameter along the groups (schools).

$$\begin{aligned}
Y_{ijk}|(\theta_{jk},\zeta_i) &\overset{ind.}{\sim} \text{Bernoulli}(p_{ijk}) \\
p_{ijk} &= c_i + (1-c_i)\Phi(a_i\theta_{jk} - d_{ik}) \\
b_{ik} &= b_i + \beta_{ik} \\
\theta_{jk} &\overset{ind.}{\sim} N(\mu_{\theta_k},\psi_{\theta_k}) \\
\beta_{ik} &\overset{i.i.d.}{\sim} N(0,\sigma^2_{\beta_i})
\end{aligned}$$

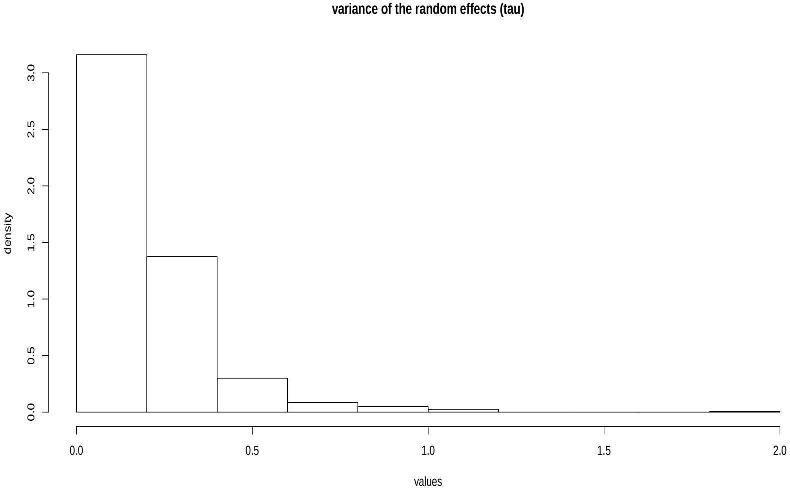
Inference

- Priors: as presented in Application 2, also in the preview slide, additionally, $\sigma^2_{\beta_i} \overset{i.i.d.}{\sim} \text{Ga}(0.1, 0.1), i = 1, 2, ..., 8$
- Model identification $a_1 = 1, b_1 = 0$.
- Show Bugs code (probit2PnormMultDIF.r file).

Results


discrimination parameter
Bayesian estimates
0.5
1.5
5
10
15
20
item
difficulty parameter
Bayesian estimates
−4
0
2
5
10
15
20
item
beta and gama
Bayesian estimates
1.0
1.5
2.0
2
4
6
8
beta − gama
s2beta
Bayesian estimates
0.0
1.0
2.0
5
10
15
20
s2beta1,....,s2beta20
psi_theta and tau
Bayesian estimates
0.1
0.3
0.5
0.7
psitheta
tau
item

Results


variance of the random effects (tau)
density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
values
0.0
0.5
1.0
1.5
2.0

Application 5 : longitudinal IRT 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.

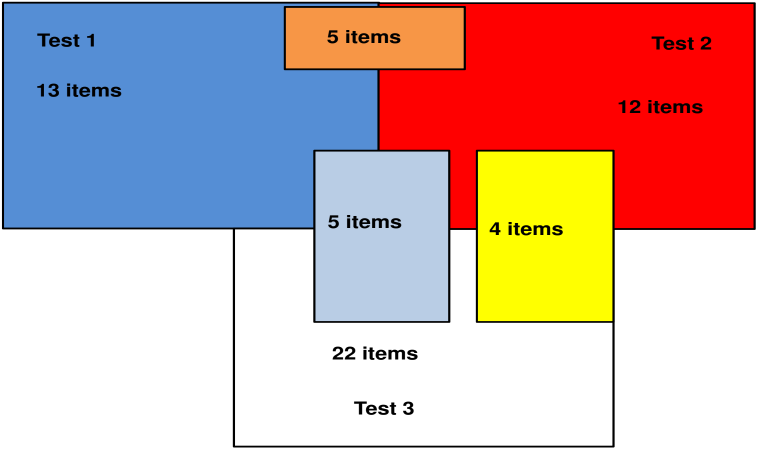
Application 5 : longitudinal IRT data

- The study was conducted from 1999 to 2003. At the start, 158 public schools were monitored, where 55 schools were selected for the program.
- Originally the students were followed along during six time-points (five grade schools - fourth to eighth). We have six tests, one for each time-point.
- One test was applied in the begin, other in the final of the first grade school, whereas the other tests were applied of the final of each grade schools.
- Other details can be found in Azevedo et al (2016).

