A multiple group Item Response Theory model with centred skew normal latent trait distributions under a Bayesian framework

Jose R. S. Santos[a], Caio L. N. Azevedo[b] & Heleno Bolfarine[c]

[a,b] *Department of Statistics, University of Campinas, Brazil,* [c] *Department of Statistics, University of São Paulo, Brazil*

Abstract

The multiple group IRT model (MGM) proposed by [4] provides a useful framework for analyzing item response data from clustered respondents. In the MGM, the selected groups of respondents are of specific interest such that group-specific population distributions need to be defined. An usual assumption for parameter estimation for this model, is to assume that the latent traits are random variables which follow possibly different symmetric normal distributions. However, many works suggest that this assumption does not apply in many cases. Furthermore, when this assumption does not hold, parameter estimates tend to be biased and misleading inference can result. Therefore, it is important to model the distribution of the latent traits properly. In this paper we present an alternative latent traits modeling, for multiple group framework, based on the so-called skew-normal distribution, see [27]. We name it SMGIRT model (skew multiple group IRT model). It extends the approach proposed by [1], [4] and [23] (concerning the latent trait distribution). We use the centred parameterization, which was proposed by [20]. This approach ensures model identifiability as pointed out by [10]. We propose and compare, concerning convergence issues, two MCMC algorithms for parameter estimation. A simulation study was performed in order to assess parameter recovery for the proposed model and the selected algorithm concerning convergence issues. The results reveals that our proposed algorithm recovers properly all model parameters. Furthermore, we analyzed a real data set which presents indication of asymmetry concerning the latent traits distribution. The results obtained by using our approach confirmed the presence of negative asymmetry of the latent traits distribution. More-

over, our model outperforms the usual symmetric normal MGM, leading to different conclusions concerning parameter estimation.

Keywords: Item Response Theory, centred skew-normal, bayesian estimation, model identifiability, multiple group model, MCMC algorithms

---

1. Introduction

In educational assessment, clinical trials, bioassays among other fields, it is common to observe examinees (subjects) from different groups. Commonly, groups can be characterized by gender, grade, social level, and so on. It is typically the case that group heterogeneity can reflect varying individual behavior. Therefore, it is important to take such heterogeneity into account. Attention will be focused on applications where the number of groups is limited and/or there is a specific interest in the sampled groups. The population distribution representing the clustered respondents completely specifies the distribution of respondents in each group, and no assumptions will be made about groups that are not selected. Then, inferences can be made with respect to the sampled groups but not at some higher levels of groups population.

[4] developed an IRT model where each group has a specific latent trait distribution. This multiple group model (MGM) has an additional set of parameters: multiple population parameters, which characterize the latent population distributions. A typical assumption for parameter estimation on this model, is to assume that the latent traits are random variables which follow a possibly different symmetric normal distributions (with possible different population parameters). However, this assumption can be unrealistic. For example, [5] shows that for many psychometric data sets the normality assumption of the latent traits distribution does not hold and it is considered only by convenience. The works of [8] and [9] also present data sets where the normality and even the symmetry assumptions are questionable. Such assumptions are also questioned by [6]. Another example can be found in [23]. Lack of normality of the latent traits distribution is related to the presence of at least one of the following characteristics: asymmetry, heavy tails and a kurtosis different from that of the normal distribution.

On the other hand, many works in the literature present evidence that lack of normality of the latent traits distribution can lead to biased estimates. One of the first works is [12], where it is shown, for dichotomous models,

that misspecification of the ability distribution can have significant impact in both item parameters and latent trait estimates. The marginal maximum likelihood (MML) method is considered (see [29]), to estimate the item parameters. Also, [13] obtained similar conclusions for these models, not only concerning departures from normality of the latent traits distribution, but also concerning lack of latent traits unidimensionality. They provided strong evidence that the absence that lack of normality for the latent traits can lead to biased item parameter estimates. Results obtained by [14], [15] and [16] for the nominal response models indicated that the latent traits distribution accounts for a reasonable percentage (around 40%) of the variability in the accuracy of item parameter estimates. Moreover, as outlined in [7], when the shape of the latent traits distribution (normal distribution, for example) is appropriate but the hyperparameters do not match, biased results can occur. Furthermore, even though suitable choices of prior distributions of the item parameters can attenuate the bias induced by the lack of some of the aforementioned assumptions, the impact of non-normality of the latent traits distribution is still significant, see [17], [18] and [19], for example.

The main goal of this work is to use a skew-normal distribution to model the latent traits in an IRT multipel group framework. We consider the centred parameterization (CP) of this distribution (see (4)), defined by [20], given that this parameterization ensures model identifiability, as established by [10]. Under this framework, it is possible to account for asymmetric behavior of the latent traits distribution in a straightforward way. In addition, we develop a MCMC framework to estimate all parameters concurrently based on an augmented data scheme with a Metropolis-Hastings within Gibbs sampling algorithm. Finally, some model fit assessment tools are considered to compare our model with the usual one (two parameter probit model with symmetric normal distribution for the latent traits). Our model extends the models proposed by [4], [1] and [23] (in terms of the latent trait distribution).

This article is organized as follows. In Section 1, we presented a literature review and the goals of this work. In Section two, we present the model and study aspects concerning its identifiability. In Section 3 two MCMC algorithms are developed to fit the model. In Section 4 we perform a simulation study and in Section 5 we conduct a real data analysis. Finally, in Section 6 we present some conclusions, comments and suggestions for future research. The main conclusion is that the model proposed can be useful in dealing with multiple groups IRT data.

2. The Multiple Group Model with centred skew normal latent trait distributions

In this paper, we deal with situations where one or more different tests are administered to the (randomly selected) examinees of each group. The tests have common items and the structure can be recognized as an incomplete block design, see [35]. We will assume that each group has a reasonable number of subjects. In summary, we are dealing with a set of n examinees clustered in K groups, with n_k examinees in group k, and $n = \sum_{k=1}^{K} n_k$. The examinees within group k answer I_k items, and $\sum_{k=1}^{K} I_K < I$, where I is the total number of items.

The following notation will be introduced: θ_{jk} is the latent trait of examinee j ($j = 1, \ldots, n_k,$) belonging to group k ($k = 1, \ldots, K$), $\boldsymbol{\theta}_{.k} = (\theta_{j1}, \ldots, \theta_{jK})^\top$ is the vector of latent traits of the examinees of group k, and $\boldsymbol{\theta}_{..} = (\boldsymbol{\theta}_{.1}, \ldots, \boldsymbol{\theta}_{.K})^\top$ is the vector with all latent traits; Y_{ijk} corresponds to the response of examinee j, of group k to item i ($i = 1, \ldots, I$), $\boldsymbol{Y}_{.jk} = (Y_{1jk}, \ldots, Y_{Ijk})^\top$ is the response vector of examinee j in group k, $\boldsymbol{Y}_{..k} = (\boldsymbol{Y}_{.1k}^\top, \ldots, \boldsymbol{Y}_{.n_k k}^\top)^\top$ is the response vector of all examinees of group k, $\boldsymbol{Y}_{...} = (\boldsymbol{Y}_{.1.}^\top, \ldots, \boldsymbol{Y}_{.n_k.}^\top)^\top$ is the whole response set and $(\boldsymbol{y}_{.1.}^\top, \ldots, \boldsymbol{y}_{.n_k.}^\top)^\top$ are the observed values; $\boldsymbol{\zeta}_i$ is the vector of parameters of item i, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^\top, \ldots, \boldsymbol{\zeta}_I^\top)^\top$ is the whole set of item parameters, $\boldsymbol{\eta}_{\theta_k}$ is the vector with the population parameters k and $\boldsymbol{\eta}_{\boldsymbol{\theta}} = (\boldsymbol{\eta}_{\theta_1}^\top, \ldots, \boldsymbol{\eta}_{\theta_K}^\top)^\top$ is the whole set of population parameters.

When modeling the grouping structure of subjects using group-specific normal population distributions for the latent traits, the SMGIRT model can be seen as a natural extension of the two-parameter (probit) item response multiple group model, which is represented by,

$$
\begin{aligned}
Y_{ijk} \mid (\theta_{jk}, \boldsymbol{\zeta}_i) &\sim \text{Bernoulli}(P_{ijk}) \\
P_{ijk} = P(Y_{ijk} = 1 \mid \theta_{jk}, \boldsymbol{\zeta}_i) &= \Phi(a_i \theta_{jk} - b_i) \qquad (1) \\
\theta_{jk} | \boldsymbol{\eta}_{\theta_k} &\sim \text{SN}_c(\mu_{\theta_k}, \sigma^2_{\theta_k}, \gamma_{\theta_k}), \qquad (2)
\end{aligned}
$$

where $\Phi(.)$ stands for the cumulative normal function. In this parametrization, the difficulty parameter $b_i = a_i b_i^*$ is a transformation of the commonly used difficulty parameter denoted as b_i^*. In addition, $\boldsymbol{\eta}_{\theta_k} = (\mu_{\theta_k}, \sigma^2_{\theta_k}, \gamma_{\theta_k})$, $SN_C(\mu_{\theta_k}, \sigma^2_{\theta_k}, \gamma_{\theta_k})$ stands for a skew normal distribution under the centred parameterization, with mean μ_{θ_k}, variance $\sigma^2_{\theta_k}$ and asymmetry coefficient γ_{θ_k}. For more details concerning the $SN_C(.,.,.,)$ distribution see [1], [2] and [3].

As in the the MGM with symmetric normal distribution, to ensure model identification, it suffices to assume that:

$$\begin{aligned} \theta_{j1} &\sim \mathrm{NA}_c(0,1,\gamma_{\theta_1}), \\ \theta_{jk} &\sim \mathrm{NA}_c(\mu_{\theta_k},\sigma^2_{\theta_k},\gamma_{\theta_k}), \quad k=2,\ldots,K, \end{aligned} \tag{3}$$

with a suitable linking design in terms of the administered tests, that is, the tests need to have some structure of common items. For more details see [3], [1] and [8].

Therefore the metric (scale) is defined and the model (1) is identified. This happens because model (1) is no longer invariant to location-scale transformations, given that the expected value and the variance of the latent distribution of the reference group (in this case, group 1) are fixed and also due to the linking design. This ensures that the metric for the latent traits is well defined. Furthermore, as stated in [22], [31], [21] and [20], the centred parameterization (CP) is more appropriate for inference purposes, model interpretation and parameter estimation than the direct parameterization (DP). More specifically, as mentioned in [21], page 588, the CP presents the following features: it removes the singularity of the information matrix at $\lambda_{\theta_k}=0$, the estimates obtained from CP are less correlated than those of DP and the likelihood shape is generally much improved. Other details can be found in [10].

On the other hand, it is possible to prove (see [3] and [10]), that the density of the skew-normal distribution under the CP is given by:

$$\begin{aligned} f(\theta_{jk}|\boldsymbol{\eta}_{\theta_k}) &= 2\frac{\sqrt{\sigma^2_{\theta_k}}}{\sigma}\phi\left[\frac{\sqrt{\sigma^2_{\theta_k}}}{\sigma}\left(y-\mu_{\theta_k}+\frac{\sigma_{\theta_k}}{\sqrt{\sigma^2_{\theta_k}}}\mu_{\theta_k}\right)\right] \\ &\times \Phi\left(\lambda\left[\frac{\sqrt{\sigma^2_{\theta_k}}}{\sigma_{\theta_k}}\left(y-\mu_{\theta_k}+\frac{\sigma_{\theta_k}}{\sqrt{\sigma^2_{\theta_k}}}\mu_{\theta_k}\right)\right]\right) \\ &= 2\omega^{-1}_{\theta_k}\phi\left(\omega^{-1}_{\theta_k}(y-\xi_{\theta_k})\right)\Phi\left[\lambda(\omega^{-1}_{\theta_k}(y-\xi_{\theta_k}))\right], \end{aligned} \tag{4}$$

which presents the stochastic representation due to Henze (1976) with parameters ξ, ω and λ where,

$$\xi = \mu - \frac{\sigma\mu_{\theta_k}}{\sigma_{\theta_k}} \qquad \text{and} \qquad \omega = \frac{\sigma_{\theta_k}}{\sigma_z}.$$

Finally, in terms of centred parameters $\mu_{\theta_k}, \sigma_{\theta_k}$ and γ_{θ_k} we have

$$\begin{aligned} \xi_{\theta_k} &= \mu_{\theta_k} - \sigma_{\theta_k}\gamma_{\theta_k}^{1/3}s, \\ \omega_{\theta_k} &= \sigma_{\theta_k}\sqrt{1+\gamma_{\theta_k}^{2/3}s^2}, \qquad (5) \\ \lambda_{\theta_k} &= \frac{\gamma_{\theta_k}^{1/3}s}{\sqrt{r^2+s^2\gamma_{\theta_k}^{2/3}(r^2-1)}} \text{where} \\ s &= (\frac{2}{4-\pi})^{1/3}. \qquad (6) \end{aligned}$$

It is also relevant to notice that the density given by (4) does not depend on the particular stochastic representation considered, either Sahu's or Henze's (see [26] and [24], respectively).

Figure 1 depicts some densities given by (4), for different values of λ_{θ_k} (γ_{θ_k}). The centred skew-normal distribution is parameterized by the skewness coefficient (γ_θ) instead of the asymmetry parameter (λ_θ). The parameter γ_θ has a straightforward interpretation, so that the closer it is to -1 or 1, the higher is the negative or the positive asymmetry of the latent traits distribution. Also, it is generally considered that if $\gamma_\theta \in (-0.13, 0.13)$ the asymmetry is not significant. Hence, it is clear that this limitation makes it simpler, for example, to define priors for γ_θ.

