Multidimensional multiple group IRT models with skew normal latent trait distributions

Juan L. Padilla[a], Caio L.N. Azevedo[a,*], Victor H. Lachos[a]

[a]*A Department of Statistics, University of Campinas, Brazil*

Abstract

Item response theory (IRT) models are one of the most important statistical tools for psychometric data analysis. Their applicability goes from educational assessment to biological essays. The IRT models combine, at least, two sets of unknown quantities: the latent traits (person parameters) and item parameters (related to measurement instruments of interest). The multidimensional item response theory (MIRT) models are quite useful to analyze data sets involving multiple skills or latent traits, which occurs in many of the applications. However, most of the works in the literature consider the usual assumption of multivariate (symmetric) normal distribution to the latent traits and do not deal with the multiple group framework (few groups with many of subjects in each one). They, in general, consider a limited number of model fit assessment tools, and do not investigate the measurement instrument dimensionality in a detailed way, while also dealing with the model nonidentifiability in a different way than that we presented here and only for one group model. In this work, we propose a MIRT multiple group model with multivariate skew normal distributions for modeling the latent traits of each group under the centered parameterization, presenting simple and feasible conditions for model identification. A full Bayesian approach for parameter estimation, structural selection (model comparison and determination of the dimensionality of the measurement instrument) and model fit assessment are developed through Markov Chain Monte Carlo (MCMC) algorithms. The developed tools are illustrated through the analysis of a real data set related to the first stage of the University of Campinas 2013 admission exam.

*Corresponding author

Email address: `cnaber@ime.unicamp.br` (Caio L.N. Azevedo)

Preprint submitted to Journal of Multivariate Analysis *May 30, 2017*

Keywords: Item response theory, multidimensional models, multivariate skew normal distribution, centered parameterization, Bayesian inference, MCMC algorithms, Model fit assessment.

---

1. Introduction

Item response theory (IRT) models are one of the most important psychometric tools for data analysis. Their applicability goes from educational assessment to biological essays. The IRT models combine, at least, two sets of unknown quantities: the latent traits (person parameters) and item parameters (related to measurement instruments of interest, that is, a cognitive test, genetic experiments or a psychiatric questionnaire, among others examples). The multidimensional item response theory (MIRT) models are quite useful to analyze data sets involving multiple skills or latent traits, which occurs in many of the applications. However, most of the works in the literature consider the usual assumption of multivariate (symmetric) normal distribution to the latent traits and do not deal with the multiple group framework (few groups with many subjects in each one). They also both do and do not consider a limited number of model fit assessment tools, and do not investigate the measurement instrument dimensionality in a detailed way, while also dealing with the model nonidentiability in a different way than that we presented here and only for one group model. Particularly, the assumption of multivariate (symmetric) normality is also considered for unidimensional models, with some underlying correlation structure for the latent traits, as in Andrade & Tavares (2005) and Azevedo et al. (2016), but not for MIRT models. Particularly, the above issues were not simultaneously considered in any work of the literature, to the best of our knowledge.

Bayesian inference, model identification and model fit assessment/model comparison tools are discussed in Béguin & Glas (2001) and Bolt & Lall (2003). A new MCMC algorithm is proposed in Fu et al. (2009). Some tools for testing dimensionality are discussed in Levy et al. (2009), Béguin & Glas (2001) and Bartolucci (2007). Torre & Patz (2005) discuss the gain in latent trait estimation when the underlying correlation structure of the latent traits is taken into account. In de Jong & Steenkamp (2003), a graded multilevel finite mixture MIRT model is presented with group-specific normal multivariate (symmetric) distributions along with developments about model identification, model fit assessment and model comparison through Bayesian inference. Sheng & Wikle (2008) presented a hierarchical one group MIRT model built from unidimensional models, along with developments about model identification, model fit assessment and model comparison through Bayesian inference. The work of Béguin & Glas (2001) presents a multiple-group MIRT model but all developments and applications are made for one group MIRT model, but in a different way that we consider in this work. Only the work of de Jong & Steenkamp (2003) deals with a multiple group MIRT model. More discussions about MIRT models can be found in Reckase (2009). To the best of our knowledge, none of the works in the literature consider a latent trait distribution different from the multivariate normal neither all the mentioned issues simultaneously.

In this work, we propose a new MIRT multiple group model with group specific multivariate skew normal distributions for the latent traits. We consider a slightly different version of the skew multivariate distribution under the centered parameterization developed by Arellano-Valle & Azzalini (2008). That is, a new parameterization of the multivariate skew normal distribution is introduced. Also, we explore two types of covariance matrix for the latent traits: a diagonal matrix and a non-structured covariance matrix. We present simple conditions for model identification, allowing to have either non-correlated or correlated latent factors and without imposing or imposing only few restrictions to the item parameters (which can be data and/or experimental design-driven). A full Bayesian approach for parameter estimation, structural selection (model comparison and the determination of the dimensionality of the measurement instrument) and model fit assessment is developed through MCMC algorithms. Approaches for comparison and assessment of the test dimensionality are proposed based on Bayesian measures of model complexity as in Spiegelhalter et al. (2002) and posterior predictive checking (see Levy et al., 2009). Also, mechanisms for measuring the global, per group and per item model fit assessment, as in Azevedo et al. (2012), are developed. The developed tools are illustrated through the analysis of a real data set related to 2013 first stage of the University of Campinas entrance exam.

This paper is outlined as follows. In Section 2, we present a new parameterization of the multivariate skew normal distribution, introduce the multiple group skew MIRT model and the model identifiability is discussed. In Section 3, the prior and posterior distributions are presented, the MCMC algorithm is given and the model fit assessment and model comparison tools are discussed. In Section 4, the analysis of a real data set is presented.

Finally, in Section 5, some additional comments are presented.

2. The Model

In this section, we present a new model which is a usual multidimensional three-parameter probit model (see Reckase, 2009) with the assumption of multivariate skew normal distribution under a new centered parameterization for the latent traits. First, we introduce this new parameterization and afterwards we propose a new MIRT model. Some useful and necessary notations are also introduced. Finally, the identification issues are discussed.

2.1. A new parameterization of the multivariate skew normal distributions

Similarly to the dichotomous unidimensional IRT (UIRT) models, for the dichotomous MIRT models, is necessary to fix the mean (vector) and the variance (covariance matrix) of the latent traits distribution and/or to impose some restrictions to the discrimination and difficulty parameters (see Reckase, 2009; Béguin & Glas, 2001; Azevedo et al., 2011). In the usual skew univariate normal distribution, the mean and variance can not be fixed without fixing the value of the asymmetry parameter, Azzalini (1985). Therefore, Azevedo et al. (2011) considered the centered version of the skew normal distribution to define a skew unidimensional IRT model, see also Azzalini (1985) and Pewsey (2000). For our model, we need to impose restrictions in the mean vector and covariance matrix of the latent traits distributions and, depending on the covariance structure, to the discrimination parameters (more details about the identification aspects are discussed in Subsection 2.3).

To the best of our knowledge, none of the parameterizations of the multivariate skew normal distributions provides mean vector and covariance matrices, simultaneously, free from the asymmetry parameters (see Genton, 2004) except those proposed by Arellano-Valle & Azzalini (2008). However, we will consider a slightly different version of this multivariate skew normal under the centered parameterization (MSNCP), since in our parameterization the respective density can be easily obtained, an useful stochastic representation can be always defined and there are no restrictions in the parameter space. Also, for the unidimensional case, the MSNCP distribution considered here becomes that proposed by Azzalini (1985), which was used by Azevedo et al. (2011) and Santos et al. (2013), within the IRT context. Therefore, our model is a generalization of those proposed by Santos et al. (2013), Azevedo et al. (2011) and Azevedo et al. (2012). The MSNCP proposed in this work is based on that presented by Lachos (2004), while the MSNCP distribution of Arellano-Valle & Azzalini (2008) is developed from the distribution proposed by Azzalini & Capitanio (1999).

First, let us present the definition of the multivariate skew normal distribution considered in Lachos (2004), henceforth denoted by MSN (multivariate skew normal) distribution.