Application 5 : longitudinal IRT data

- In the present study, Math's performances of 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


Test 1
13 items
5 items
Test 2
12 items
5 items
4 items
22 items
Test 3

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

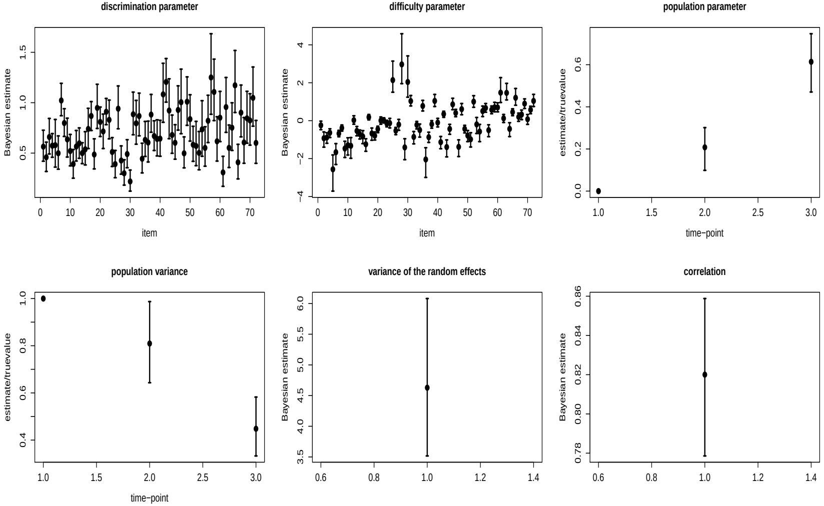
	Grade four	**Grade five**	**Grade six**
Grade four	1.000	**.723**	**.629**
Grade five	.659	1.152	**.681**
Grade six	.540	.641	1.071

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

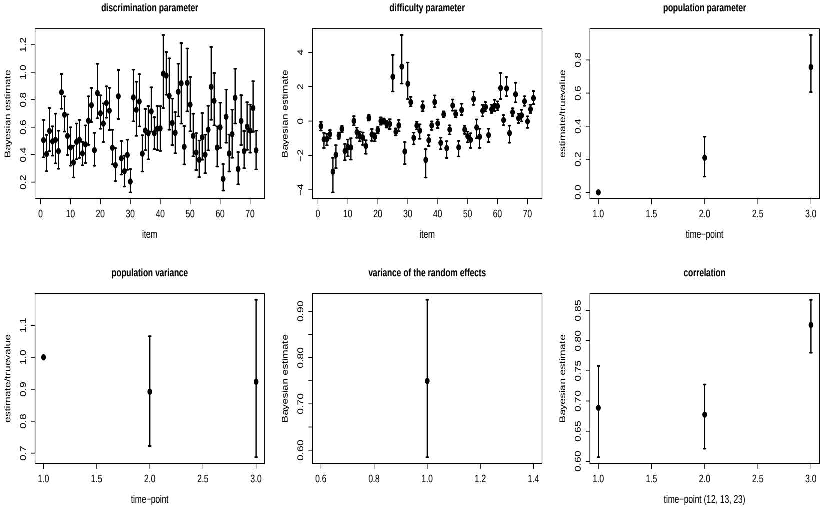
Inference

- We fitted two longitudinal IRT models: uniform (slide 52-53) and Hankel (slide 54-55).
- Priors: some of them were already defined in slides 52-55. Additionally $\mu_{\theta_1} = 0, \psi_{\theta_1} = 1$ (for model identification, reference time-point: 1), $a_i \overset{i.i.d.}{\sim}$ lognormal$(0, 0.25)$, $d_i \overset{i.i.d}{\sim} N(0, 4)$, $\mu_{\theta_t} \overset{ind.}{\sim} N(0, 10)$, $\psi_{\theta_t} \overset{ind.}{\sim}$ Ga$(0.1, 0.1)$ and $\sigma^2 \sim$ Ga$(0.1, 0.1)$.
- Show Bugs code (probit2PnormLongUnif.r and probit2PnormLongHankel.r files).

Results: uniform model



Results: Hankel model



discrimination parameter

difficulty parameter

population parameter

population variance

variance of the random effects

correlation