As a final comment, we can mention that in the literature, other item response functions (IRFs) such as the skew-probit, logit, or log-log are considered as well as other latent trait population distributions. To include this flexibility in response functions and latent trait distributions, a generalized MGM is defined as a mixture (indexed $l = 1, \ldots, L$) of different response functions based on different cumulative distribution functions (indexed $h = 1, \ldots, H$), and different latent trait distributions across items and groups of subjects. Then, the success probability is stated as

$$\begin{aligned} P_{ijk} = P(Y_{ijk} = 1 \mid \theta_{jk}, \boldsymbol{\zeta}_i, \boldsymbol{\nu}) &= \sum_{l=1}^{L}\prod_{h=1}^{H} F_{lh}\left(\theta_{jk}, \boldsymbol{\zeta}_i, \boldsymbol{\nu}\right) \qquad (7) \\ \theta_{jk} \mid \boldsymbol{\eta}_{\theta_k} &\sim D(\boldsymbol{\eta}_\theta), \qquad (8) \end{aligned}$$

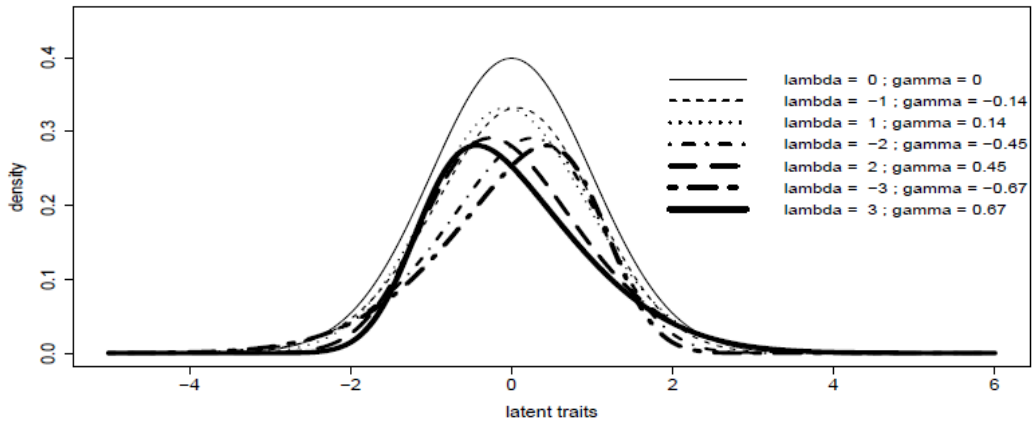

Figure 1: Skew-normal plots for different values of λ_θ (γ_θ).

where $F(.)$ represents a cumulative distribution function with parameters $\boldsymbol{\nu}$. $D(.)$ represents a continuous population distribution function, where $\boldsymbol{\eta}_\theta$ denotes the population parameters.

This modeling framework comprehends the well-known one-, two- and three-parameters item response models using probit, logit, log-log, Student's t, skew probit, or skew-probit link function, among others. In addition, the latent trait population distribution can be characterized by a normal, Student's t, skew normal, skew Student's t or finite mixture of normals, among others. Note also that extensions to nominal and ordinal response data can be made by defining a different response model at level 1 of the MGM. In the same way, the MGM for mixed response data will contain response models for discrete binary and polytomous response data.

Following, we present a Bayesian approach for parameter estimation.

3. Bayesian estimation and MCMC algorithms

Bayesian inference is based on the posterior distribution of the model parameters, which is proportional to the product of the likelihood and a prior distribution. Unfortunately, it is not possible to obtain the posterior distributions, analytically, for any IRM. However, MCMC algorithms provide empirical approximations for them, under some conditions, see [30] for example. An augmented data scheme is introduced to sample continuous normally

distributed item response data, denoted as $\mathbf{Z}$, given discrete observed item response data, denoted as $\mathbf{y}$. Following [28],

$$Z_{ijk} \mid (\theta_{jk}, \boldsymbol{\zeta}_i, Y_{ijk}) \quad \sim \quad N(a_i\theta_{jk} - b_i, 1), \tag{9}$$

where Y_{ijk} is the indicator of Z_{ijk} being greater than zero.

To handle an incomplete block design, an indicator variable $\mathbf{I}$ is defined that defines the set of administered items according to the design. For each administered item response the corresponding information is recorded. This indicator variable is described by,

$$I_{ijk} = \begin{cases} 1, & \text{if item i is administered for examinee } j \text{ of group } k \\ 0, & \text{if missing by design} \end{cases} \tag{10}$$

The indicator matrix $\mathbf{I}$ describes the patterns of missing data that is deliberately allowed to be missing. These missing data are missing by design.

In the same way an indicator variable can be defined to describe the missingness due to uncontrolled events as nonresponse or errors in recoding data. The missing data indicator variable is defined as,

$$V_{ijk} = \begin{cases} 1, & \text{if response of examinee } j \text{ of group } k \text{ on item i is recorded} \\ 0, & \text{if response is missing} \end{cases} \tag{11}$$

Indicator variables $\boldsymbol{V}_{...} = (V_{111}, \ldots, V_{In_KK})$ refer to observed data that could be missing.

In case of MAR, the missing indicator matrix, $\mathbf{V}$, and the augmented response data are conditionally independently distributed.

The object is to derive the conditional posterior density of $(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\boldsymbol{\theta}})$. It follows that the posterior distribution of $(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\boldsymbol{\theta}})$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, \boldsymbol{\delta} | \boldsymbol{z}_{...}, \boldsymbol{y}_{...}, \boldsymbol{\upsilon}_{...}) &\propto p(\boldsymbol{z}_{...}|\boldsymbol{y}_{...}, \boldsymbol{\upsilon}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, \boldsymbol{\delta}) p(\boldsymbol{\upsilon}_{...}|\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\delta}) \\ &= p(\boldsymbol{z}_{...}|\boldsymbol{y}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}) p(\boldsymbol{\upsilon}_{...}|\boldsymbol{\delta}) \\ &\propto p(\boldsymbol{z}_{...}|\boldsymbol{y}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}) \qquad (12) \\ &\propto \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} \exp\left\{-0,5(z_{ijk} - a_i\theta_{jk} + b_i)^2\right\} \\ &\times \mathbb{1}_{(Z_{ijk}, y_{ijk})}, \qquad (13) \end{aligned}$$

where $\boldsymbol{z}_{..} = (z_{11}, ..., z_{In})'$ and $\mathcal{I}_{jk}$ is the subset of items presented to (or answered by) examinee j from group k (they are related do the V_{ijk}, which

are known). On the other hand, the prior distribution of the parameters is assumed to be

$$p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) = \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(\theta_{jk} | \boldsymbol{\eta}_{\theta_k}) \right\} \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_\zeta) \right\} \left\{ \prod_{k=1}^{K} p(\boldsymbol{\eta}_{\theta_k} | \boldsymbol{\eta}_\eta) \right\}, \tag{14}$$

where, $\boldsymbol{\eta}_\zeta$ e $\boldsymbol{\eta}_\eta$ are the hyperparameters associated with $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}_\theta$, respectively. Moreover, we are assuming independency between items and population parameters.

The prior distribution of the latent trait given by (4) can be written as

$$p(\boldsymbol{\zeta}_i) \propto \exp\left[-0,5(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta)^\top \boldsymbol{\Psi}_\zeta{}^{-1} (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta)\right] 1\!\!1_{(a_i>0)}. \tag{15}$$

For the transformed population parameters, we consider the following prior density:

$$p(\xi_{\theta_k}, \omega_{\theta_k}, \lambda_{\theta_k}) = p(\xi_{\theta_k}, \omega_{\theta_k}) p(\lambda_{\theta_k}) \propto \frac{1}{\omega_{\theta_k}} \left(1 + \frac{\lambda_{\theta_k}^2}{d\varphi^2}\right)^{-\frac{d+1}{2}}, \tag{16}$$

where,

$$p(\xi_{\theta_k}, \omega_{\theta_k}) \propto \frac{1}{\omega_{\theta_k}}, \tag{17}$$

$$p(\lambda_{\theta_k}) \propto \left(1 + \frac{\lambda_{\theta_k}^2}{d\varphi^2}\right)^{-\frac{d+1}{2}}. \tag{18}$$

From equation (16) we can obtain an approximation to the Jeffreys prior (considering $d = \frac{1}{2}$ and $\varphi^2 = \frac{\pi^2}{4}$) and the prior induced by $\delta_{\theta_k} \sim U(-1, 1)$ (making, $d = 2$ e $\varphi^2 = \frac{1}{2}$). For simplicity, these two priors will be called Jeffreys prior and Uniform prior. Another choice is based on the following joint prior density:

$$p(\xi_{\theta_k}, \omega_{\theta_k}, \lambda_{\theta_k}) = p(\xi_{\theta_k}) p(\omega_{\theta_k}) p(\lambda_{\theta_k}), \tag{19}$$

considering,

$$\xi_{\theta_k} \sim N(\mu_{\xi_\theta}, \sigma_{\xi_\theta}^2), \tag{20}$$

$$\omega_{\theta_k}^2 \sim \mathrm{IG}(\alpha_{\omega_\theta}, \beta_{\omega_\theta}), \tag{21}$$

$$\lambda_{\theta_k} \sim N(\mu_{\lambda_\theta}, \sigma_{\lambda_\theta}^2). \tag{22}$$

where IG stands for the inverse-gamma distribution. In this way, considering the augmented likelihood (13) and the priors (15) and (16), we have that the posterior distribution of $(\boldsymbol{Z}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta)$ is given by:

$$
\begin{aligned}
p(\boldsymbol{Z}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{y}_{...}, \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) &\propto p(\boldsymbol{Z}_{...}|\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{y}_{...}) p(\boldsymbol{\theta}_{..}|\boldsymbol{\eta}_\theta) p(\boldsymbol{\zeta}|\boldsymbol{\eta}_\zeta) p(\boldsymbol{\eta}_\theta|\boldsymbol{\eta}_\eta) \\
&= \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} p(z_{ijk}|\theta_{jk}, \boldsymbol{\zeta}_i, y_{ijk}) \right\} \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(\theta_{jk}|\boldsymbol{\eta}_{\theta_k}) \right\} \\
&\times \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i) \right\} \left\{ \prod_{k=1}^{K} p(\xi_{\theta_k}, \omega_{\theta_k}, \lambda_{\theta_k}) \right\} \\
&\propto \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} \exp\left\{ -0,5(z_{ijk} - a_i\theta_{jk} + b_i)^2 \right\} 1\!\!1_{(Z_{ijk}, y_{ijk})} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \omega_{\theta_k}^{-n_k} \prod_{j=1}^{n_k} \phi\left(\frac{\theta_{jk} - \xi_{\theta_k}}{\omega_{\theta_k}} \right) \Phi\left[\left(\frac{\theta_{jk} - \xi_{\theta_k}}{\omega_{\theta_k}} \right) \right] \right\} \\
&\times \left\{ \prod_{i=1}^{I} \exp\left[-0,5(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta)^\top \boldsymbol{\Psi}_\zeta{}^{-1} (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta) \right] 1\!\!1_{(a_i > 0)} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \frac{1}{\omega_{\theta_k}} \left(1 + \frac{\lambda_{\theta_k}^2}{d\varphi^2} \right)^{-\frac{d+1}{2}} \right\}. \qquad (23)
\end{aligned}
$$

The posterior distribution, under the prior density (19), can be seen in Appendix equation (38). Due to the augmented data scheme, the full conditional distributions of the item parameters and of the augmented data themselves are known and easy to sample from. However, for the latent traits and population parameters conditionals are not simple to sample from. Therefore we need to use auxiliary algorithms to sample from such distributions. We consider the *Metropolis-Hastings* within Gibbs sampling algorithm (MHWGS), see [32] and [33]. However, we are using the augmented data likelihood instead of the original likelihood. Then, we call our algorithm ADMHWGS (Augmented data Metropolis-Hastings within Gibbs sampling), as in [1].

To implement the *Metropolis-Hastings* steps we need to consider suitable proposal densities, see [30], for both latent traits and population parameters. To the latent traits we consider:

$$J_t(\theta_{jk}^{(*)}|\theta_{jk}^{(t-1)}) = N(\theta_{jk}^{(t-1)}, \sigma_\theta^2), \tag{24}$$

where, $J_t(.)$ stands for the proposal density at the iteration t. For the population parameters we consider:

$$\begin{aligned} J_t(\xi_k^{(*)}|\xi_k^{(t-1)}) &= N(\xi_k^{(t-1)}, \sigma_0^2), & (25)\\ J_t(\omega_k^{(*)}|\omega_k^{(t-1)}) &= \text{Log-Normal}(\omega_k^{(t-1)}, \sigma_0^2), & (26)\\ J_t(\lambda_k^{(*)}|\lambda_k^{(t-1)}) &= N(\lambda_k^{(t-1)}, \sigma_0^2). & (27) \end{aligned}$$

Denoting (.) the set of all other parameters, the proposed algorithm (ADMHWGS), for $t = 1, 2, ..., B, ..., M$, where B is the burn-in and M is the generated sample size, simulates iteratively all unknown quantities in the following order:

- Start the algorithm by choosing suitable starting values;
- Simulate Z_{ijk} from $Z_{ijk}|(.)$, $k = 1, \ldots, K$, $j = 1, \ldots, n_k$ and $i \in I_{jk}$;
- Simulate θ_{jk} from $\theta_{jk}|(.)$, $k = 1, \ldots, K$ e $j = 1, \ldots, n_k$;
- Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(.)$, $i \in I_{jk}$;
- Simulate ξ_{θ_k} from $\xi_{\theta_k}|(.)$, $k = 1, \ldots, K$;
- Simulate ω_{θ_k} from $\omega_{\theta_k}|(.)$, $k = 1, \ldots, K$;
- Simulate λ_{θ_k} from $\lambda_{\theta_k}|(.)$, $k = 1, \ldots, K$.