Definition 1. *A D-dimensional random vector, say, $\boldsymbol{Z}$ has a MSN distribution if its density is given by:*

$$p_Z(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_D(\boldsymbol{z}; \boldsymbol{\mu}; \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{z} - \boldsymbol{\mu}))\mathbb{1}_{\mathbb{R}^D}(\boldsymbol{z}), \tag{1}$$

where $\phi_D(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of a D-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, Φ is the cdf of the standard (symmetric) normal distribution, $\boldsymbol{\mu}$ is the location parameter, $\boldsymbol{\Sigma}$ is the dispersion matrix and $\boldsymbol{\lambda}$ is the vector of the asymmetry parameters. Let us denote by $\boldsymbol{Z} \sim SN_D(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ a random vector that follows a d-dimensional MSN distribution. Notice that, if $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_D$, we have that $E(\boldsymbol{Z}) = \boldsymbol{\mu}_{\boldsymbol{Z}} = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$ and $Cov(\boldsymbol{Z}) = \boldsymbol{\Sigma}_{\boldsymbol{Z}} = \boldsymbol{I}_D - \boldsymbol{\mu}_{\boldsymbol{Z}}\boldsymbol{\mu}_{\boldsymbol{Z}}^\top$, where $\boldsymbol{I}_D$ stands for a D-dimensional identity matrix and $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}$. Therefore, this parameterization is not useful to build our MIRT model, since it does not allow to determinate the latent trait scale without fixing the asymmetry parameter (see also Azevedo et al., 2011; Santos et al., 2013). In addition, since we want to let the data indicate the entire behavior of the latent traits distributions, this parameterization is not useful, even fixing the location mean and the dispersion matrix. For further details about the MSN distributions, see Genton (2004) and Lachos (2004).

Now let us introduce our centered parametrization for the MSN distribution. The idea is similar to the unidimensional case (see Azevedo et al., 2012; Santos et al., 2013; Azzalini, 1985). First, we consider a random vector such that $\boldsymbol{Z} \sim SN_D(\boldsymbol{0}, \boldsymbol{I}_D, \boldsymbol{\lambda})$, and then, we define the following transformation:

$$\boldsymbol{\theta} = \boldsymbol{\mu}_{\boldsymbol{\theta}} + \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{1/2\top} \left[\boldsymbol{\Sigma}_{\boldsymbol{Z}}{}^{1/2\top}\right]^{-1} (\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}}), \tag{2}$$

where $(.)^{1/2}$ stands for the Cholesky decomposition, $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is the mean vector and $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$ is the covariance matrix. Let us define $\boldsymbol{\theta} \sim SNCP_D(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Psi}_{\boldsymbol{\theta}}, \boldsymbol{\delta}_{\boldsymbol{\theta}})$, where $SNCP_D$ represents a D-variate skew normal distribution under the centered parameterization and $\boldsymbol{\delta}_{\boldsymbol{\theta}} \equiv \boldsymbol{\delta}$. Therefore, in our parameterization, the mean vector and covariance matrix are directly defined and the identification restrictions can be easily considered. In addition, these quantities can be directly estimated. Finally, if $D = 1$, we have the distribution defined in Azevedo et al. (2012) and Santos et al. (2013). Also, we have:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = \begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \vdots \\ \mu_{\boldsymbol{\theta}_D} \end{bmatrix}, \boldsymbol{\Psi}_{\boldsymbol{\theta}} = \begin{bmatrix} \psi_{\theta_1} & \psi_{\theta_{12}} & \dots & \psi_{\theta_{1D}} \\ \psi_{\theta_{12}} & \psi_{\theta_2} & \dots & \psi_{\theta_{2D}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{1D}} & \psi_{\theta_{2D}} & \dots & \psi_{\theta_D} \end{bmatrix} \text{ and } \boldsymbol{\delta}_{\boldsymbol{\theta}} = \begin{bmatrix} \delta_{\theta_1} \\ \delta_{\theta_2} \\ \vdots \\ \delta_{\boldsymbol{\theta}_D} \end{bmatrix}. \quad (3)$$

Another way to write the expression (2) is

$$\boldsymbol{\theta} = \boldsymbol{\alpha}_{\boldsymbol{\theta}} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\top} \boldsymbol{Z}, \quad (4)$$

where $\boldsymbol{\alpha}_{\boldsymbol{\theta}} = \boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{1/2\top} \left(\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{1/2\top}\right)^{-1} \boldsymbol{\mu}_{\boldsymbol{Z}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{t} = \boldsymbol{\Psi}_{\boldsymbol{\theta}}^{1/2\top} \left(\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{1/2\top}\right)^{-1}$. Therefore, by using some properties of the MSN distribution, see Lachos (2004), we may conclude that $\boldsymbol{\theta} \sim SN_D\left(\boldsymbol{\alpha}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}, \boldsymbol{\lambda}\right)$.

Theorem 2. *From equations (1) to (4) and from some properties of the MSN distribution, see Lachos (2004), it is possible to conclude that the density of the MSNCP is given by*

$$p(\boldsymbol{\theta}|\boldsymbol{\eta}_{\boldsymbol{\theta}}) = 2\phi_D\left(\boldsymbol{\theta}; \boldsymbol{\alpha}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right) \Phi\left\{\boldsymbol{\lambda}_{\boldsymbol{\theta}}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\top} \left(\boldsymbol{\theta} - \boldsymbol{\alpha}_{\boldsymbol{\theta}}\right)\right\} \mathbb{1}_{\mathbb{R}^D}(\boldsymbol{\theta}), \quad (5)$$

where $\boldsymbol{\eta}_{\boldsymbol{\theta}} = \left(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \text{vech}\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}}\right), \boldsymbol{\delta}_{\boldsymbol{\theta}}\right)^{\top}$, *vech* extracts all elements in and below the main diagonal.

The following lemma regarding the distribution of an affine transformation of the MSN distribution will be quiet useful for finding the marginal distributions of a vector distributed as a MSNCP. Its proof can be found in Lachos (2004).

Lemma 3. *Let* $\boldsymbol{\theta} \sim SN_D\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Psi}_{\boldsymbol{\theta}}, \boldsymbol{\delta}_{\boldsymbol{\theta}}\right)$, *and* $\boldsymbol{C}$ *a* $(D \times k)$ *full rank matrix. Then* $\boldsymbol{C}^{\top}\boldsymbol{\theta} \sim SN_k\left(\boldsymbol{C}^{\top}\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{C}^{\top}\boldsymbol{\Psi}_{\boldsymbol{\theta}}\boldsymbol{C}, \boldsymbol{\delta}_*\right)$, *where* $\boldsymbol{\delta}_* = \left(\boldsymbol{C}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{C}\right)^{-\frac{1}{2}\top}$ $\boldsymbol{C}^{\top}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{\frac{1}{2}\top}\boldsymbol{\delta}_{\boldsymbol{\theta}}$.

Theorem 4. *Let us consider two sets of random variables* $(\theta_1, \dots, \theta_q)$ *and* $(\theta_{q+1}, \dots, \theta_D)$ *forming the vectors* $\boldsymbol{\theta}^{(1)} = (\theta_1, \dots, \theta_q)^{\top}$ *and* $\boldsymbol{\theta}^{(2)} = (\theta_{q+1}, \dots, \theta_D)^{\top}$. *These variables form the random vector*

$$\boldsymbol{\theta} = \left[\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right]^\top = (\theta_1, \ldots, \theta_q, \theta_{q+1}, \ldots, \theta_D)^\top . \quad (6)$$

Now let us assume that the D variables have a joint skew normal distribution under the centered parameterization with null mean vector, identity covariance matrix and skewness coefficient $\boldsymbol{\delta_\theta} = \left[\boldsymbol{\delta}_{\boldsymbol{\theta}}^{(1)}, \boldsymbol{\delta}_{\boldsymbol{\theta}}^{(2)}\right]^\top$. Then the distribution of the partitioned vector $\boldsymbol{\theta}^{(i)}$, for $i = 1, 2$ and such that $D_1 + D_2 = D$, is $SNCP_{D_{(i)}}\left(\mathbf{0}, \boldsymbol{I}_{D_{(i)}}, \boldsymbol{\delta}_{\boldsymbol{\theta}}^{(i)}\right)$. With $D_{(i)}$ being the rank of the vector $\boldsymbol{\theta}^{(i)}$.

Proof. In order to prove this result, recall the equivalence between the MSN and MSNCP distributions. That is, if $\boldsymbol{\theta} \sim SNCP_D(\boldsymbol{\mu_\theta}, \boldsymbol{\Psi_\theta}, \boldsymbol{\delta})$ then $\boldsymbol{\theta} \sim SN_D\left(\boldsymbol{\alpha_\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^\top \boldsymbol{\Sigma_\theta}, \boldsymbol{\lambda}\right)$ where $(\boldsymbol{\alpha_\theta}, \boldsymbol{\Sigma_\theta}, \boldsymbol{\lambda})$ were already defined. Taking $\boldsymbol{C}^\top$ from Lemma 3 to be the matrix such that it chooses from $\boldsymbol{\theta}$ the elements to form the partition i we arrive to the desired conclusion. □

Figure 1 presents the contour plots for a bivariate MSNCP distribution (that is, $D = 2$) for different values of the correlation between θ_1 and θ_2 (Corre(θ_1, θ_2)), and $\boldsymbol{\delta_\theta}$ with $\boldsymbol{\mu} = (0, 0)^\top$ and $\psi_{\boldsymbol{\theta}_{11}} = \psi_{\boldsymbol{\theta}_{22}} = 1$. It can be seen that the contours do not depict an elliptic behaviour, except for the cases in which the skewness coefficient is the same for both components.

An important result is the stochastic representation of density (5), which can be deduced from that presented in Lachos (2004) for a random vector with MSN distribution. That is, if $\boldsymbol{\theta} \sim SNCP_D(\boldsymbol{\mu_\theta}, \boldsymbol{\Psi_\theta}, \boldsymbol{\delta_\theta})$, then

$$\boldsymbol{\theta}|(T = t) \sim N_D\left(\boldsymbol{\alpha_\theta} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^\top \boldsymbol{\delta_\theta} t, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^\top\left(\boldsymbol{I}_D - \boldsymbol{\delta_\theta}\boldsymbol{\delta}_{\boldsymbol{\theta}}^\top\right)\boldsymbol{\Sigma_\theta}\right), \quad (7)$$

where $\boldsymbol{\alpha_\theta}$ and $\boldsymbol{\Sigma_\theta}$ are as defined before, $T \sim HN(0, 1)$ and $N_D(\boldsymbol{\mu}, \boldsymbol{\Psi})$ stands for a D-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Psi}$. More details about the MSNCP can be found in Padilla (2014). Next, we present our multiple group MIRT model and the related identification issues.

2.2. A new MIRT multiple group model with skew normal latent traits distributions under the centered parameterization

One or more different tests are administered to the (randomly selected) subjects of each group. The tests have common items and the structure can be recognized as an incomplete block design (see Montgomery, 2004). We will assume that each group has a reasonable number of subjects. In

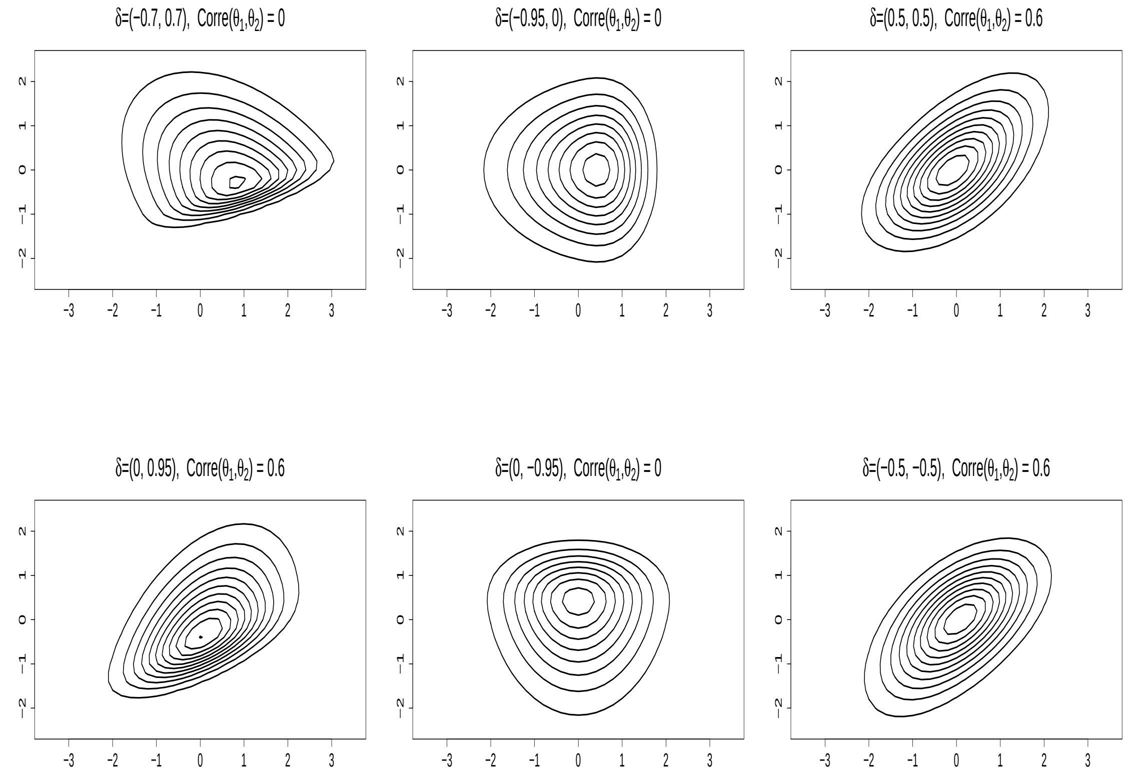

Figure 1: Contour plots of the bivariate MSNCP for different values of Corre(θ_1, θ_2) and δ.

summary, we are dealing with a set of n subjects clustered in K groups, with n_k subjects in group k, and $n = \sum_{k=1}^{K} n_k$. The subjects of each 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_{djk} is the latent trait of subject j $(j = 1, \ldots, n_k,)$ belonging to group k $(k = 1, \ldots, K)$, related to the dimension d , $(d = 1, ..., D)$, $\boldsymbol{\theta}_{.jk} = (\theta_{1jk}, \ldots, \theta_{DjK})^\top$ is the vector of the latent traits of subject j of group k, $\boldsymbol{\theta}_{..k} = (\boldsymbol{\theta}_{.1k}, ..., \boldsymbol{\theta}_{.n_k k})$ is the vector of all latent traits of the subjects of group k and $\boldsymbol{\theta}_{...} = (\boldsymbol{\theta}_{..1}, \ldots, \boldsymbol{\theta}_{..K})^\top$ is the vector with all latent traits; Y_{ijk} is the response of the subject j of group k to item i $(i = 1, \ldots, I_k)$, $\boldsymbol{Y}_{.jk} = (Y_{1jk}, \ldots, Y_{I_k jk})^\top$ is the response vector of subject j of group k, $\boldsymbol{Y}_{..k} = (\boldsymbol{Y}_{.1k}^\top, \ldots, \boldsymbol{Y}_{.n_k k}^\top)^\top$ is the response vector of all subjects of group k, $\boldsymbol{Y}_{...} = (\boldsymbol{Y}_{.1.}^\top, \ldots, \boldsymbol{Y}_{.n_k.}^\top)^\top$ is the whole response set and $y_{ijk}, \boldsymbol{y}_{.jk}, \boldsymbol{y}_{..k}, \boldsymbol{y}_{...}$ are the respective observed values; $\boldsymbol{\zeta}_i$ is the vector of parameters of the item i, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^\top, \ldots, \boldsymbol{\zeta}_I^\top)^\top$ is the whole set of item parameters, $\boldsymbol{\eta}_{\boldsymbol{\theta}_k}$ is the vector with the population parameters of group k and $\boldsymbol{\eta}_{\boldsymbol{\theta}} = (\boldsymbol{\eta}_{\boldsymbol{\theta}_1}^\top, \ldots, \boldsymbol{\eta}_{\boldsymbol{\theta}_K}^\top)^\top$ is the whole set of population parameters.