On the hand, we can consider the usual hierarchical representation of the centred skew normal distribution, (see [1]), that is:

$$\theta_{jk} = \xi_{\theta_k} + \omega_{\theta_k}\left(\delta_{\theta_k} H_{jk} + \sqrt{1-\delta_{\theta_k}^2} Q_{jk}\right), \tag{28}$$

where, $H_{jk} \sim \text{HN}(0,1)$, $Q_{jk} \sim N(0,1)$, $H_{jk} \perp Q_{jk}$, $\forall j, k$. We can also rewrite (28) as:

$$\begin{aligned} \theta_{jk}|(h_{jk}, \boldsymbol{\eta}_{\theta_k}) &\sim N(\xi_{\theta_k} + \tau_{\theta_k} h_{jk}, \varsigma_{\theta_k}^2), & (29)\\ H_{jk} &\sim \text{HN}(0,1), & (30) \end{aligned}$$

where,

$$\begin{aligned}\tau_{\theta_k} &= \omega_{\theta_k}\delta_{\theta_k},\\ \varsigma^2_{\theta_k} &= \omega^2_{\theta_k}(1-\delta^2_{\theta_k}).\end{aligned}$$

Considering now the population parameters, we can notice that the prior (18) can be rewritten as follows:

$$\lambda_{\theta_k}|t_k \sim N(0,\frac{\varphi^2}{t_k}), \tag{31}$$

$$T_k \sim \text{gamma}(\frac{d}{2},\frac{d}{2}). \tag{32}$$

Thus, the prior (16) can be rewritten as,

$$p(\lambda_{\theta_k},\xi_{\theta_k},\omega_{\theta_k},t_k) \propto \frac{1}{\omega_{\theta_k}}\exp\left(-\frac{1}{2}\frac{\lambda^2_{\theta_k}t_k}{\varphi^2}\right)t_k^{\frac{d+1}{2}-1}\exp\left(-\frac{d}{2}t_k\right). \tag{33}$$

Considering the aforementioned reparametrization, we have,

$$p(\xi_{\theta_k},\tau_{\theta_k},\varsigma_{\theta_k},t_k) \propto \frac{1}{\omega_{\theta_k}}\exp\left(-\frac{1}{2}\frac{\tau^2_{\theta_k}t_k}{\varsigma^2_{\theta_k}\varphi^2}\right)t_k^{\frac{d+1}{2}-1}\exp\left(-\frac{d}{2}t_k\right). \tag{34}$$

Equation (34) is a joint priori for the parameters $(\xi_{\theta_k},\tau_{\theta_k},\varsigma_{\theta_k})$ and the latent variables T_k. On the other hand, we may notice that equation (29) can be viewed as a regression model with the response variable θ_{jk}, intercept ξ_{θ_k} and slope τ_{θ_k}. Therefore, analogously to what was made for the item parameters, we can consider a bivariate normal distribution as prior distribution for the vector $(\xi_{\theta_k},\tau_{\theta_k})^\top$, which we will denote by $\boldsymbol{\beta}_{\theta_k}$, and an inverse gamma prior for parameter $\varsigma^2_{\theta_k}$. That is,

$$\boldsymbol{\beta}_{\theta_k} \sim N\left(\boldsymbol{\mu}_\theta,\Sigma_\theta\right), \tag{35}$$

$$\varsigma^2_{\theta_k} \sim \text{IG}(\alpha_{\varsigma_\theta},\beta_{\varsigma_\theta}), \tag{36}$$

where,

$$\boldsymbol{\beta}_{\theta_k} = \begin{pmatrix} \xi_{\theta_k} \\ \tau_{\theta_k} \end{pmatrix}, \quad \boldsymbol{\mu}_\theta = \begin{pmatrix} \mu_{\xi_\theta} \\ \mu_{\tau_\theta} \end{pmatrix} \quad \text{e} \quad \Sigma_\theta = \begin{pmatrix} \sigma^2_{\xi_\theta} & \rho\sigma^2_{\xi_\theta}\sigma^2_{\tau_\theta} \\ \rho\sigma^2_{\xi_\theta}\sigma^2_{\tau_\theta} & \sigma^2_{\tau_\theta} \end{pmatrix}.$$

We will assume the following structure for the joint prior distribution of the parameters:

$$\begin{aligned} p(\boldsymbol{\theta}_{..}, \boldsymbol{h}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, \boldsymbol{t}_{.}|\boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) &= \left\{\prod_{k=1}^{K}\prod_{j=1}^{n_k} p(\theta_{jk}|\boldsymbol{\eta}_{\theta_k})\right\}\left\{\prod_{k=1}^{K}\prod_{j=1}^{n_k} p(h_{jk})\right\} \\ &\times \left\{\prod_{i=1}^{I} p(\boldsymbol{\zeta}_i|\boldsymbol{\eta}_\zeta)\right\}\left\{\prod_{k=1}^{K} p(\boldsymbol{\eta}_{\theta_k}|\boldsymbol{\eta}_\eta)\right\} \\ &\times \left\{\prod_{k=1}^{K} p(t_k)\right\}, \end{aligned} \tag{37}$$

where, $\boldsymbol{h}_{..} = (\boldsymbol{h}_{1.}, \ldots, \boldsymbol{h}_{n_k.})^\top$ and $\boldsymbol{t}_{.} = (t_1, \ldots, t_K)^\top$. The prior distributions to the latent traits and the variables $\boldsymbol{h}$ are given by (29) and (30), respectively. To the item parameters, we will consider the prior (15) as before. For the population parameters, we can consider either the joint prior (34) or the prior distributions (35) and (36). Depending on the prior distribution chosen the posterior distribution will be different. They are given in Appendix (equations (39) and (40)). It is important to observe that, the hierarchical representation makes it possible to obtain the complete conditional distributions with known forms, to all parameters. Therefore, it is possible the using of the full Gibbs sampling algorithm. This algorithm is called ADGS (augmented data Gibbs sampling)

For each one of the two posterior distributions we will have the following algorithm:

- Start the algorithm by choosing suitable initial values;
- Simulate Z_{ijk} from $Z_{ijk}|(.)$, $k = 1, \ldots, K$, $j = 1, \ldots, n_k$ and $i \in I_{jk}$;
- Simulate h_{jk} from $H_{jk}|(.)$, $k = 1, \ldots, K$ and $j = 1, \ldots, n_k$;
- Simulate θ_{jk} from $\theta_{jk}|(.)$, $k = 1, \ldots, K$ and $j = 1, \ldots, n_k$;
- Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(.)$, $i \in I_{jk}$;

- Simulate t_k from $T_k|(.)$, $k = 1, \ldots, K$;
- Simulate ξ_{θ_k} from $\xi_{\theta_k}|(.)$, $k = 1, \ldots, K$;
- Simulate τ_{θ_k} from $\tau_{\theta_k}|(.)$, $k = 1, \ldots, K$;
- Simulate $\varsigma^2_{\theta_k}$ from $\varsigma^2_{\theta_k}|(.)$, $k = 1, \ldots, K$;

if we consider the posterior distribution (39). On the other hand, if we consider the posterior distribution (40), we have the following algorithm:

- Start the algorithm by choosing suitable initial values;
- Simulate Z_{ijk} from $Z_{ijk}|(.)$, $k = 1, \ldots, K$, $j = 1, \ldots, n_k$ and $i \in I_{jk}$;
- Simulate h_{jk} from $H_{jk}|(.)$, $k = 1, \ldots, K$ and $j = 1, \ldots, n_k$;
- Simulate θ_{jk} from $\theta_{jk}|(.)$, $k = 1, \ldots, K$ and $j = 1, \ldots, n_k$;
- Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(.)$, $i \in I_{jk}$;
- Simulate $\boldsymbol{\beta}_{\theta_k}$ from $\boldsymbol{\beta}_{\theta_k}|(.)$, $k = 1, \ldots, K$;
- Simulate $\varsigma^2_{\theta_k}$ from $\varsigma^2_{\theta_k}|(.)$, $k = 1, \ldots, K$.

For more details see the Appendix.

4. Simulation study

In this section we conducted a simulation study in order to asses convergence issues and the accuracy of parameter estimates based on the two proposed algorithms. First, we discuss convergence aspects and, after that, we will evaluate the accuracy of the parameter estimation for the MCMC algorithm with better convergence proprieties. To accomplish for that we simulated a set of response based on the following structure:

- test 1: 20 items;
- test 2: test 1 + 20 other items;
- test 3: the last 20 items of test 2 + 20 other items;
- test 4: the last 20 items of test 3 + 20 other items.

The item parameters were fixed in the following intervals: $a_i \in [0.7, 1.4]$ and $b_i^* \in [-2, 4]$. The latent traits were simulated from independent skew normal distributions under CP with means $\boldsymbol{\mu}_\theta = (0, 1, 1.4, 2)^\top$, standard deviation $\boldsymbol{\sigma}_\theta = (1, 0.88, 0.62, 0.77)^\top$ and asymmetry coefficients $\boldsymbol{\gamma}_\theta = (0, 0.14, 0.5, -0.5)^\top$ for groups 1,2,3 and 4, respectively. The number of subjects per group were $(n_1 = 556, n_2 = 556, n_3 = 401, n_4 = 294)$. We tried to mimic the results obtained by [8], once that will analyze the same data set. The values of the asymmetry parameter were chosen in order to consider symmetry, weak asymmetry, left strong asymmetry and right strong asymmetry.
For the item parameters we considered $(1, 0)$ as a vector mean and $\text{diag}(0.5, 9)$ as the covariance matrix. For the population parameters we considered the Jeffreys prior. For this simulated data set, we generated three chains, based on different set of starting values, for each one of the algorithms. The Gelman-Rubin statistics ranged from 1.00 to 1.02 for both algorithms. Also, an inspection of trace plots indicated that the chains mixed very well for both algorithms, see Figures 2 and 3. Also, the correlograms (not shown) revealed that the samples produced by storing values at every 30-th iteration is enough to produce samples with negligible autocorrelations.

Furthermore, we compared the two algorithms concerning the effective sample size (ESS), see [25]. Table 1 presents the values of the ESS and ESS per minute. We can see that, in a general way, the AGDS algorithm presented better results than the ADMHWGS algorithm.

Table 1: ESS and ESS per minute for the two algorithms

	ADMHWGS		ADGS		
Prior	ESS	ESS/m	ESS	ESS/m	Ratio
1	4474.178	25.228	4771.270	36.361	1.441
2	4549.746	20.190	4724.984	36.014	1.784
3	5054.523	32.316	4955.711	35.533	1.010
4	4569.048	27.968	5103.164	23.491	0.840

Concerning parameter recovery we generated other R = 10 replicas (data sets), for each one of the situation defined by crossing the levels of different factors. These factors were (with the levels within parenthesis): number of examinees per group (NE) (1500, 3000), number of items per group (NI) (20, 40), number of common items (NCI) (25%, 50%), prior (P) (prior1, prior2, prior3, prior4). The priors 1 to 4 are, in fact, sets of priors. For simplicity,

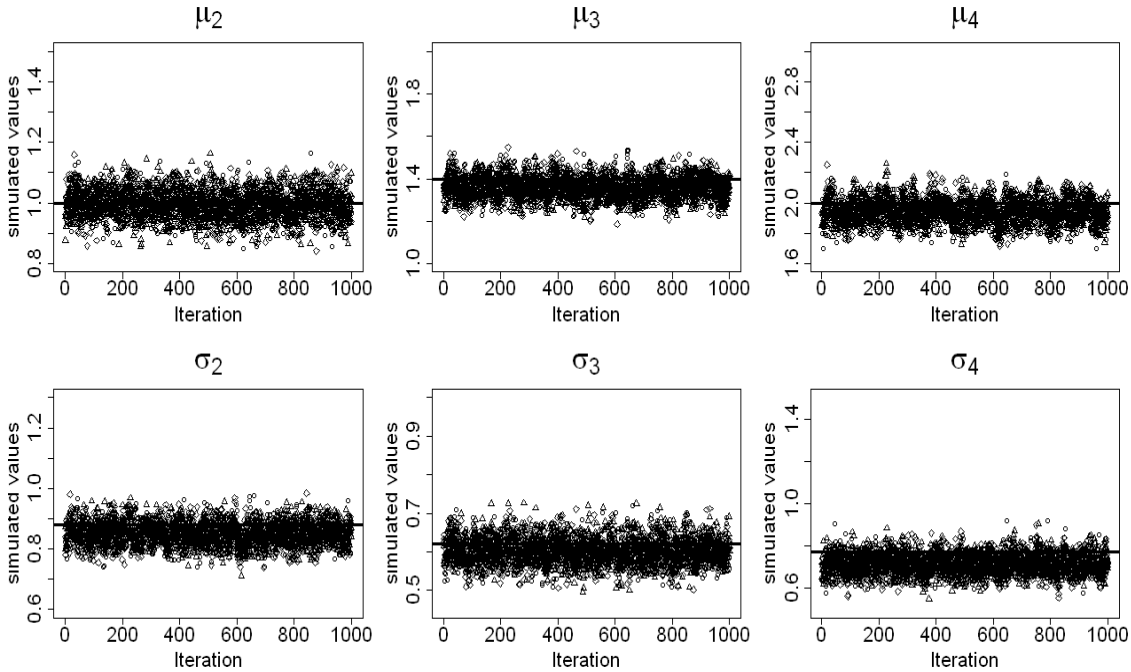

Figure 2: Trace plots for the population parameters: ADMHWGS algorithm. Legend: ($\circ$ set 1, $\triangle$ set 2, $\diamond$ set 3, — true value)

these will be named priors only (without mention that they are sets of priors explicitly). These sets differ only in terms of the priors for the population parameters. That is, the prior 1 and 2 are equivalent to the Jeffreys prior and uniform prior for the population parameters, respectively. On the other hand, the priors 3 and 4 are the conjugate priors, see equations (35) and (36), respectively. In the last case, were fixed hyperparameters in order to obtain ($\mathbb{E}(\mu_{\theta_k}) = \mathbb{E}(\gamma_{\theta_k}) = 0$ and $\mathbb{E}(\sigma_{\theta_k}) = 1$) and ($\mathbb{V}\text{ar}(\mu_{\theta_k}) = \mathbb{V}\text{ar}(\sigma_{\theta_k}) = \mathbb{V}\text{ar}(\gamma_{\theta_k}) = 1$), considering the prior 3, and same averages with ($\mathbb{V}\text{ar}(\mu_{\theta_k}) = \mathbb{V}\text{ar}(\sigma_{\theta_k}) = \mathbb{V}\text{ar}(\gamma_{\theta_k}) = 10$), considering the prior 4. In addition, we defined another factor called asymmetry of the reference group (ARG) (0; 0,6 e -0,6). The other population parameters were fixed as follows: the means and the standard deviations were fixed in $\boldsymbol{\mu}_\theta = (0, -1, 1)^\top$ and $\boldsymbol{\sigma}_\theta = (1, 0.8, 1.2)^\top$ for the groups 1, 2 e 3, respectively. The asymmetry coefficients for the groups 2 and 3 were fixed in $\gamma_{\theta_2} = 0.6$ and $\gamma_{\theta_3} = -0.6$. Table 2 presents values for the population parameters chosen for each group in each level of ARG factor.