The MIRT multiple group model with multivariate skew normal distribution under the centered parameterization (MSNCP) is given by

$$\begin{aligned}
Y_{ijk} \mid (\boldsymbol{\theta}_{.jk}, \boldsymbol{\zeta}_i) &\sim \text{Bernoulli}(P_{ijk}), \\
P_{ijk} = P(Y_{ijk} = 1 \mid \boldsymbol{\theta}_{.jk}, \boldsymbol{\zeta}_i) &= c_i + (1 - c_i)\Phi\left(\boldsymbol{a}_i^\top \boldsymbol{\theta}_{.jk} - b_i\right), \\
&= c_i + (1 - c_i)\Phi\left(\sum_{d=1}^{D} a_{id}\theta_{djk} - b_i\right), \\
\theta_{.jk} \mid \boldsymbol{\eta}_{\theta_k} &\sim SNCP_D\left(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}, \boldsymbol{\delta}_{\boldsymbol{\theta}_k}\right),
\end{aligned}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}_k}$, $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}$ and $\boldsymbol{\delta}_{\boldsymbol{\theta}_k}$ are as in (3), considering the index k, $\boldsymbol{\zeta}_i = (\boldsymbol{a}_i^\top, b_i, c_i)^\top$, $\boldsymbol{a}_i = (a_{i1}, ..., a_{iD})^\top$, $\boldsymbol{\eta}_{\boldsymbol{\theta}_k} = (\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, vech(\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}), \boldsymbol{\delta}_{\boldsymbol{\theta}_k})$. For more details concerning the interpretation of item parameters and for the so-called multidimensional item parameters, the reader is referred to Reckase (2009). Notice that for $K = 1$, we also have a new model, that is, a one group MIRT model with a skew normal multivariate distribution under the centered parameterization. As mentioned before, two structures for the covariance matrix will be considered in this work. In the first case, we assume that the covariance matrix of the reference group is an identity matrix, that is, $\boldsymbol{\Psi}_{\boldsymbol{\theta}_1} = \boldsymbol{I}_D$, while the other matrices are diagonal, that is, $\boldsymbol{\Psi}_{\theta_k} = \text{diag}\left(\psi_{\theta_{k1}}, ..., \psi_{\theta_{kD}}\right), k = 2, ..., K$. In the second case, the covariance matrix of the reference group is assumed to be a correlation one, whereas for the other groups, a usual covariance matrix is considered.

2.3. Model identification

Similarly to the usual multiple group model (MGM), it is necessary to establish a reference group, for example, the first. To accomplish that, we can fix the mean vector and the covariance matrix of the first group in some specific values and/or to impose restrictions to the difficulty and discrimination parameters.

In this work, we consider two scenarios of interest, concerning the covariance structure of the latent traits, as mentioned before, and each situation must be treated in a different way, in terms of model identification, as we will show ahead.

For the class of MIRT models, two conditions must hold, in order to ensure the model identification, namely: invariance against linear transformations (IALT) and invariance against rotations (IAR) (see Rivers, 2003; Matos, 2008). For MIRT models, these aspects are related to the linear predictor, that is, to $\sum_{d=1}^{D} a_{id}\theta_{djk} - b_i = \boldsymbol{a}_i^{\top}\boldsymbol{\theta}_{.jk} - b_i$. Being $\boldsymbol{A}$ a non-singular real matrix and $\boldsymbol{\beta}$ a real vector, we have:

$$
\begin{aligned}
\boldsymbol{a}_i^{\top}\boldsymbol{\theta}_{.jk} - b_i &= \boldsymbol{a}_i^{\top}(\boldsymbol{\theta}_{.jk} - \boldsymbol{\beta} + \boldsymbol{\beta}) - b_i \\
&= \sum_{d=1}^{D} a_{id}\theta^{*}_{djk} - b^{*}_i = \boldsymbol{a}_i^{\top}\boldsymbol{\theta}^{*}_{.jk} - b^{*}_i,
\end{aligned}
\tag{8}
$$

where $\boldsymbol{\theta}^{*}_{.jk} = \boldsymbol{\theta}_{.jk} + \boldsymbol{\beta}$ and $b^{*}_i = \boldsymbol{a}_i^{\top}\boldsymbol{\beta} + b_i$. The second type of transformation, that is, the IAR, is related, being $\boldsymbol{A}$ a orthogonal matrix, to

$$
\begin{aligned}
\boldsymbol{a}_i^{\top}\boldsymbol{\theta}_{.jk} - b_i &= \boldsymbol{a}_i\boldsymbol{A}\boldsymbol{A}^{\top}\boldsymbol{\theta}_{.jk} - b_i \\
&= \sum_{d=1}^{D} a^{*}_{id}\theta^{*}_{djk} - b_i = (\boldsymbol{a}^{*}_i)^{\top}\boldsymbol{\theta}^{*}_{.jk} - b_i,
\end{aligned}
\tag{9}
$$

where $\boldsymbol{\theta}^{*}_{.jk} = \boldsymbol{A}\boldsymbol{\theta}_{.jk}$ and $\boldsymbol{a}^{*}_i = \boldsymbol{A}\boldsymbol{a}_i$. Notice that, combining the two sets of transformations and due to some proprieties of the multivariate skew normal distribution, see Lachos (2004), we have that

$$
\boldsymbol{\theta}^{*}_{.jk}|\boldsymbol{\eta}_{\boldsymbol{\theta}_k} \sim SNCP_D(\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, \boldsymbol{A}\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}\boldsymbol{A}^{\top}, \boldsymbol{\delta}_{\boldsymbol{\theta}_k}).
\tag{10}
$$

Therefore, these sort of transformations change the mean vector and the covariance matrix, but does not affect the vector of the asymmetry parameters of the latent trait distribution. Then, the idea to identify the model, is to restrict the mean vector and the covariance matrix of the reference group (similarly to the unidimensional MGM) and/or to restrict some (or all) item parameters belonging to the test applied to the reference group, in such way that transformations as in (8) and (9) are no longer possible. Therefore, by combining these restrictions with the linking design (a structure of common items among the tests) the model will be identified, see Bock & Zimowski (1997) and Santos et al. (2013) for further details. When $D = 1$, that is, for the one group MIRT model, the restrictions are similar see Azevedo et al. (2011) but, in this case, the reference group is the unique group.

To solve the problem of IALT, it suffices to fix the mean vector of the latent trait distribution of the reference group, whatever the selected structure of the covariance matrix, as we can see from (10). On the other hand, the approach for solving the IAR depends on the structure adopted for the covariance matrix.

Diagonal covariance matrix: uncorrelated factors

In this case, as explained in Béguin & Glas (2001), some additional restrictions on the item discrimination parameters are necessary. Essentially, since any orthogonal transformation, as in (9) is feasible. For example, we can impose that some items load in specific latent traits dimensions and/or that they are positively or negatively related to some specific dimensions. This choice depends on the situation. For example, in an educational assessment, it is reasonable to expect that each item either loads in some specific dimensions (or even in all of them) and/or they are positive correlated to some (or all) latent traits (since it is not expected that having high values in any latent trait will decrease the probability of correct response). On the other hand, in psychiatric studies, the items of a questionnaire (measurement instrument) can be grouped in such a way that they are related to specific symptoms. Therefore, each item will load in only one specific dimension which, in its turn, is related to some specific symptoms. Another situation can be where a specific symptom prevents the presence of another symptom. Therefore, in this case, it is expected that the higher the latent trait in the former symptom, smaller the probability of manifesting the latter. In this case, some discrimination parameters must be positive and others negative. In conclusion, by assuming a diagonal covariance matrix for the groups, some particular interactions between item and latent traits need to be considered. However, in general, this is not a difficult task and can be drawn from the data and/or from the experiment and/or from the specialist. In conclusion, if we fix some discrimination parameters to zero for some items or fix the signal for some of them (related to the reference group) the model is identified. Naturally, if we consider these two sets of restrictions, simultaneously, the model is also identified.

Full covariance matrix: correlated factors

Recall that, is this the case, the covariance matrix of the reference group is, in fact, a correlation matrix. In this case, orthogonal transformations as in (9) related to the covariance matrix of the reference group, are not possible. As a result, the model is identified, provided that all correlations are different from zero, regardless any restrictions imposed on the item parameters. If at least one is equals to zero, the model is no longer identified, and we have a similar pattern to that of the previous section. We shall prove this result in the sequence.