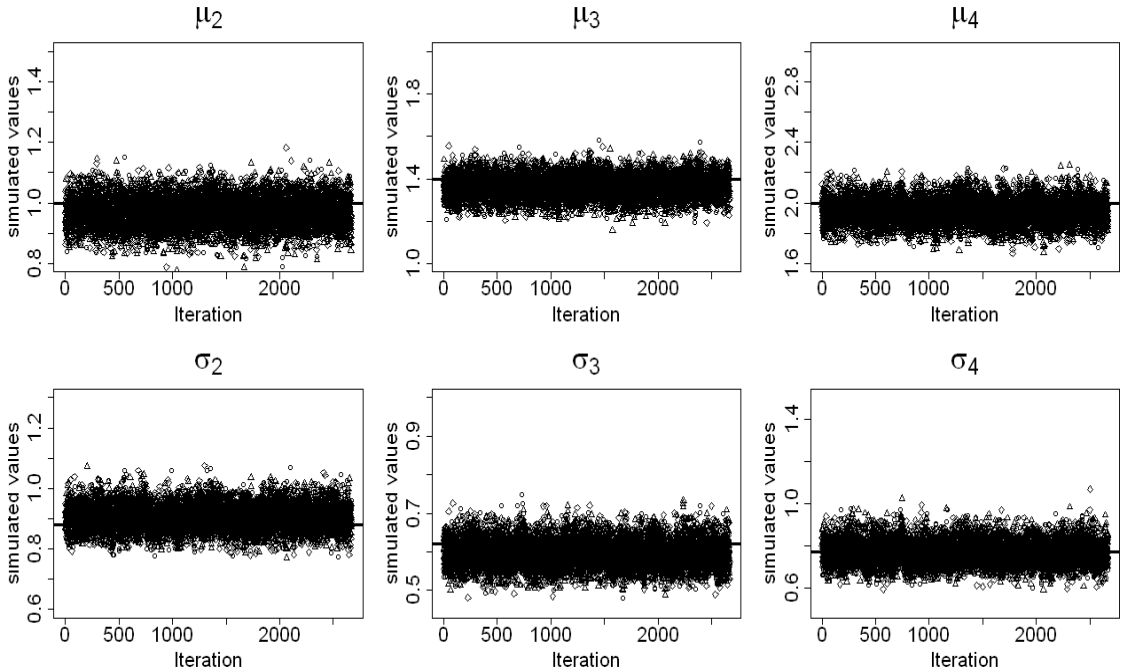

Figure 3: Trace plots for the population parameters: ADGS algorithm. Legend: (○ set 1, △ set 2, ⋄ set 3, — true value)

Table 2: Population parameters chosen for each group and each level of the ARG factor

		ARG= 1	ARG= 2	ARG= 3
Group 1	μ_{θ_1}	0	0	0
	σ_{θ_1}	1	1	1
	γ_{θ_1}	0	0.6	-0.6
Group 2	μ_{θ_2}	-1	-1	-1
	σ_{θ_2}	0.8	0.8	0.8
	γ_{θ_2}	0.6	0.6	0.6
Group 3	μ_{θ_3}	1	1	1
	σ_{θ_3}	1.2	1.2	1.2
	γ_{θ_3}	-0.6	-0.6	-0.6

The values of the item parameters were chosen considering the values of the population parameters and in order to have items with different discrimination powers and different difficulty levels. We chose less groups than in

the convergence assessment study related to the time necessary to run all the replicas. We considered the RMSE (square root of the mean square error) and the AVRB (average value of the relative bias) to measure the accuracy of the estimates, based on the estimates obtained with the 10 replicas. In addition, we considered only the ADGS algorithm (due to the results of the convergence assessment study).

From an inspection of the Figures 4 to 9 we can conclude that the ADGS algorithm recovered all parameters properly. In addition, one can see that there is a slight difference between the results obtained from the different priors. This difference is much higher for the asymmetry coefficient and in this case the more accurate results were obtained by using the Jeffreys prior. Moreover, there is a strong indication that the higher are NI and NCI values (simultaneously) the more accurate are the results.

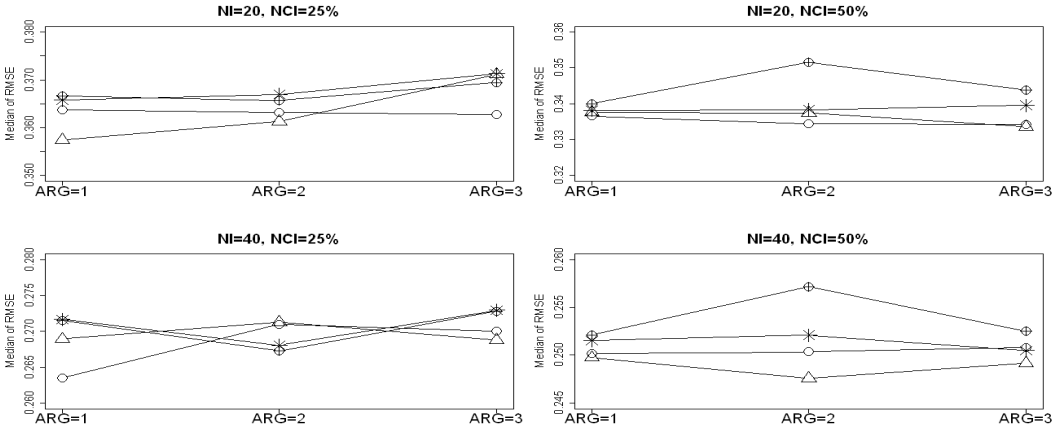

Figure 4: Median of RMSE for the latent traits estimates. Legend: (−○− Prior 1, −△− Prior 2, −∗− Prior 3 and −⊕− Prior 4)

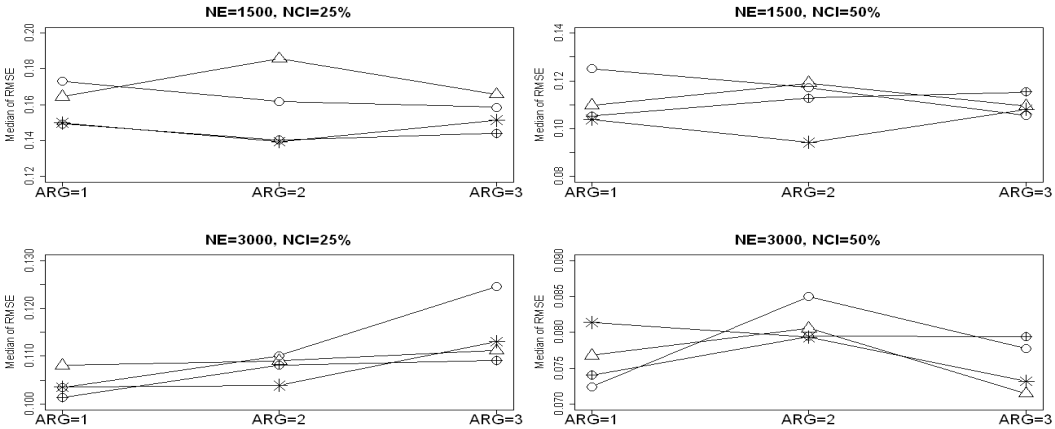

Figure 5: Median of RMSE for the discrimination parameters. Legend: (–○– Prior 1, –△– Prior 2, –∗– Prior 3 and –⊕– Prior 4)

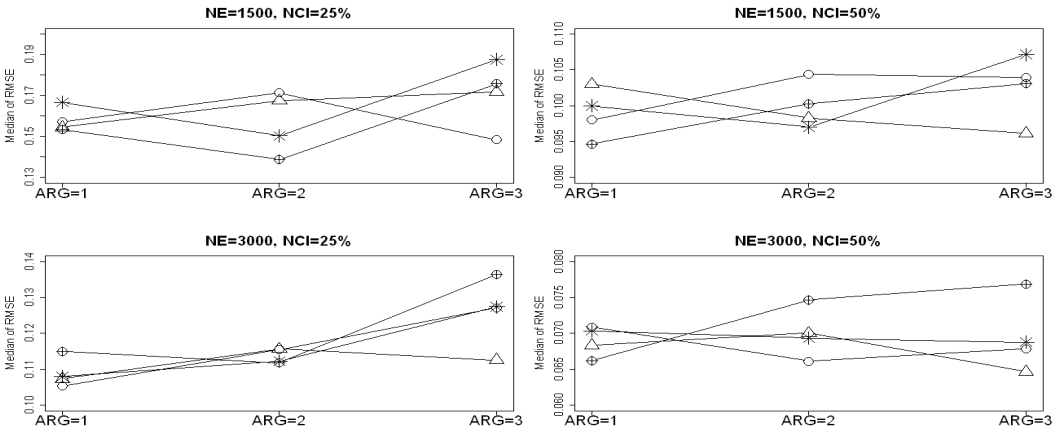

Figure 6: Median of RMSE for the difficulty parameters. Legend: (–○– Prior 1, –△– Prior 2, –∗– Prior 3 and –⊕– Prior 4)

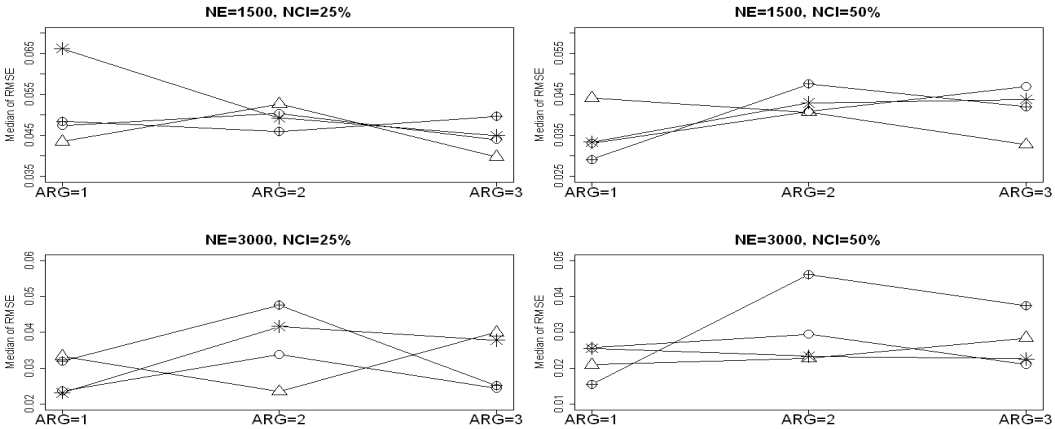

Figure 7: Median of RMSE for the population means. Legend: (–○– Prior 1, –△– Prior 2, –∗– Prior 3 and –⊕– Prior 4)

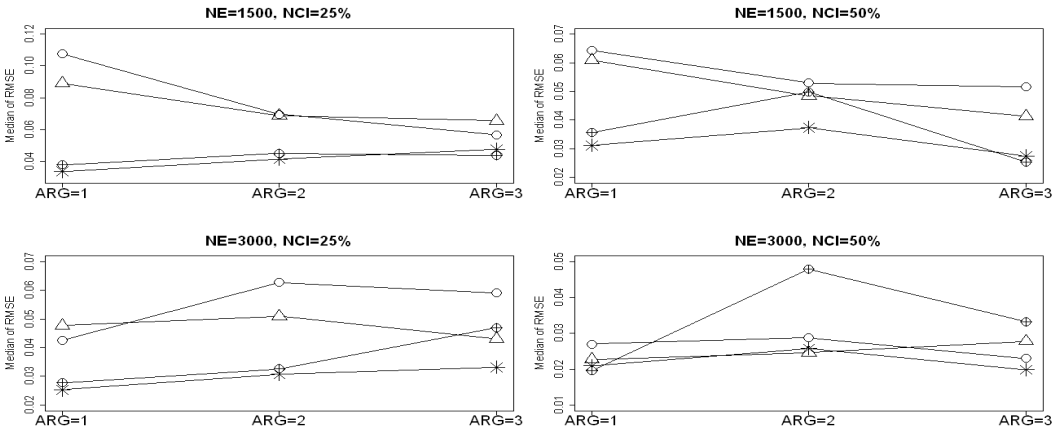

Figure 8: Median of RMSE for the population standard deviations. Legend: (–○– Prior 1, –△– Prior 2, –∗– Prior 3 and –⊕– Prior 4)

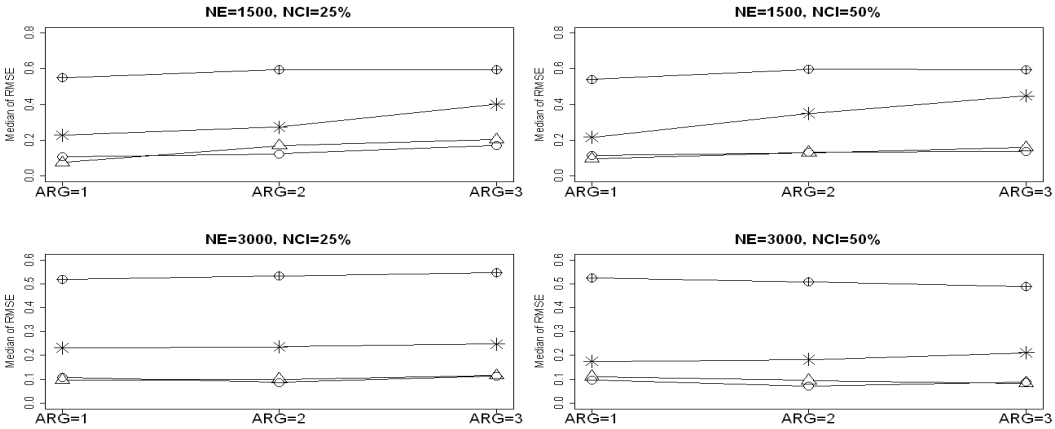

Figure 9: Median of RMSE for the population asymmetry coefficients. Legend: (−○− Prior 1, −△− Prior 2, −∗− Prior 3 and −⊕− Prior 4)

5. Real data analysis

Our data set is part of a major longitudinal educational assessment coordinated by the University of Exeter, England, regarding to Mathematics achievement. It consists of a sample of 557 second grade students from eight public schools of the city of Londrina, southern of Brazil. They were submitted to a test of 40 open items that were corrected as right/wrong. More details can be obtained in [36]. The whole data set was previously analyzed by [8] using the symmetric normal MGM. We compared those results with the ones obtained by using the model proposed in this paper, estimated by using the ADGS algorithm with the Jeffreys prior. We considered the model fit assessment tools described in [8] and for model comparison we used the statistics considered in [1].

From Figures 10 to 19 and Tables 3 to 4, we can see that, in general, the results from the two models were similar. However, for the second group, which presented the highest absolute value for the asymmetry coefficient (-0.805), the results were significantly different. For example, in Figure 16 we can see that there is a considerable difference between the estimates of the latent traits of the group 2. Figures 18 and 19 show the smoothed histogram of the latent trait estimates (we called estimated curves) and the densities of skew-normal by considering the estimates of the population parameters

(we called theoretical curves). These Figures suggest that the skew model models the observed latent trait distributions more properly than the symmetric model. Also, the values of the Bayesian p-values for both Pearson and deveiance residuals discrepancy measures, showed in Figure 13, indicate the two models fit all almost items properly. However, Item 22 were not well fitted to the data under the symmetric model, occuring the opposite with the skew model. In addition, the skew model tends to present higher estimates for the discrimination parameter comparing with the symmetric model. That is, under the skew model, the items discriminate examinees with different latent trait values, that they do under the symmetric model.Finally, according to Table 4, all statistics indicated that our model fitted better the data set. Therefore, inferences based on our model are more reliable than the ones obtained by using the symmetric normal model.