Proposition 5. *Let us suppose an orthogonal matrix, say,* $\boldsymbol{R}$*, different from the identity or a permutation matrix, and be* $\boldsymbol{\Gamma}$ *a correlation matrix. Therefore, the product* $\boldsymbol{\Gamma}^* = \boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^\top$ *is such that at least one element of its main diagonal is different from 1. That is, the matrix* $\boldsymbol{\Gamma}^*$*, is not a correlation matrix.*

Proof. We seek to prove that, for any orthogonal $\boldsymbol{R}$, different from the identity or a permutation matrix, and any correlation matrix, $\boldsymbol{\Gamma}$, the matrix $\boldsymbol{\Gamma}^* = \boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^\top$ will be no longer a correlation matrix. We will present the proof for the cases $D = 2$ and $D = 3$. For the other cases, the proof is straightforward. However, notice that, in general, it is usual to consider, at most, a five dimensional model, that is, $D = 5$. To prove that $\boldsymbol{\Gamma}^*$ is not a correlation matrix, it suffices to prove that it has at least one element in its main diagonal different from one.

- $\mathbf{2 \times 2}$ matrices. We have that:

$$
\begin{aligned}
\boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^\top &= \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \gamma \\ \gamma & 0 \end{pmatrix} \right) \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} r_{12}\gamma & r_{11}\gamma \\ r_{22}\gamma & r_{21}\gamma \end{pmatrix} \begin{pmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2r_{11}r_{12}\gamma & r_{12}r_{21}\gamma + r_{11}r_{22}\gamma \\ r_{12}r_{21}\gamma + r_{11}r_{22}\gamma & 2r_{11}r_{12}\gamma \end{pmatrix}.
\end{aligned}
$$

Then, we want to prove that there are not real numbers r_{11}, r_{12} and γ, such that

$$\begin{aligned} 2\gamma r_{12}r_{11} &= 0; \\ 2\gamma r_{21}r_{22} &= 0, \end{aligned} \tag{11}$$

under the restrictions:

$$\begin{aligned} r_{11}^2 + r_{12}^2 &= 1; \\ r_{21}^2 + r_{22}^2 &= 1; \\ r_{11}r_{21} + r_{12}r_{22} &= 0, \end{aligned}$$

which are valid since since $\boldsymbol{R}$ is an orthogonal matrix. Let us assume that there are real numbers r_{11}, r_{12} and γ such that (11) holds. However, this is only possible if $\gamma = 0$, which leads to the matrix $\boldsymbol{\Gamma}$ being an identity matrix, which, in its turn, clearly violates one of the assumptions of the theorem; or if at least two elements of the matrix $\boldsymbol{R}$ are equal to zero, which also violates one of the assumptions of the theorem, since that $\boldsymbol{R}$ would be either an identity matrix or a permutation matrix. Therefore, there are no real numbers r_{11}, r_{12} and γ such that (11) holds, which implies that $\boldsymbol{\Gamma}^*$ can not be a correlation matrix. It is worthwhile to mention that the identity matrix does not change the covariance matrix and the permutation matrix can only permutate the dimension positions in the covariance matrix, see Equation (10). Therefore, these two cases are not relevant for the model identification. Also, the results obtained in the simulation studies (see Padilla, 2014) indicates that the model is identified, under the aforementioned restrictions.

- For **3 × 3** matrices, we have:

$$\boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} 1 & \gamma_1 & \gamma_2 \\ \gamma_1 & 1 & \gamma_3 \\ \gamma_2 & \gamma_3 & 1 \end{pmatrix},$$

the elements of the main diagonal, $\boldsymbol{R}\boldsymbol{\Gamma}\boldsymbol{R}^\top$ will be:

$$\begin{aligned} &1 + r_{12}r_{11}\gamma_1 + r_{11}r_{13}\gamma_2 + r_{12}r_{13}\gamma_3; \\ &1 + r_{21}r_{22}\gamma_1 + r_{21}r_{23}\gamma_2 + r_{23}r_{22}\gamma_3; \\ &1 + r_{31}r_{32}\gamma_1 + r_{31}r_{33}\gamma_2 + r_{32}r_{33}\gamma_3. \end{aligned} \tag{12}$$

We seek to prove that, provided $\boldsymbol{R}$ is a orthogonal matrix and $\gamma_i \neq 0$, $\forall i = 1,2,3$, there are not real numbers r_{ij}, $\forall i,j = 1,2,3$ and γ_i, $\forall i = 1,2,3$, such that

$$\begin{aligned} r_{12}r_{11}\gamma_1 + r_{11}r_{13}\gamma_2 + r_{12}r_{13}\gamma_3 = 0; \\ r_{21}r_{22}\gamma_1 + r_{21}r_{23}\gamma_2 + r_{23}r_{22}\gamma_3 = 0; \\ r_{31}r_{32}\gamma_1 + r_{31}r_{33}\gamma_2 + r_{32}r_{33}\gamma_3 = 0, \end{aligned} \tag{13}$$

holds. However, similarly to the 2×2 matrices, the orthogonal matrices $\boldsymbol{R}$ that satisfy (13), are the permutation and the identity matrices, which would be a contradiction. Therefore, the result follows. □

Then, in this case, that is, when the covariance matrix of the reference group is a correlation matrix with non zero off diagonal elements, it is not necessary to impose restrictions on the discrimination parameters.

In conclusion, the structure assumed for the latent traits distribution is:

$$\begin{aligned} \boldsymbol{\theta}_{.j1}|\boldsymbol{\eta}_{\theta_1} &\sim SNCP_D\left(\mathbf{0}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_1}, \boldsymbol{\delta}_{\boldsymbol{\theta}_1}\right), \\ \boldsymbol{\theta}_{.jk}|\boldsymbol{\eta}_{\theta_k} &\sim SNCP_D\left(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}, \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}, \boldsymbol{\delta}_{\boldsymbol{\theta}_k}\right), k = 2, ..., K. \end{aligned}$$

Two situations, as mentioned before, are considered for the covariance matrices of the latent traits: 1) $\boldsymbol{\Psi}_{\boldsymbol{\theta}_1} = \boldsymbol{I}_D; \boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \text{diag}\left(\psi_{\theta_{k11}}, ..., \psi_{\theta_{kDD}}\right), k = 2, ..., K$ and 2) for $k = 1, ..., K$.

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}_1} = \begin{bmatrix} 1 & \psi_{\theta_{112}} & \dots & \psi_{\theta_{11D}} \\ \psi_{\theta_{112}} & 1 & \dots & \psi_{\theta_{12D}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{11D}} & \psi_{\theta_{12D}} & \dots & 1 \end{bmatrix}; \boldsymbol{\Psi}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \psi_{\theta_{k1}} & \psi_{\theta_{k12}} & \dots & \psi_{\theta_{k1D}} \\ \psi_{\theta_{k12}} & \psi_{\theta_{k2}} & \dots & \psi_{\theta_{k2D}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\theta_{k1D}} & \psi_{\theta_{k2D}} & \dots & \psi_{\theta_{kD}} \end{bmatrix}.$$

While in the situation 1) it is necessary to impose additional restrictions in the discrimination parameters of some items, besides the restrictions imposed on the latent traits distribution of the reference group, in the situation 2) no further restrictions are necessary.

3. Bayesian inference and Gibbs sampling algorithm

Despite the prior distributions adopted, the structure for the covariance matrices and the likelihood considered (original or augmented), the marginal posterior distributions of interest are not analytically obtainable. The use of MCMC algorithms, however, enables one to obtain numerical approximations. Some MCMC algorithms were compared in Padilla et al. (2017), according to Effective Sample Size criterion, see Sahu (2002). In this work, we consider the selected algorithm by that work, which corresponds to the augmented data scheme proposed by Sahu (2002), combined with the convergence acceleration algorithm proposed by Gonzalez (2004). More details of this algorithm can be found in Appendix 8.

3.1. Augmented likelihood and prior and posterior distributions

The initial step, following Sahu (2002), is to define two sets of augmented variables, say $\boldsymbol{Z}_{\ldots} = (Z_{111}, ..., Z_{I_k n_k K})^\top$ and $\boldsymbol{U}_{\ldots} = (U_{111}, ..., U_{I_k n_k K})^\top$, such that: $Z_{ijk} \overset{i.i.d.}{\sim} N(\boldsymbol{a}_i^\top \boldsymbol{\theta}_{.jk} - b_i, 1) \perp U_{ijk} \overset{i.i.d.}{\sim} \text{Bernoulli}(c_i), \forall i, j, k$, here *i.i.d.* means that the variables are independent and identically distributed.