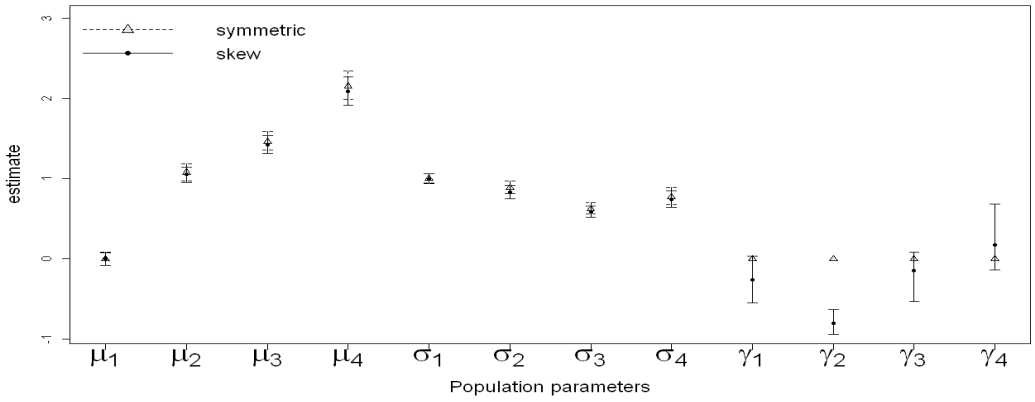

Figure 10: Population parameter estimates and 95% credibility intervals (skew and symmetric models)

Table 3: Estimates and credibility intervals for the population parameters (skew and symmetric models)

	Skew						Symmetric					
Parameter	EAP	PSD	CI (95%)		HPD (95%)		EAP	PSD	CI (95%)		HPD (95%)	
μ_{θ_1}	0	-	-	-	-	-	0	-	-	-	-	-
μ_{θ_2}	1.049	0.050	0.956	1.150	0.958	1.153	1.078	0.054	0.974	1.186	0.981	1.193
μ_{θ_3}	1.427	0.057	1.317	1.539	1.316	1.536	1.463	0.060	1.356	1.588	1.356	1.589
μ_{θ_4}	2.093	0.088	1.926	2.270	1.917	2.258	2.156	0.091	1.986	2.346	1.996	2.354
σ_{θ_1}	1	-	-	-	-	-	1	-	-	-	-	-
σ_{θ_2}	0.828	0.045	0.740	0.917	0.750	0.923	0.887	0.042	0.811	0.970	0.810	0.967
σ_{θ_3}	0.588	0.036	0.520	0.659	0.519	0.659	0.623	0.037	0.556	0.703	0.555	0.702
σ_{θ_4}	0.744	0.055	0.642	0.869	0.636	0.853	0.772	0.056	0.673	0.892	0.679	0.895
γ_{θ_1}	-0.283	0.153	-0.587	-0.001	-0.552	0.001	0.000	-	-	-	-	-
γ_{θ_2}	-0.805	0.092	-0.960	-0.605	-0.975	-0.630	0.000	-	-	-	-	-
γ_{θ_3}	-0.160	0.160	-0.484	0.045	-0.448	0.060	0.000	-	-	-	-	-
γ_{θ_4}	0.140	0.184	-0.079	0.569	-0.098	0.537	0.000	-	-	-	-	-

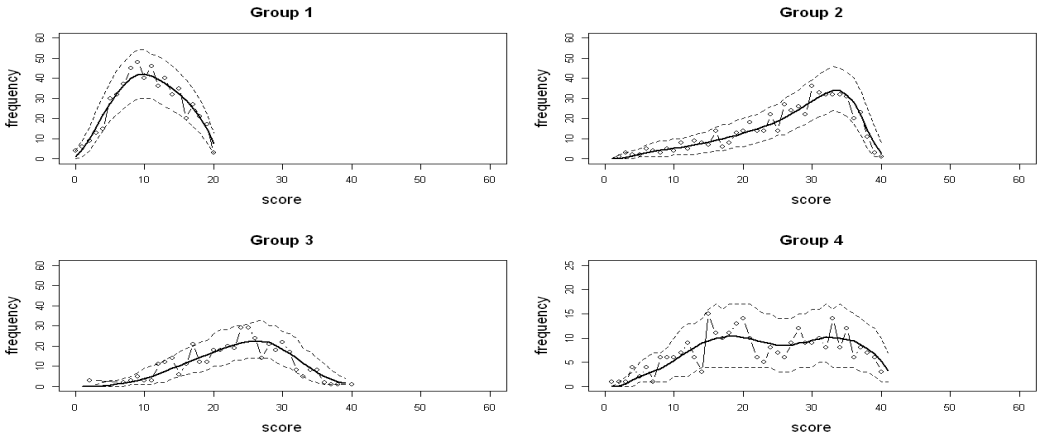

Figure 11: Observed and predictive scores distribution (skew model). Legend: solid line: predict scores, –o– observed scores, – – – 95% credibility interval

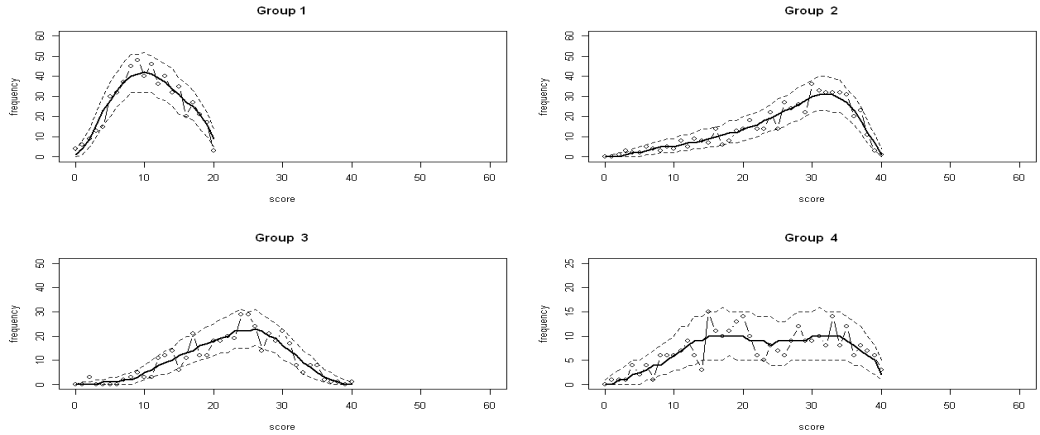

Figure 12: Observed and predictive scores distribution (skew model). Legend: solid line: predict scores, –o– observed scores, – – – 95% credibility interval

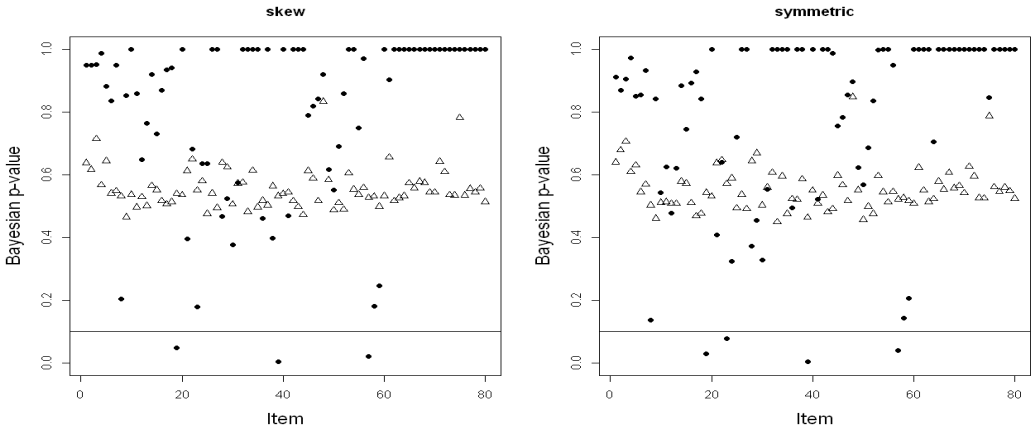

Figure 13: Bayesian p-values for the discrepancy measure based on the Pearson and deviance residuals. Legend: (• Pearson, △ deviance)

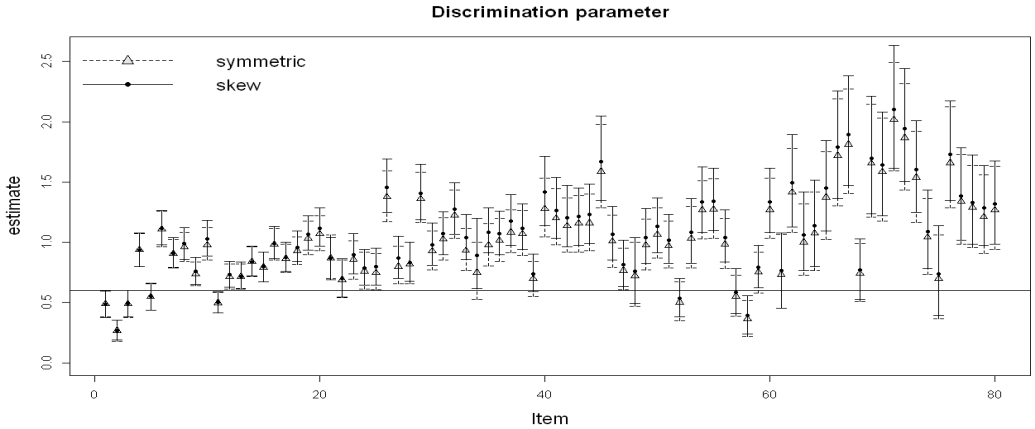

Figure 14: Discrimination parameter estimates and 95% credibility intervals (skew and symmetric models)

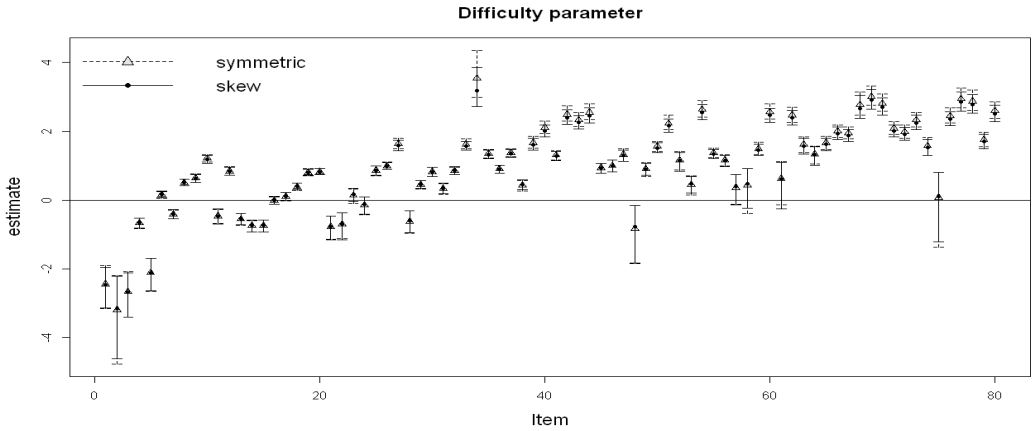

Figure 15: Difficulty parameter estimates and 95% credibility intervals (skew and symmetric models)

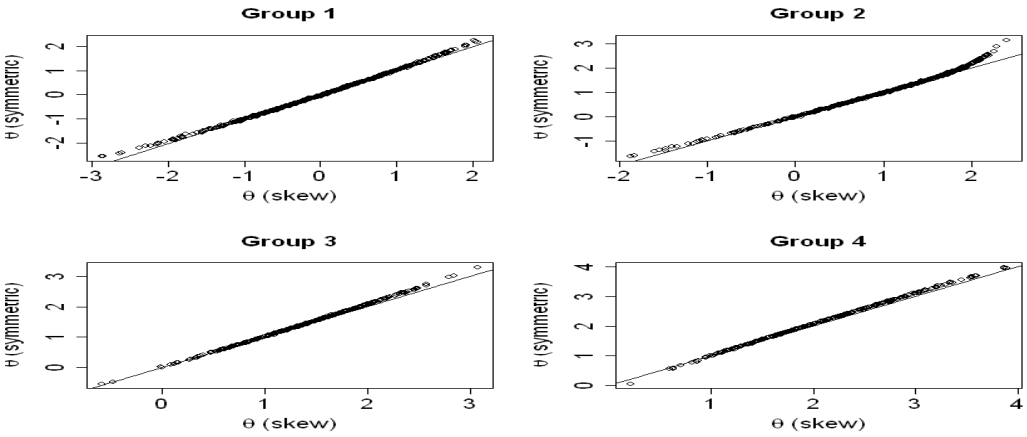

Figure 16: Scatter plot for the latent trait estimates (skew and symmetric models)

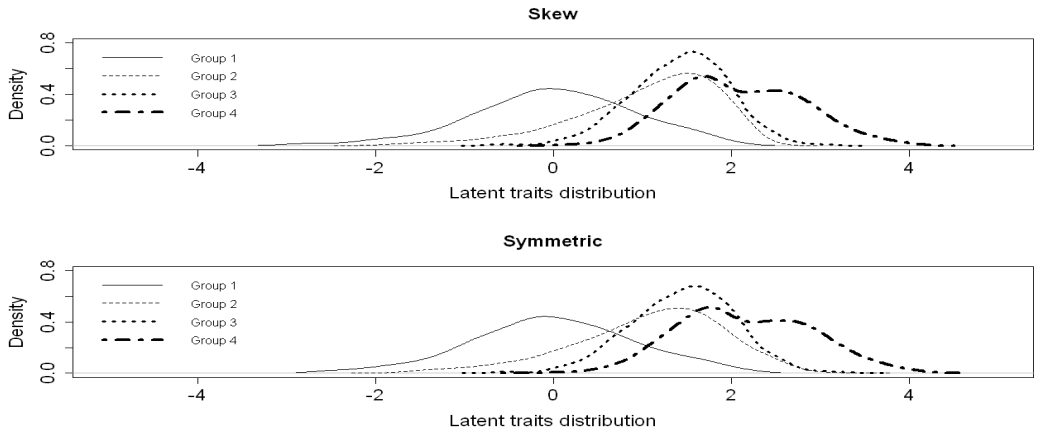

Figure 17: Latent scores distributions per group on a common scale (skew and symmetric models)

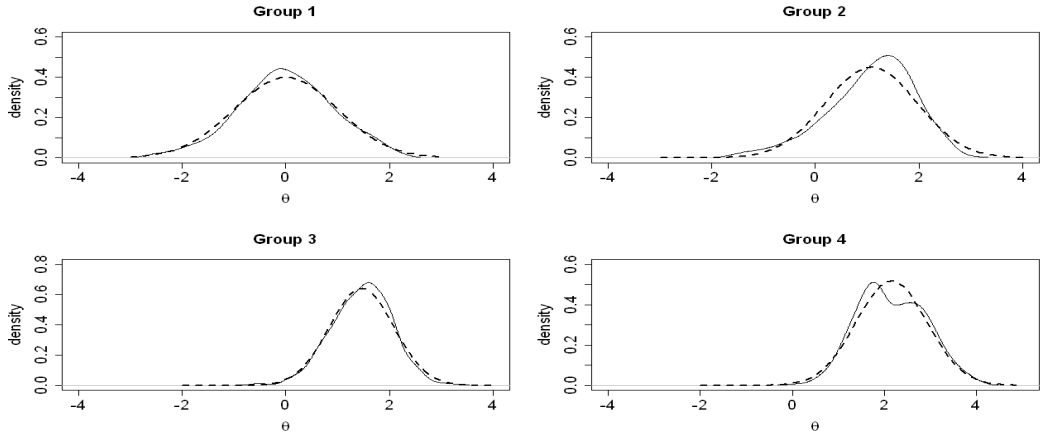

Figure 18: Latent scores distributions with theoretical curves (symmetric model). Legend: (— estimated curve, − − − theoretical curve)

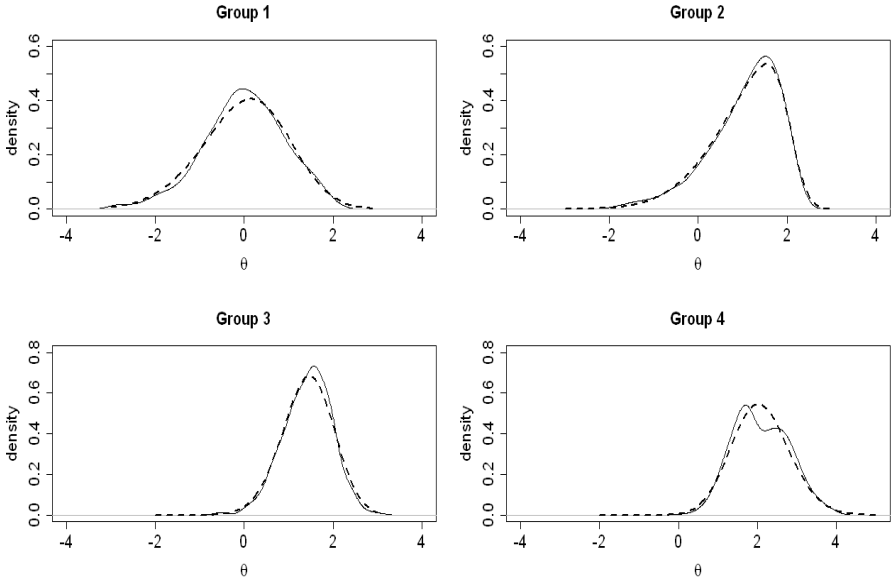

Figure 19: Latent scores distributions with theoretical curves (skew model). Legend: (— estimated curve, − − − theoretical curve)

Table 4: Statistics for model comparisons

Model	$\widehat{D(\bar{\vartheta})}$	$\widehat{\overline{D(\vartheta)}}$	$\widehat{\rho}_D$	$\widehat{\text{DIC}}$	$\widehat{\text{EAIC}}$	$\widehat{\text{EBIC}}$
Symmetric	58103.00	60073.00	1969.80	62042.60	64012.60	83477.08
Skew	57873.00	59836.00	1962.60	61798.20	63761.20	83154.54

6. Final conclusions and remarks

We presented an multiple group IRT model with a centred skew-normal assumption for the latent trait distributions. Such approach is more flexible than the usual onde (by considering symmetric normal distributions) and more straightforwardly interpretable than the nonparametric ones (in terms of the density obtained in the final step of the estimation process). Moreover, it leads to an identified model. We developed two MCMC algorithms for parameter estimation and compared them concerning convergence issues. The selected algorithm (AGDS, which is a full Gibbs sampling algorithm) showed to be more efficient in terms of parameter recovery, according to the simulation study. Our model fitted better the real data set studied than the standard normal multiple group model. Also, our algorithm indicated that three groups, present negative asymmetric behavior. In conclusion: our approach showed to be a promising alternative to the usual ones in analyzing multiple groups IRT data sets. For future research we intend to study extensions of our model by considering other Item Response Functions, longitudinal data sets and multidimensional tests. It is also called for the investigation of alternative estimation algorithms.

7. Acknowledgements

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support.

8. Appendix

In the following we present more details concerning the resulting posterior distribution (obtained using the density prior (19) and hierarchical representations) and the MCMC algorithms.

Using the augmented likelihood (13) and the priori distribution (19), we have the following posterior distribution for $(\boldsymbol{Z}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{h}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, \boldsymbol{t}_{.})$:

$$
\begin{aligned}
p(\mathbf{Z}_{...}, \boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \mathbf{y}_{...}, \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) &\propto p(\mathbf{Z}_{...}|\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \mathbf{y}_{...}) p(\boldsymbol{\theta}_{..}|\boldsymbol{\eta}_\theta) p(\boldsymbol{\zeta}|\boldsymbol{\eta}_\zeta) p(\boldsymbol{\eta}_\theta|\boldsymbol{\eta}_\eta) \\
&= \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} p(z_{ijk}|\theta_{jk}, \boldsymbol{\zeta}_i, y_{ijk}) \right\} \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(\theta_{jk}|\boldsymbol{\eta}_{\theta_k}) \right\} \\
&\times \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i) \right\} \left\{ \prod_{k=1}^{K} p(\xi_{\theta_k}, \omega_{\theta_k}, \lambda_{\theta_k}) \right\} \\
&\propto \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} \exp\left\{ -0,5(z_{ijk} - a_i\theta_{jk} + b_i)^2 \right\} 1\!\!1_{(Z_{ijk}, y_{ijk})} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \omega_{\theta_k}^{-n_k} \prod_{j=1}^{n_k} \phi\left(\frac{\theta_{jk} - \xi_{\theta_k}}{\omega_{\theta_k}} \right) \Phi\left[\left(\frac{\theta_{jk} - \xi_{\theta_k}}{\omega_{\theta_k}} \right) \right] \right\} \\
&\times \left\{ \prod_{i=1}^{I} \exp\left[-0,5(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta)^\top \boldsymbol{\Psi}_\zeta{}^{-1} (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta) \right] 1\!\!1_{(a_i > 0)} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \exp\left\{ -\frac{1}{2\sigma^2_{\xi_\theta}} (\xi_{\theta_k} - \mu_{\xi_\theta})^2 \right\} \right\} \\
&\times \left\{ \prod_{k=1}^{K} (\omega^2_{\theta_k})^{-\alpha_{\omega_\theta} - 1} \exp(-\beta_{\omega_\theta} \frac{1}{\omega^2_{\theta_k}}) \right\} \\
&\times \left\{ \prod_{k=1}^{K} \exp\left\{ -\frac{1}{2\sigma^2_{\lambda_\theta}} (\lambda_{\theta_k} - \mu_{\lambda_\theta})^2 \right\} \right\}. \qquad (38)
\end{aligned}
$$

Now, using the augmented likelihood (13) and the prior distribution (37) we have that:

$$
\begin{aligned}
p(\boldsymbol{\theta}_{..}, \boldsymbol{h}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, \boldsymbol{t}_{.}|\boldsymbol{y}_{...}, \boldsymbol{\eta}_{\boldsymbol{\zeta}}, \boldsymbol{\eta}_{\boldsymbol{\eta}}) &\propto p(\boldsymbol{Z}_{...}|\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{y}_{...}) p(\boldsymbol{\theta}_{..}|\boldsymbol{h}_{..}, \boldsymbol{\eta}_\theta) p(\boldsymbol{\zeta}|\boldsymbol{\eta}_{\boldsymbol{\zeta}}) p(\boldsymbol{\eta}_\theta|\boldsymbol{t}_{.}, \boldsymbol{\eta}_{\boldsymbol{\eta}}) p(\boldsymbol{t}_{.}) \\
&= \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} p(z_{ijk}|\theta_{jk}, \boldsymbol{\zeta}_i, y_{ijk}) \right\} \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(\theta_{jk}|h_{jk}, \boldsymbol{\eta}_{\theta_k}) \right\} \\
&\times \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(h_{jk}) \right\} \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i) \right\} \left\{ \prod_{k=1}^{K} p(\eta_{\theta_k}|\boldsymbol{\eta}_{\eta_k}) \right\} \left\{ \prod_{k=1}^{K} p(t_k) \right\} \\
&\propto \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} \exp\left\{ -0,5(z_{ijk} - a_i\theta_{jk} + b_i)^2 \right\} 1\!\!1_{(Z_{ijk}, y_{ijk})} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \varsigma_{\theta_k}^{-\frac{n_k}{2}} \prod_{j=1}^{n_k} \exp\left\{ -\frac{1}{2\varsigma_{\theta_k}^2} \left[\theta_{jk} - \xi_{\theta_k} - \tau_{\theta_k} h_{jk}\right]^2 \right\} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \exp\left(-\frac{1}{2} h_{jk}^2 \right) 1\!\!1_{(h_{jk}>0)} \right\} \\
&\times \left\{ \prod_{i=1}^{I} \exp\left[-0,5(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta)^\top \boldsymbol{\Psi}_\zeta{}^{-1} (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta) \right] 1\!\!1_{(a_i>0)} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \frac{1}{\varsigma_{\theta_k}^2} \exp\left(-\frac{1}{2} \frac{\tau_{\theta_k}^2 t_k}{\varsigma_{\theta_k}^2 \varphi^2} \right) t_k^{\frac{d+1}{2}-1} \exp\left(-\frac{b}{2} t_k \right) \right\}, \qquad (39)
\end{aligned}
$$

if we use the prior (34), or

$$
\begin{aligned}
p(\boldsymbol{\theta}_{..}, \boldsymbol{h}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{y}_{...}, \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) &\propto p(\boldsymbol{Z}_{...}|\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{y}_{...}) p(\boldsymbol{\theta}_{..}|\boldsymbol{h}_{..}, \boldsymbol{\eta}_\theta) p(\boldsymbol{\zeta}|\boldsymbol{\eta}_\zeta) p(\boldsymbol{\eta}_\theta|\boldsymbol{\eta}_\eta) \\
&= \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} p(z_{ijk}|\theta_{jk}, \boldsymbol{\zeta}_i, y_{ijk}) \right\} \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(\theta_{jk}|h_{jk}, \boldsymbol{\eta}_{\theta_k}) \right\} \\
&\times \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} p(h_{jk}) \right\} \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i) \right\} \left\{ \prod_{k=1}^{K} p(\eta_{\theta_k}|\boldsymbol{\eta}_{\eta_k}) \right\} \left\{ \prod_{k=1}^{K} p(t_k) \right\} \\
&\propto \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i \in I_{jk}} \exp\left\{ -0,5(z_{ijk} - a_i\theta_{jk} + b_i)^2 \right\} 1\!\!1_{(Z_{ijk}, y_{ijk})} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \varsigma_{\theta_k}^{-\frac{n_k}{2}} \prod_{j=1}^{n_k} \exp\left\{ -\frac{1}{2\varsigma_{\theta_k}^2} \left[\theta_{jk} - \xi_{\theta_k} - \tau_{\theta_k} h_{jk} \right]^2 \right\} \right\} \\
&\times \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_k} \exp\left(-\frac{1}{2} h_{jk}^2 \right) 1\!\!1_{(h_{jk}>0)} \right\} \\
&\times \left\{ \prod_{i=1}^{I} \exp\left[-0,5(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta)^\top \boldsymbol{\Psi}_\zeta{}^{-1} (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_\zeta) \right] 1\!\!1_{(a_i>0)} \right\} \\
&\times \left\{ \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\beta}_\theta - \boldsymbol{\mu}_\theta \right)^\top \Sigma_\theta^{-1} \left(\boldsymbol{\beta}_\theta - \boldsymbol{\mu}_\theta \right) \right\} \right\} \\
&\times \left\{ (\varsigma_{\theta_k}^2)^{-(\alpha_{\varsigma_\theta}+1)} \exp\left(-\frac{\beta_{\varsigma_\theta}}{\varsigma_{\theta_k}^2} \right) \right\}, \qquad (40)
\end{aligned}
$$

if we consider the priors (35) and (36). From the posterior distributions (39) and (40) we obtain all full conditional distributions, which are known and easy to sample from, see Appendix. Thus, it is possible to develop a full Gibbs sampling scheme as follows:

In the following we present a description od the ADMHWGS algorithm (using the original density):

- Step 1: Simulate the augmented variables $Z_{ijk}^{(t)}$ from $Z_{ijk}|(.) \sim N(a_i^{(t-1)}\theta_{jk}^{(t-1)} - b_i^{(t-1)}, 1) 1\!\!1_{(y_{ijk}, z_{ijk})}$, $\forall\ k = 1, \ldots, K,\ j = 1, \ldots, n_k$ e $i \in I_{jk}$ independently.

- Step 2: Simulate $\theta_{jk}^{(t)}$ de $\theta_{jk}|(.)$ independently, by considering:

  (a) Simulate $\theta_{jk}^{(*)}|\theta_{jk}^{(t-1)} \sim N(\theta_{jk}^{(t-1)}, \sigma_0^2)$,

(b) Accept $\theta_{jk}^{(t)} = \theta_{jk}^{(*)}$ with probability

$$\pi_{jk}(\theta_{jk}^{(t-1)}, \theta_{jk}^{(*)}) = \min\left\{R_{\theta_{jk}}, 1\right\} \text{ where,}$$

$$\begin{aligned} R_{\theta_{jk}} &= \frac{\exp\left\{-\frac{1}{2}\left\{(\theta_{jk}^{(*)})^2 \sum_{i\in I_{jk}}(a_i^{(t-1)})^2 - 2\theta_{jk}^{(*)}\left[\sum_{i\in I_{jk}}(z_{ijk}^{(t)}a_i^{(t-1)}) + \sum_{i\in I_{jk}}(a_i^{(t-1)}b_i^{(t-1)})\right]\right\}\right\}}{\exp\left\{-\frac{1}{2}\left\{(\theta_{jk}^{(t-1)})^2 \sum_{i\in I_{jk}}(a_i^{(t-1)})^2 - 2\theta_{jk}^{(t-1)}\left[\sum_{i\in I_{jk}}(z_{ijk}^{(t)}a_i^{(t-1)}) + \sum_{i\in I_{jk}}(a_i^{(t-1)}b_i^{(t-1)})\right]\right\}\right\}} \\ &\times \frac{\exp\left\{-\frac{1}{2(\omega_{\theta_k}^{(t-1)})^2}\left[(\theta_{jk}^{(*)})^2 - 2\theta_{jk}^{(*)}\xi_{\theta_k}^{(t-1)}\right]\right\}\Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(*)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t-1)}}\right)\right]}{\exp\left\{-\frac{1}{2(\omega_{\theta_k}^{(t-1)})^2}\left[(\theta_{jk}^{(t-1)})^2 - 2\theta_{jk}^{(t-1)}\xi_{\theta_k}^{(t-1)}\right]\right\}\Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(t-1)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t-1)}}\right)\right]}. \end{aligned} \tag{41}$$

Otherwise fix $\theta_{jk}^{(t)} = \theta_{jk}^{(t-1)}$.