To handle incomplete block designs, an indicator variable is defined that defines the set of administered items for each occasion and subject. This indicator variable is defined as follows,

$$I_{ijk} = \begin{cases} 1, & \text{item } i \text{ administered for subject } j \text{ of group } k, \\ 0, & \text{missing by design.} \end{cases}$$

The nonselective missing responses due to uncontrolled events are marked, as nonresponse or errors in recording data, by another indicator, which is defined as,

$$V_{ijk} = \begin{cases} 1, & \text{observed response of subject } j \text{ of group } k \text{ on item } i, \\ 0, & \text{otherwise.} \end{cases}$$

It is assumed that the missing data are missing at random (MAR), such that the distribution of patterns of missing data does not depend on the unobserved data. When the MAR assumption does not hold and the missing data cannot be ignored, a missing data model can be defined to model explicitly the pattern of missingness. In case of MAR, the observed data can be used to make valid inferences about the model parameters.

To ease the notation, let the indicator matrix $\mathbf{I} = (I_{111}, ..., I_{I_K n_k K})^\top$ represent both cases of missing data (which can not be confounded with the identity matrix $\boldsymbol{I}_{(.)}$).

Therefore, using the usual conditional independence assumptions, we have that the augmented likelihood is given by:

$$
\begin{aligned}
p(\boldsymbol{z}_{...}, \boldsymbol{u}_{...}|\boldsymbol{\theta}_{...}, \boldsymbol{\zeta}, \boldsymbol{y}_{...}) \;\propto\; & \prod_{k=1}^{K}\prod_{j=1}^{n_k}\prod_{i|I_{ijk=1}} \left\{ \exp\left\{-0.5\left(z_{ijk} - \boldsymbol{a}_i^\top \boldsymbol{\theta}_{.jk} + b_i\right)^2\right\}\right. \\
& \times \left. c_i^{u_{ijk}}(1-c_i)^{1-u_{ijk}} \mathbb{1}_{\left(z_{ijk}, u_{ijk}, y_{ijk}\right)} \right\}, \qquad (14)
\end{aligned}
$$

where $\boldsymbol{z}_{...} = (z_{111}, ..., z_{I_k n_k K})^\top$ and $\boldsymbol{u}_{...} = (u_{111}, ..., u_{I_k n_k K})^\top$. Here $\mathbb{1}_{\left(z_{ijk}, u_{ijk}, y_{ijk}\right)}$ stands for the indicator function representing the sample space defined in Sahu (2002). That is, if $y_{ijk} = 0$, we have that $u_{ijk} = 0$ and z_{ijk} must be negative, that is, $Z_{ijk} \sim N(\boldsymbol{a}_i^\top \boldsymbol{\theta}_{.jk} - b_i, 1)\mathbb{1}_{(z_{ijk}<0)}$. If $y_{ijk} = 1$ and, if $u_{ijk} = 0$, then $Z_{ijk} \sim N(\boldsymbol{a}_i^\top \boldsymbol{\theta}_{.jk} - b_i, 1)\mathbb{1}_{(z_{ijk}\geq 0)}$, otherwise $Z_{ijk} \sim N(\boldsymbol{a}_i^\top \boldsymbol{\theta}_{.jk} - b_i, 1)$. Once z_{ijk} has been sampled, we verify if it is negative if it is, then we simply set $u_{ijk} = 1$. Otherwise u_{ijk} is drawn from a Bernoulli(c_i).

The joint prior distribution of the parameters is given by: $p(\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = p(\boldsymbol{\theta}|\boldsymbol{\eta}_{\boldsymbol{\theta}})p(\boldsymbol{\eta}_{\boldsymbol{\theta}}|\boldsymbol{\eta}_{\boldsymbol{\eta}})p(\boldsymbol{\zeta}|\boldsymbol{\eta}_{\boldsymbol{\zeta}})$
where $\boldsymbol{\eta}_{\boldsymbol{\eta}}$ and $\boldsymbol{\eta}_{\boldsymbol{\zeta}}$ are the hyperparameters associated with $\boldsymbol{\eta}_{\boldsymbol{\theta}}$ and $\boldsymbol{\eta}_{\boldsymbol{\zeta}}$, respectively.

The prior distribution of the latent traits will be considered through the stochastic representation given by (7), that is:

$$
\begin{aligned}
p(\boldsymbol{\theta}_{...}|\boldsymbol{t}_{..}, \boldsymbol{\eta}_{\boldsymbol{\theta}}) &= p(\boldsymbol{\theta}_{...}|\boldsymbol{t}_{..}, \boldsymbol{\eta}_{\boldsymbol{\theta}})p(\boldsymbol{t}) = \prod_{k=1}^{K}\prod_{j=1}^{n_k} p(\boldsymbol{\theta}_{.jk}|t_{jk}, \boldsymbol{\eta}_{\boldsymbol{\theta}_k})p(t_{jk}) \\
&\propto \prod_{k=1}^{K}\prod_{j=1}^{n_k}\left\{ \exp\left\{-0.5\left(\boldsymbol{\theta}_{.jk} - \boldsymbol{\mu}^*_{\boldsymbol{\theta}_{jk}}\right)^\top \left(\boldsymbol{\Psi}^*_{\boldsymbol{\theta}_k}\right)^{-1}\left(\boldsymbol{\theta}_{.jk} - \boldsymbol{\mu}^*_{\boldsymbol{\theta}_{jk}}\right)\right\} \mathbb{1}_{\mathbb{R}^D}(\boldsymbol{\theta}_{.jk}) \right. \\
&\times \left. \exp\left\{-\frac{t_{jk}^2}{2}\right\} \mathbb{1}_{(0,\infty)}(t_{jk})\right\},
\end{aligned}
$$

where $\boldsymbol{t}_{..} = (t_{11}, ..., t_{n_K K})^\top$, $\boldsymbol{\mu}^*_{\boldsymbol{\theta}_{jk}} = \boldsymbol{\alpha}_{\boldsymbol{\theta}_k} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}_k}\boldsymbol{\delta}_{\boldsymbol{\theta}_k} t_{jk}$, $\boldsymbol{\Psi}^*_{\boldsymbol{\theta}_k} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}_k}^\top\left(\boldsymbol{I} - \boldsymbol{\delta}_{\boldsymbol{\theta}_k}\boldsymbol{\delta}_{\boldsymbol{\theta}_k}^\top\right)\boldsymbol{\Sigma}_{\boldsymbol{\theta}_k}$, $\boldsymbol{\alpha}_{\boldsymbol{\theta}_k} = \boldsymbol{\mu}_{\boldsymbol{\theta}_k} - \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{1/2\top}\left(\boldsymbol{\Sigma}_{\boldsymbol{Z}_k}^{1/2\top}\right)^{-1}\boldsymbol{\mu}_{\boldsymbol{Z}_k}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_k}^\top = \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{1/2\top}\left(\boldsymbol{\Sigma}_{\boldsymbol{Z}_k}^{1/2\top}\right)^{-1}$. Then, the joint prior distribution for $(\boldsymbol{\theta}_{...}^{\top}, \boldsymbol{t}_{..}^{\top}, \boldsymbol{\zeta}^{\top}, \boldsymbol{\eta}_{\boldsymbol{\theta}}^{\top})^{\top}$, assumed here, is

$$
\begin{aligned}
p(\boldsymbol{\theta}_{...}, \boldsymbol{t}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\boldsymbol{\theta}}) &\propto p(\boldsymbol{\theta}_{...}|\boldsymbol{t}_{..}, \boldsymbol{\eta}_{\boldsymbol{\theta}})p(\boldsymbol{t}_{..})p(\boldsymbol{\zeta}|\boldsymbol{\eta}_{\boldsymbol{\zeta}})p(\boldsymbol{\eta}_{\boldsymbol{\theta}}) \\
&= \prod_{k=1}^{K}\prod_{j=1}^{n_k}\left\{p(\boldsymbol{\theta}_{.jk}|t_{jk}, \boldsymbol{\eta}_{\boldsymbol{\theta}_k})p(t_{jk})\right\}\prod_{k=1}^{k}p(\boldsymbol{\eta}_{\boldsymbol{\theta}_k})\prod_{i=1}^{I}\left\{p(\boldsymbol{\zeta}_i|\boldsymbol{\eta}_{\boldsymbol{\zeta}})\right\} \\
&= \prod_{k=1}^{K}\prod_{j=1}^{n_k}\left\{p(\boldsymbol{\theta}_{.jk}|t_{jk}, \boldsymbol{\eta}_{\boldsymbol{\theta}_k})p(t_{jk})\right\}\prod_{k=1}^{K}\left\{p(\boldsymbol{\mu}_{\boldsymbol{\theta}_k})p(\boldsymbol{\Psi}_{\boldsymbol{\theta}_k})p(\boldsymbol{\delta}_{\boldsymbol{\theta}_k})\right\} \\
&\times \prod_{i=1}^{I}\left\{p(\boldsymbol{a}_i, b_i)p(c_i)\right\}, \qquad (15)
\end{aligned}
$$