- Step 3: Simulate the item parameters $\boldsymbol{\zeta}_i^{(t)}$ from $\boldsymbol{\zeta}_i|(.) \sim N(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}^{(t)}\widehat{\boldsymbol{\zeta}}_i^{(t)}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}^{(t)})$, see [11], independently, where

$$\begin{aligned} \widehat{\boldsymbol{\zeta}}_i^{(t)} &= \left(\boldsymbol{\Theta}_{i.}^{(t)}\right)^\top \boldsymbol{z}_{i.}^{(t)} + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\zeta}}, \\ \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}^{(t)} &= \left[\left(\boldsymbol{\Theta}_{i.}^{(t)}\right)^\top (\boldsymbol{\Theta}_{i.}) + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1}\right]^{-1}, \\ \boldsymbol{\Theta}_{i.} &= \left[\boldsymbol{\theta}^{(t)} - \mathbf{1}_n\right] \bullet \mathbb{1}_i, \end{aligned}$$

  where $\mathbb{1}_i$ ia $(n \times 2)$ matrix with the elements, in each line, equal to 1 or 0, according to observe or not the answer of the examinee j from the group k to the item i and $\bullet$ stands for the *Hadamard* product, see [34].

- Step 4: Choose either the Steps 4.1 and 4.2 or the Steps 4.3 to 4.5.

  (**Simulation of the population parameters considering the prior given by equation** (16))

- Step 4.1: Simulate $\xi_{\theta_k}^{(t)}$ de $\xi_{\theta_k}|(.)$, by considering:

(a) Simulate $\xi_{\theta_k}^{(*)}|\xi_{\theta_k}^{(t-1)} \sim N(\xi_{\theta_k}^{(t-1)}, \sigma_0^2)$

(b) Accept $\xi_{\theta_k}^{(t)} = \xi_{\theta_k}^{(*)}$ with probability

$$\pi_k\left[\xi_{\theta_k}^{(t-1)}, \xi_{\theta_k}^{(*)}\right] = \min\left\{R_{\xi_{\theta_k}}, 1\right\} \text{ where,}$$

$$R_{\xi_{\theta_k}} = \frac{\exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(*)}}{\omega_{\theta_k}^{(*)}}\right)^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(*)}}{\omega_{\theta_k}^{(*)}}\right)\right]}{\exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t-1)}}\right)^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t-1)}}\right)\right]}$$

Otherwise, fix $\xi_{\theta_k}^{(t)} = \xi_{\theta_k}^{(t-1)}$

- Step 4.2: Simulate $\omega_{\theta_k}^{(t)}$ de $\omega_{\theta_k}|(.)$ by considering:

  (a) Simulate $\omega_{\theta_k}^{(*)}|\omega_{\theta_k}^{(t-1)} \sim \text{Log-Normal}(\log(\omega_{\theta_k}^{(t-1)}), \sigma_0^2)$,

  (b) Accept $\omega_{\theta_k}^{(t)} = \omega_{\theta_k}^{(*)}$ with probability

$$\pi_k(\omega_{\theta_k}^{(t-1)}, \omega_{\theta_k}^{(*)}) = \min\left\{R_{\omega_{\theta_k}}, 1\right\} \text{ where,}$$

$$R_{\omega_{\theta_k}} = \frac{(\omega^{(*)})^{-n_k}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t)}}{\omega_{\theta_k}^{(*)}}\right)^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t)}}{\omega_{\theta_k}^{(*)}}\right)\right]}{(\omega^{(t-1)})^{-n_k}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t)}}{\omega_{\theta_k}^{(t-1)}}\right)^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t-1)}}\right)\right]}$$

- Step 4.3: Simulate $\lambda_{\theta_k}^{(t)}$ de $\lambda_{\theta_k}|(.)$ by considering:

  (a) Simulate $\lambda_{\theta_k}^{(*)}|\lambda_{\theta_k}^{(t-1)} \sim N(\lambda_{\theta_k}^{(t-1)}, \sigma_0^2)$,

  (b) Accept $\lambda_{\theta_k}^{(t)} = \lambda_{\theta_k}^{(*)}$ with probability

$$\pi_k(\lambda_{\theta_k}^{(t-1)}, \lambda_{\theta_k}^{(*)}) = \min\left\{R_{\lambda_{\theta_k}}, 1\right\} \text{ where,}$$

$$R_{\lambda_{\theta_k}} = \frac{\left(1+\frac{(\lambda_{\theta_k}^{(*)})^2}{d\varphi^2}\right)^{-\frac{d+1}{2}} \prod_{j=1}^{n_k} \Phi\left[\lambda_{\theta_k}^{(*)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t)}}{\omega_{\theta_k}^{(t)}}\right)\right]}{\left(1+\frac{(\lambda_{\theta_k}^{(t-1)})^2}{d\varphi^2}\right)^{-\frac{d+1}{2}} \prod_{j=1}^{n_k} \Phi\left[\lambda_{\theta_k}^{(t-1)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t)}}{\omega_{\theta_k}^{(t)}}\right)\right]}.$$

Otherwise, fix $\lambda_{\theta_k}^{(t)} = \lambda_{\theta_k}^{(t-1)}$.

(**Simulation of the population parameters considering the priors given by equations** (20), (21) e (22))

- Step 4.4: Simulate $\xi_{\theta_k}^{(t)}$ de $\xi_{\theta_k}|(.)$ by considering:

  (a) Simulate $\xi_{\theta_k}^{(*)}|\xi_{\theta_k}^{(t-1)} \sim N(\xi_{\theta_k}^{(t-1)}, \sigma_0^2)$,

  (b) Accept $\xi_{\theta_k}^{(t)} = \xi_{\theta_k}^{(*)}$ with probability

$$\pi_k(\xi_{\theta_k}^{(t-1)}, \xi_{\theta_k}^{(*)}) = \min\left\{R_{\xi_{\theta_k}}, 1\right\} \text{ where,}$$

$$R_{\xi_{\theta_k}} = \frac{\exp\left\{-\frac{1}{2\sigma_{\xi_\theta}^2}(\xi_{\theta_k}^{(*)}-\mu_{\xi_\theta})^2\right\} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(*)}}{\omega_{\theta_k}^{(t)}}\right)^2\right\} \prod_{j=1}^{n_k} \Phi\left[\lambda_{\theta_k}^{(t)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(*)}}{\omega_{\theta_k}^{(t)}}\right)\right]}{\exp\left\{-\frac{1}{2\sigma_{\xi_\theta}^2}(\xi_{\theta_k}^{(t-1)}-\mu_{\xi_\theta})^2\right\} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t)}}\right)^2\right\} \prod_{j=1}^{n_k} \Phi\left[\lambda_{\theta_k}^{(t)}\left(\frac{\theta_{jk}^{(t)}-\xi_{\theta_k}^{(t-1)}}{\omega_{\theta_k}^{(t)}}\right)\right]}.$$

Otherwise, fix $\xi_{\theta_k}^{(t)} = \xi_{\theta_k}^{(t-1)}$.

- Step 4.5: Simulate $(\omega_{\theta_k}^2)^{(t)}$ de $\omega_{\theta_k}^2|(.)$ by considering:

  (a) Simulate $(\omega_{\theta_k}^2)^{(*)}|(\omega_{\theta_k}^2)^{(t-1)} \sim \text{Log-Normal}(\log((\omega_{\theta_k}^2)^{(t-1)}), \sigma_0^2)$,

  (b) Accept $(\omega_{\theta_k}^2)^{(t)} = (\omega_{\theta_k}^2)^{(*)}$ with probability

$$\pi_k((\omega_{\theta_k}^2)^{(t-1)}, (\omega_{\theta_k}^2)^{(*)}) = \min\left\{R_{\omega_{\theta_k}^2}, 1\right\} \text{ where,}$$

$$R_{\omega^2_{\theta_k}} = \frac{\left((\omega^2_{\theta_k})^{(*)}\right)^{-(\alpha_{\omega_\theta}+\frac{n_k}{2})} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta^{(t)}_{jk}-\xi^{(t)}_{\theta_k}}{\omega^{(*)}_{\theta_k}}\right)^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda^{(t)}_{\theta_k}\left(\frac{\theta^{(t)}_{jk}-\xi^{(t)}_{\theta_k}}{\omega^{(*)}_{\theta_k}}\right)\right]}{\left((\omega^2_{\theta_k})^{(t-1)}\right)^{-(\alpha_{\omega_\theta}+\frac{n_k}{2})} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_k}\left(\frac{\theta^{(t)}_{jk}-\xi^{(t)}_{\theta_k}}{\omega^{(t-1)}_{\theta_k}}\right)^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda^{(t)}_{\theta_k}\left(\frac{\theta^{(t)}_{jk}-\xi^{(t)}_{\theta_k}}{\omega^{(t-1)}_{\theta_k}}\right)\right]}.$$

Otherwise, fix $(\omega^2_{\theta_k})^{(t)} = (\omega^2_{\theta_k})^{(t-1)}$.

- Step 4.5: Simulate $\lambda^{(t)}_{\theta_k}$ de $\lambda_{\theta_k}|(.)$ by considering:

  (a) Simulate $\lambda^{(*)}_{\theta_k}|\lambda^{(t-1)}_{\theta_k} \sim N(\lambda^{(t-1)}_{\theta_k}, \sigma^2_0)$,

  (b) Accept $\lambda^{(t)}_{\theta_k} = \lambda^{(*)}_{\theta_k}$ with probability

$$\pi_k(\lambda^{(t-1)}_{\theta_k}, \lambda^{(*)}_{\theta_k}) = \min\left\{R_{\lambda_{\theta_k}}, 1\right\} \text{ where,}$$

$$R_{\lambda_{\theta_k}} = \frac{\exp\left\{-\frac{1}{2\sigma^2_{\lambda_\theta}}(\lambda^{(*)}_{\theta_k}-\mu_{\lambda_\theta})^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda^{(*)}_{\theta_k}\left(\frac{\theta^{(t)}_{jk}-\xi^{(t)}_{\theta_k}}{\omega^{(t)}_{\theta_k}}\right)\right]}{\exp\left\{-\frac{1}{2\sigma^2_{\lambda_\theta}}(\lambda^{(t-1)}_{\theta_k}-\mu_{\lambda_\theta})^2\right\}\prod_{j=1}^{n_k}\Phi\left[\lambda^{(t-1)}_{\theta_k}\left(\frac{\theta^{(t)}_{jk}-\xi^{(t)}_{\theta_k}}{\omega^{(t)}_{\theta_k}}\right)\right]}.$$

Otherwise, fix $\lambda^{(t)}_{\theta_k} = \lambda^{(t-1)}_{\theta_k}$.

<u>**Note**</u>: The simulated values for the population parameters $(\mu_{\theta_k}, \sigma_{\theta_k}, \gamma_{\theta_k})$ are obtained in every iteration from the simulated values of the parameters $(\xi_{\theta_k}, \omega_{\theta_k}, \lambda_{\theta_k})$ by doing:

$$\begin{aligned}
\gamma^{(t)}_{\theta_k} &= r\left(\delta^{(t)}_{\theta_k}\right)^3\left(\frac{4}{\pi}-1\right)\left(1-r^2\left(\delta^{(t)}_{\theta_k}\right)^2\right)^{-3/2}, \\
\sigma^{(t)}_{\theta_k} &= \frac{\omega^{(t)}_{\theta_k}}{\sqrt{1+\left(\gamma^{(t)}_{\theta_k}\right)^{2/3}s^2}}, \\
\mu^{(t)}_{\theta_k} &= \xi^{(t)}_{\theta_k} + \sigma^{(t)}_{\theta_k}\left(\gamma^{(t)}_{\theta_k}\right)^{1/3}s,
\end{aligned}$$

for $t = 1, 2, \ldots, B, \ldots, M$, where B is the *burn-in* and M and the size of the generated sample and,

$$\begin{aligned}
\delta_{\theta_k}^{(t)} &= \frac{\lambda_{\theta_k}^{(t)}}{\sqrt{1 + \left(\lambda_{\theta_k}^{(t)}\right)^2}}, \\
s &= \left(\frac{2}{4-\pi}\right)^{1/3}.
\end{aligned}$$

In the following we describe the ADGS algorithm.

- Step 1: Simulate the augmented variables $Z_{ijk}^{(t)}$ de $Z_{ijk}|(.) \sim N(a_i^{(t-1)}\theta_{jk}^{(t-1)} - b_i^{(t-1)}, 1)\mathbb{1}_{(y_{ijk}, z_{ijk})}$, independently $\forall\ k = 1, \ldots, K,\ j = 1, \ldots, n_k$ and $i \in I_{jk}$.

- Step 2: Simulate $\theta_{jk}^{(t)}$ de $\theta_{jk}|(.) \sim N(\widehat{\tau}_{\theta_{jk}}^{(t-1)}\widehat{\theta}_{jk}^{(t-1)}, \widehat{\tau}_{jk}^{(t-1)})\ \forall\ k = 1, \ldots, K$, $j = 1, \ldots, n_k$ independently, where:

$$\begin{aligned}
\widehat{\theta}_{jk}^{(t-1)} &= \sum_{i \in I_{jk}} z_{ijk}^{(t)} a_i^{(t-1)} + \sum_{i \in I_{jk}} a_i^{(t-1)} b_i^{(t-1)} + \frac{\xi_{\theta_k}^{(t-1)} + \tau_{\theta_k}^{(t-1)} h_{jk}^{(t-1)}}{(\varsigma_{\theta_k}^{(t-1)})^2}, \\
\widehat{\tau}_{\theta_{jk}}^{(t-1)} &= \left(\frac{1}{(\varsigma_{\theta_k}^{(t-1)})^2} + \sum_{i \in I_{jk}} (a_i^{(t-1)})^2\right)^{-1}.
\end{aligned}$$

- Step 3: Simulate the latent variables $H_{jk}^{(t)}$ de

$$H_{jk}|(.) \sim \text{HN}\left(\frac{\tau_{\theta_k}^{(t-1)}(\theta_{jk}^{(t)} - \xi_{\theta_k}^{(t-1)})}{(\tau_{\theta_k}^{(t-1)})^2 + (\varsigma_{\theta_k}^{(t-1)})^2}, \frac{(\varsigma_{\theta_k}^{(t-1)})^2}{(\varsigma_{\theta_k}^{(t-1)})^2 + (\tau_{\theta_k}^{(t-1)})^2}\right),$$

  independently $\forall\ k = 1, \ldots, K$ e $j = 1, \ldots, n_k$.