where $\boldsymbol{\eta}_{\boldsymbol{\zeta}}$ are the hyperparameters associated with the vector $\boldsymbol{\zeta}$. We assume that $\boldsymbol{\mu}_{\boldsymbol{\theta}_k} \overset{i.i.d.}{\sim} N_D\left(\boldsymbol{\mu}_{\boldsymbol{\mu}}, \boldsymbol{\Psi}_{\boldsymbol{\mu}}\right)$, $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} \overset{i.i.d.}{\sim} IW(\tau, \boldsymbol{\Psi}_{\boldsymbol{\Psi}})$, $k = 1, 2, ..., K$, where $IW(\tau, \boldsymbol{\Psi}_{\boldsymbol{\Psi}})$ stands for a Inverse-Wishart distribution with degrees of freedom κ and dispersion matrix $\boldsymbol{\Psi}_{\boldsymbol{\Psi}}$. The priors for the asymmetry vectors are based on the beta distribution so let $\delta_{d\boldsymbol{\theta}_k}$ denote the element on the dth position in the vector $\boldsymbol{\delta}_{\boldsymbol{\theta}_k}$ and $\boldsymbol{\delta}_{(-d)\boldsymbol{\theta}_k}$ denote the vector $\boldsymbol{\delta}_{\boldsymbol{\theta}_k}$ removing the element on position d, this way

$$
\begin{aligned}
p(\delta_{d\boldsymbol{\theta}_k}|\alpha_{\delta_1}, \alpha_{\delta_2}) \propto &\left(\sqrt{1 - \boldsymbol{\delta}_{(-d)\boldsymbol{\theta}_k}^{\top}\boldsymbol{\delta}_{(-d)\boldsymbol{\theta}_k}} + \delta_{d\boldsymbol{\theta}_k}\right)^{\alpha_{\delta_1}-1} \times \\
&\times \left(\sqrt{1 - \boldsymbol{\delta}_{(-d)\boldsymbol{\theta}_k}^{\top}\boldsymbol{\delta}_{(-d)\boldsymbol{\theta}_k}} - \delta_{d\boldsymbol{\theta}_k}\right)^{\alpha_{\delta_2}-1} \mathbb{1}_{\left(\delta_{d\boldsymbol{\theta}_k} \in A_{\delta_{dk}}\right)},
\end{aligned} \qquad (16)
$$

where $(\alpha_{\delta_1}, \alpha_{\delta_2})$ is a set oh hyperparameters and $A_{\delta_{dk}} = (-1, 1)$, for $d = 1, ..., D$ and $k = 1, ..., K$. The prior chosen for the item parameters vector $\boldsymbol{\zeta}_i$ is $\boldsymbol{\zeta}_i = (\boldsymbol{a}_i, b_i)^{\top} \overset{i.i.d.}{\sim} N_D(\boldsymbol{\mu}_{\boldsymbol{\zeta}_i}, \boldsymbol{\Psi}_{\boldsymbol{\zeta}_i})\mathbb{1}_{\boldsymbol{A}_{\boldsymbol{a}_i}}(\boldsymbol{a}_i)\mathbb{1}_{(-\infty,\infty)}(b_i)$, where $\mathbb{1}_{\boldsymbol{A}_{\boldsymbol{a}_i}}(\boldsymbol{a}_i)$ and $\mathbb{1}_{(-\infty,\infty)}(b_i)$ are the indicator functions associated with the item parameters $\boldsymbol{a}_i$ and b_i, respectively, and $\boldsymbol{A}_{\boldsymbol{a}_i}$ is an appropriate set, for example, $\boldsymbol{A}_{\boldsymbol{a}_i} = \mathbb{R}^{+D}$ or $\boldsymbol{A}_{\boldsymbol{a}_i} = \mathbb{R}^{D}$, depending on the situation (see Subsection 2.3), and $c_i \overset{i.i.d.}{\sim}$ beta(κ_1, κ_2), for i=1,..,I. Therefore, from (14) to (15), the joint posterior distribution is given in equation (17).

$$
\begin{aligned}
& p(\boldsymbol{z}_{\ldots}, \boldsymbol{u}_{\ldots}, \boldsymbol{\theta}_{\ldots}, \boldsymbol{\zeta}, \boldsymbol{\eta_\theta} | \boldsymbol{y}_{\ldots}) \\
& \quad \propto \prod_{k=1}^{K} \prod_{j=1}^{n_k} \prod_{i|I_{ijk=1}} \left\{ \exp\left\{ -0.5 \left(z_{ijk} - \boldsymbol{a}_i^{\top} \boldsymbol{\theta}_{.jk} - b_i \right)^2 \right\} \right. \\
& \quad \left. \times\, c_i^{u_{ijk}} (1-c_i)^{1-u_{ijk}} \mathbb{1}_{\left(z_{ijk}, u_{ijk}, y_{ijk} \right)} \right\} \\
& \quad \times \prod_{k=1}^{K} \prod_{j=1}^{n_k} \left\{ \exp\left\{ -0.5 \left(\boldsymbol{\theta}_{.jk} - \boldsymbol{\mu}^*_{\boldsymbol{\theta}_{jk}} \right)^{\top} (\boldsymbol{\Psi}^*_{\boldsymbol{\theta}})^{-1} \left(\boldsymbol{\theta}_{.jk} - \boldsymbol{\mu}^*_{\boldsymbol{\theta}_{jk}} \right) \right\} \mathbb{1}_{I\!R^D}(\boldsymbol{\theta}_{.jk}) \right. \\
& \quad \left. \times \exp\left\{ -\frac{t_{jk}^2}{2} \right\} \mathbb{1}_{(0,\infty)}(t_{jk}) \right\} \\
& \quad \times \prod_{k=1}^{K} \exp\left\{ -0.5 \left(\boldsymbol{\mu}_{\boldsymbol{\theta}_k} - \boldsymbol{\mu}_{\boldsymbol{\mu}} \right)^{\top} \boldsymbol{\Psi}_{\boldsymbol{\mu}}^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{\theta}_k} - \boldsymbol{\mu}_{\boldsymbol{\mu}} \right) \mathbb{1}_{I\!R^D}(\boldsymbol{\mu}_{\boldsymbol{\theta}_k}) \right\} \\
& \quad \times \prod_{k=1}^{K} |\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}|^{-\frac{\kappa+D+1}{2}} exp\left\{ -\frac{1}{2} tr\left(\boldsymbol{\Psi}_{\boldsymbol{\Psi}} \boldsymbol{\Psi}_{\boldsymbol{\theta}_k}^{-1} \right) \right\} \\
& \quad \times \prod_{i=1}^{I} \exp\left\{ -0.5 \left(\boldsymbol{a}_i - \boldsymbol{\mu}_{\boldsymbol{a}} \right)^{\top} \boldsymbol{\Psi}_{\boldsymbol{a}}^{-1} \left(\boldsymbol{a}_i - \boldsymbol{\mu}_{\boldsymbol{a}} \right) \right\} \mathbb{1}_{\boldsymbol{A}_{\boldsymbol{a}_i}}(\boldsymbol{a}_i) \\
& \quad \times \prod_{i=1}^{I} \exp\left\{ -\frac{(b_i - \mu_b)^2}{2\psi_{\boldsymbol{b}}} \right\} \mathbb{1}_{(-\infty,\infty)}(b_i) \\
& \quad \times \prod_{i=1}^{I} c_i^{\kappa_1 - 1} (1-c_i)^{\kappa_2 - 1} \mathbb{1}_{(0,1)}(c_i).
\end{aligned}
\tag{17}
$$