- Step 4: Simulate the item parameters $\boldsymbol{\zeta}_i^{(t)}$ de $\boldsymbol{\zeta}_i|(.) \sim N(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}^{(t)}\widehat{\boldsymbol{\zeta}}_i^{(t)}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}^{(t)})$, veja [11], independently, where

$$
\begin{aligned}
\widehat{\boldsymbol{\zeta}}_i^{(t)} &= \left(\boldsymbol{\Theta}_{i..}^{(t)}\right)^\top \boldsymbol{z}_{i..}^{(t)} + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\zeta}}, \\
\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}^{(t)} &= \left[\left(\boldsymbol{\Theta}_{i..}^{(t)}\right)^\top (\boldsymbol{\Theta}_{i..}) + \boldsymbol{\Psi}_{\boldsymbol{\zeta}}^{-1}\right]^{-1}, \\
\boldsymbol{\Theta}_{i..} &= \left[\boldsymbol{\theta}^{(t)} - \mathbf{1}_n\right] \bullet \mathbb{1}_i,
\end{aligned}
$$

where $\mathbb{1}_i$ ia $(n \times 2)$ matrix with the elements, in each line, equal to 1 or 0, according to observe or not the answer of the examinee j from the group k to the item i and $\bullet$ stands for the *Hadamard* product, see [34].

- Step 5: Choose either the Steps 5.1 to 5.4 or the Steps 5.5 and 5.6.

  (**Simulation of the population parameters considering the prior given by equation** (34)).

- Step 5.1: Simulate the latent variables $T_k^{(t)}$ de

$$
T_k|(.) \sim \text{Gamma}\left(\frac{d+1}{2}, \frac{(\tau_{\theta_k}^{(t-1)})^2}{(\varsigma_{\theta_k})^2\varphi^2} + d\right),
$$

  independently $\forall\ k = 1, \ldots, K$.

- Step 5.2: Simulate $\xi_{\theta_k}^{(t)}$ de

$$
\xi_{\theta_k}|(.) \sim N\left(\frac{1}{n_k}\sum_{j=1}^{n_k}(\theta_{jk}^{(t)} - \tau_{\theta_k}^{(t-1)}h_{jk}^{(t)}), \frac{(\varsigma_{\theta_k}^{(t-1)})^2}{n_k}\right)
$$

  independently $\forall\ k = 1, \ldots, K$.

- Step 5.3: Simulate $\tau_{\theta_k}^{(t)}$ de $\tau_{\theta_k}|(.) \sim N(\widehat{\mu}_{\tau_{\theta_k}}^{(t)}, (\widehat{\sigma}_{\theta_k}^{(t)})^2)$ independently, where:

$$
\begin{aligned}
\widehat{\mu}_{\tau_{\theta_k}}^{(t)} &= \frac{\sum_{j=1}^{n_k}(\theta_{jk}^{(t)} - \xi_{\theta_k}^{(t)})h_{jk}^{(t)}}{\sum_{j=1}^{n_k}(h_{jk}^{(t)})^2 + \frac{t_k^{(t)}}{(\varsigma_{\theta_k}^{(t-1)})^2\varphi^2}}, \\
(\widehat{\sigma}_{\tau_{\theta_k}}^{(t)})^2 &= \frac{(\varsigma_{\theta_k}^{(t-1)})^2}{\sum_{j=1}^{n_k}(h_{jk}^{(t)})^2 + \frac{t_k^{(t)}}{(\varsigma_{\theta_k}^{(t-1)})^2\varphi^2}},
\end{aligned}
$$

$\forall\ k = 1, \ldots, K$.

- Step 5.4: Simulate $(\varsigma_{\theta_k}^{(t)})^2$ de $\varsigma_{\theta_k}^2 \sim$ gamma-Inv.$(\widehat{\alpha}_{\varsigma_{\theta_k}}^{(t)}, \widehat{\beta}_{\varsigma_{\theta_k}}^{(t)})$ independently, where:

$$
\begin{aligned}
\widehat{\alpha}_{\varsigma_{\theta_k}}^{(t)} &= \frac{n_k + 4}{2}, \\
\widehat{\beta}_{\varsigma_{\theta_k}}^{(t)} &= \frac{1}{2}\left[\sum_{j=1}^{n_k}\left(\theta_{jk}^{(t)} - \xi_{\theta_k}^{(t)} - \tau_{\theta_k}^{(t)}h_{jk}^{(t)}\right)^2 + \frac{(\tau_{\theta_k}^{(t)})^2 t_k^{(t)}}{\varphi^2}\right],
\end{aligned}
$$

$\forall\ k = 1, \ldots, K$.

(**Simulation of the population parameters considering the prior given by equation** (35) e (36)).

- Step 5.5: Simulate $\beta_{\theta_k}^{(t)} = (\xi_{\theta_k}^{(t)}, \tau_{\theta_k}^{(t)})^\top$ de $\beta_{\theta_k}|(.) \sim N\left[\widehat{\boldsymbol{\tau}}_{\theta_k}^{(t)}\widehat{\boldsymbol{\xi}}_{\theta_k}^{(t)}, \widehat{\boldsymbol{\tau}}_{\theta_k}^{(t)}\right]$ independently, where:

$$
\begin{aligned}
\widehat{\boldsymbol{\tau}}_{\theta_k}^{(t)} &= \left[\left(\boldsymbol{H}_{.k}^{(t)}\right)^\top \left(\boldsymbol{D}_{n_k}^{(t-1)}\right)^{-1}\left(\boldsymbol{H}_{.k}^{(t)}\right) + \Sigma_{\beta_\theta}^{-1}\right]^{-1}, \\
\widehat{\boldsymbol{\xi}}_{\theta_k}^{(t)} &= \left(\boldsymbol{H}_{.k}^{(t)}\right)\left(\boldsymbol{D}_{n_k}^{(t-1)}\right)^{-1}\boldsymbol{\theta}_{.k}^{(t)} + \Sigma_{\beta_\theta}^{-1}\boldsymbol{\mu}_{\beta_\theta}, \\
\boldsymbol{H}_{.k}^{(t)} &= \begin{bmatrix}\mathbf{1}_{n_k} & \boldsymbol{h}_{.k}^{(t)}\end{bmatrix} \text{ e} \\
\boldsymbol{D} &= \begin{pmatrix} \left(\varsigma_{\theta_k}^{(t-1)}\right)^2 & & \\ & \ddots & \\ & & \left(\varsigma_{\theta_k}^{(t-1)}\right)^2 \end{pmatrix},
\end{aligned}
$$

is a matrix of order n_k, $\forall\ k = 1, \ldots, K$.

- Step 6: Simulate $\left(\varsigma_{\theta_k}^{(t)}\right)^2$ from

$$\varsigma_{\theta_k}^2 |(.) \sim \text{IG}\left(\frac{n_k}{2} + \alpha_{\varsigma_\theta}, \frac{1}{2}\sum_{j=1}^{n_k}\left(\theta_{jk}^{(t)} - \xi_{\theta_k}^{(t)} - \tau_{\theta_k}^{(t)} h_{jk}^{(t)}\right)^2 + \beta_{\varsigma_\theta}\right),$$

  independently $\forall\ k = 1, \ldots, K$.

<u>**Note**</u>: The simulated values of the population parameters $(\mu_{\theta_k}, \sigma_{\theta_k}, \gamma_{\theta_k})$ are obtained at every iteration from the simulated values of the parameters $(\xi_{\theta_k}, \tau_{\theta_k}, \varsigma_{\theta_k})$ by doing:

$$\begin{aligned}
\gamma_{\theta_k}^{(t)} &= r\left(\delta_{\theta_k}^{(t)}\right)^3\left(\frac{4}{\pi} - 1\right)\left(1 - r^2\left(\delta_{\theta_k}^{(t)}\right)^2\right)^{-3/2}, \\
\sigma_{\theta_k}^{(t)} &= \frac{\omega_{\theta_k}^{(t)}}{\sqrt{1 + \left(\gamma_{\theta_k}^{(t)}\right)^{2/3} s^2}}, \\
\mu_{\theta_k}^{(t)} &= \xi_{\theta_k}^{(t)} + \sigma_{\theta_k}^{(t)}\left(\gamma_{\theta_k}^{(t)}\right)^{1/3} s,
\end{aligned}$$

for $t = 1, 2, \ldots, B, \ldots, M$, where B is the *burn-in* and M is the size of the sample generated,

$$\begin{aligned}
\delta_{\theta_k}^{(t)} &= \frac{\tau_{\theta_k}^{(t)}}{\sqrt{\left(\tau_{\theta_k}^{(t)}\right)^2 + \left(\varsigma_{\theta_k}^{(t)}\right)^2}}, \\
\omega_{\theta_k}^{(t)} &= \frac{\tau_{\theta_k}^{(t)}}{\delta_{\theta_k}^{(t)}}.
\end{aligned}$$

[1] Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F. (2011), Bayesian inference for a skew-normal IRT model under the centred parameterization, *Computational Statistics & Data Analysis, 55,1, 353-365.*

[2] Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F. (2011), Parameter recovery for a skew-normal IRT model under a Bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size,*Accepted for publication, DOI:10.1080/00949655.2011.591798.*

[3] Santos, J. R. S. (2012), *A Multiple Group Item Response Theory Model with skew normal latent trait distributions under centred parameterization*, Master Thesis (In Portuguese), Department of Statistics, University of Campinas.

[4] Bock, D. R. and Zimowski, M. F. (1997), The Multiple Groups IRT, *in Handbook of modern Item Response Theory, Wim J. van der Linden and Ronald K. Hambleton eds, Springer-Verlag*

[5] Micceri, T. (1989), *The unicorn, the normal curve, and other improbable creatures*, Psychological bulletin, (105), 1, 156–166.

[6] Samejima, F., (1997), *Departure from normal assumptions: a promise for future psychometrics with substantive mathematical modeling*, Psychometrika, (62), 4, 471-493.

[7] Azevedo, C. L. N. and Andrade, D. F. (2009), *An estimation method for latent traits and population paremeter for the nominal response model*, Brazilian Journal of Probability ans Statistics, to appear.

[8] Azevedo, C. L. N. and Andrade, D. F. and Fox, J.-P. (2012), *A Bayesian generalized multiple group IRT model with model-fit assessment tools, accepted for publication, Computational Statistics & Data Analysis, http://dx.doi.org/10.1016/j.csda.2012.03.017* .

[9] Azevedo, C. L. N., Fox, J.-P., Andrade, D. F. Andrade (2009), *Bayesian general multivariate latent variable modeling of longitudinal item response data*, under review, Journal of Multivariate Analysis.

[10] Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F. (2009), *A note on identifiability and metric issues for skew IRT models.*, technical report.

[11] Azevedo, C. L. N., (2008) *Multilevel multiple group longitudinal models in Item Response Theory: estimation methods anb structural selection under a Bayesian perspective*, Unpublished PhD Thesis, In Portuguese.

[12] Seong, Tae-Je (1990), *Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions*, Applied psychological measurement, 14, 3, 299-311.

[13] Kirisci, L., Hsu, Tse-chi and Yu, L. (2001), *Robustness of item parameter estimation programs to assumptions of unidimensionality and normality*, Applied psychological measurement, 25,2, 146-162.

[14] De Ayala, R. J and Sava - Bolesta, M. (1999), *Item Parameter Recovery for the Nominal Response Model*, Applied psychological measurement, 23, 1, 3-19.

[15] DeMars, C. E. (2003), *Sample Size and the Recovery of Nominal Response Model Item Parameters*, Applied psychological measurement, 27, 4, 275-288.

[16] Wollack, J. A., Bolt, D. M., Cohen, A. S., and Lee, Young-Sun. (2002), *Recovery of Item Parameters in the Nominal Response Model: A Comparison of Marginal Maximum Likelihood Estimation and Markov Chain Monte Carlo Estimation*, Applied psychological measurement, 26, 3, 339-352.

[17] Swaminathan, H., Hambleton, R. K. , Sireci, S. G. , Xing, D. and Rizavi, S. M. (2003), *Small ample size estimation in dichotomous item respjnse models: effect of prioris based on judgmental information on the accuracy of item parameter estimates*, Applied psychological measurement, 27, 1, 27-51.

[18] Harwell, M. R. and Janosky, J. E. (1991), *An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in Bilog*, Applied psychological measurement, 15, 3, 279-291.

[19] Gifford, J. A. and Swaminathan, H. (1990), *Bias and the effect of priors in bayesian estimation of item response models*, Applied psychological measurement, 14, 33, 33-43.

[20] Azzalini, A. (1985), *A class of distribution which includes the normal ones*, Scandinavian journal of statistics. 12, pp. 171-178.

[21] Azzalini, A. and Capitanio, A. (1999), *Statistical Applications of the Multivariate Skew Normal Distribution*, SJournal of the Royal Statistical Society. Series B. 61, pp. 579-602.

[22] Pewsey, A. (2000) *Problems of inference for Azzalini's skew-normal distribution*, Journal of applied statistics. 27, pp. 859-870.

[23] Bazán, J. L., Branco, M. D. and Bolfarine, H. (2006), *A Skew Item Response Model*, Bayesian analysis, pp. 861-892.

[24] Henze, N. (1986), *A probabilistic representation of the skew-normal distribution*, Scandinavian journal of statistics. 13, pp. 271–275 .

[25] Sahu, S. K. (2002). Bayesian Estimation and Model Choice in Item Response Models, *Journal of Statistical Computation and Simulation, 72, 3, 2002, 217-232.*

[26] Sahu, S. K., Dey, D. K. and Branco, M. D. (2003), *A new class of multivariate skew distributions with applications to Bayesian regression models*, The Canadian Journal of Statistics. 31, pp. 129–150.

[27] Genton, M. (2004), Skew-elliptical distrivbutions and their applications, *Chapman & Hall, London.*

[28] Albert, J. (1992), *Bayesian estimation of normal ogive item response curves using Gibbs sampling*, Journal of educational statistics. 17, pp. 251–269.

[29] Bock, R. D. and Aitkin, M. (1981), *Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm*, Psychometrika. 46, pp. 433–459.

[30] Gamerman, D. and Lopes, H. F. (2006), Markov Chain Monte Carlo : Stochastic simulation for bayesian inference, second edition, *Chapman & Hall, London.*

[31] Arellano-Valle, R. B. and Azzalini, A. (2008), *The centred parameterization for the multivariate skew-normal distribution*, Journal of multivariate analysis. 99, pp. 1362–1382.

[32] R. J. Patz and B. W. Junker. (1999 a), *A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models*, Journal of educational and behavioral Statistics. 24, 2, pp. 146–178.

[33] Patz, R. J. and Junker, B. W. (1999 b), *The Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses*, Journal of educational and behavioral Statistics. 24, 4, pp. 342-366.

[34] Horn, R. A. and Johnson, C. R. (1991), *Topics in Matrix Analysis*. Cambridge university press, United Kingdom, first edition.

[35] Montgomery, D. C. (2004), Design and Analysis of Experiments, sixth edition, *Chapman & Hall, London*

[36] Poli, E. C. (2007), Longitudinal study in Mathematics: possibilities and reading of a reality of elementary school. Doctoral Thesis in Education. State University of Campinas. (In Portuguese)