This posterior distribution has an intractable form, and it is not possible to obtain the marginal posterior distributions analytically. Some full conditional distributions, however, are either known, and thus easily sampled from, or can be sampled, using an auxiliary algorithm, such as the Metropolis-Hastings, see Gamerman & Lopes (2006). Here; we use Metropolis-Hastings when the full conditional distribution is unknown. Therefore, our algorithm is a sort of a Metropolis-Hastings within Gibbs Sampling, as in Patz & Junker (1999). The technical details about the full posterior distributions and the MCMC steps can be found in Appendix 8. To develop our algorithm, we need to define a kernel density for the parameters $\boldsymbol{\Psi}_{\boldsymbol{\theta}k}$ and $\boldsymbol{\delta}_{\boldsymbol{\theta}k}$, that is: $q(\boldsymbol{\Psi}^{(t-1)}_{\boldsymbol{\theta}_k}, \boldsymbol{\Psi}_{\boldsymbol{\theta}k}) \sim IW\left(2D; \boldsymbol{\Psi}^{(t-1)}_{\boldsymbol{\theta}k}\right)$, $q(\boldsymbol{\delta}^{(t-1)}_{\boldsymbol{\theta}_k}, \boldsymbol{\delta}_{\boldsymbol{\theta}_k}) \sim U\left(g_1(\boldsymbol{\delta}^{(t-1)}_{\boldsymbol{\theta}k}), g_2(\boldsymbol{\delta}^{(t-1)}_{\boldsymbol{\theta}k})\right)$, where $g_1(\boldsymbol{\delta}^{(t-1)}_{\boldsymbol{\theta}k}) = \max\{-\sqrt{1 - \boldsymbol{\delta}^{\top(t-1)}_{(-d)k}\boldsymbol{\delta}^{(t-1)}_{(-d)k}}, \delta^{(t-1)}_{dk} - 0.01\}$, $g_2(\boldsymbol{\delta}^{(t-1)}_{\boldsymbol{\theta}k}) = \min\{\sqrt{1 - \boldsymbol{\delta}^{\top(t-1)}_{(-d)k}\boldsymbol{\delta}^{(t-1)}_{(-d)k}}, \delta^{(t-1)}_{dk} + 0.01\}$ and $\boldsymbol{\delta}_{(-d)k}$ stands for the vector $\boldsymbol{\delta}_{\boldsymbol{\theta}k}$ after removing the element on position d.

Let (.) denote the set of all necessary parameters. The Metropolis-Hastings within Gibbs sampling algorithm, where GS indicates that the full conditional distribution is known and can be simulated directly and MH indicates that this distribution is not known and it is simulated by using the Metropolis-Hastings algorithm, is defined as follows:

1. Start the algorithm by choosing suitable initial values.
   Repeat steps 2–11:
2. Simulate U_{ijk} from $U_{ijk} \mid (.), i = 1, ..., I_k,\ j = 1, ..., n_k,\ k = 1, ..., K$ (GS).
3. Simulate Z_{ijk} from $Z_{ijk} \mid (.), i = 1, ..., I_k,\ j = 1, ..., n_k,\ k = 1, ..., K$ (GS).
4. Simulate T_{jk} from $T_{jk} \mid (.), j = 1, ..., n_k,\ k = 1, ..., K$(GS).
5. Simulate $\boldsymbol{\theta}_{.jk}$ from $\boldsymbol{\theta}_{.jk} \mid (.), j = 1, ..., n_k,\ k = 1, ..., K$(GS).
6. Simulate $\boldsymbol{\mu}_{\boldsymbol{\theta}_k}$ from $\boldsymbol{\mu}_{\boldsymbol{\theta}_k} \mid (.),\ k = 1, ..., K$ (GS).
7. Simulate $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k}$ from $\boldsymbol{\Psi}_{\boldsymbol{\theta}_k} \mid (.),\ k = 1, ..., K$ (MH).
8. Simulate $\boldsymbol{\delta}_{\boldsymbol{\theta}_k}$ from $\boldsymbol{\delta}_{\boldsymbol{\theta}_k} \mid (.),\ k = 1, ..., K$ (MH).
9. Simulate $(\boldsymbol{a}_i^\top, b_i)^\top$, from $(\boldsymbol{a}_i^\top, b_i)^\top \mid (.)$, i =1,...,I(GS).
10. Use the convergence acceleration algorithm of Gonzalez (2004) to update the values $(\boldsymbol{a}_i^\top, b_i)^\top$, i =1,...,I(GS).
11. Simulate c_i, from $c_i \mid (.)$, i =1,...,I(GS).

We can notice that the MH algorithm is only necessary to simulate from $\boldsymbol{\eta}_{\boldsymbol{\theta}_k}, k = 1, .., K$ (the population parameter). Simulation studies presented (not presented here) induced that the above MCMC algorithm converges with a burn-in of 5,000, a spacement of 50 and a total number of simulations of 55,000. This produces a valid MCMC sample of size 1,000. In addition, in that work, many scenarios of interest, related to number of subjects, test size, number of dimensions of the test, underlying latent trait distribution and number of groups were considered. In all scenarios, all parameters were properly recovered.

3.2. Model fit assessment and model comparison: posterior predictive checking and statistics of model comparison

Besides using model selection criteria for selecting the best model, as in Azevedo et al. (2012), in our case, concerning the test dimensionality, the latent trait distribution and the item response function (probit model), the fit of the general MIRT model can be evaluated using Bayesian posterior predictive tests and/or appropriate plots based on the observed and the replicated data (see Sinharay et al., 2006). The literature about posterior predictive checks for Bayesian item response models shows several diagnostics for evaluating the model fit. A general discussion can be found in, among others, Stern & Sinharay (2005), Sinharay (2006), and Fox (2004, 2005, 2010). Examples where these techniques were successfully applied are Béguin & Glas (2001), Sheng & Wikle (2007), Fragoso & Curi (2013), Azevedo et al. (2012), Santos et al. (2013), Azevedo et al. (2015) and Azevedo et al. (2011).

The usual posterior predictive tests and plots can be generalized to make them applicable for the MIRT model. Each posterior predictive test is based on a discrepancy measure, where this discrepancy measure is defined in such a way that a specific assumption or the general fit of the model can be evaluated. The main idea is to generalize the well-known discrepancy measures to a multidimensional multiple group structure. On the other hand, the plots, in general, display a comparison between the predicted and observed data. These procedures can be done at the population level (general fit), as in Sinharay et al. (2006), per group, as in Santos et al. (2013) and Azevedo et al. (2012), or per item, as in Sinharay (2006) and Azevedo et al. (2012a). An example where the model fit was considered for each one of these three levels can be found in Santos et al. (2013).

In general, let $\boldsymbol{y}_{(.)}^{obs}$ be the matrix of observed responses, and $\boldsymbol{y}_{(.)}^{rep}$ the matrix of replicated responses generated from its posterior predicted distribution, where (.) represents a convenient index, employed whenever it is necessary. The posterior predicted distribution of the response data of group k is represented by $p\left(\boldsymbol{y}_k^{rep} \mid \boldsymbol{y}_k^{obs}\right) = \int p\left(\boldsymbol{y}_k^{rep} \mid \boldsymbol{\vartheta}_k\right) p\left(\boldsymbol{\vartheta}_k \mid \boldsymbol{y}_k^{obs}\right) d\boldsymbol{\vartheta}_k$, where $\boldsymbol{\vartheta}_k$ denotes the set of model parameters related to group k. From $\boldsymbol{y}_k^{(.)}$, $\boldsymbol{y}_{ilk}^{(.)}$ is available, which is the response of the subjects belonging to group k with a score l to item i. One approach commonly employed is to plot an appropriate comparison between the replicated and observed data. For example, the predicted and observed score distributions (at population and/or at group level) and the predicted and observed proportions of correct answer for each item. Another approach is, given a discrepancy measure

$D\left(\boldsymbol{y}_{(.)}, \boldsymbol{\vartheta}_{(.)}\right)$, to use the replicated data to evaluate whether the discrepancy value given the observed data is typical under the model. A p-value can be defined in order to quantify the extremeness of the observed discrepancy value $p_0\left(\boldsymbol{y}_{(.)}^{(obs)}\right) = P\left(D\left(\boldsymbol{y}_{(.)}^{(rep)}, \boldsymbol{\theta}_{(.)}\right) \geq D\left(\boldsymbol{y}_{(.)}^{(obs)}, \boldsymbol{\theta}_{(.)}\right) \mid \boldsymbol{y}_{(.)}^{(obs)}\right)$, where the probability is taken over the joint posterior of $(\boldsymbol{y}_{(.)}^{(rep)}, \boldsymbol{\theta}_{(.)})$. The discrepancy measure can be defined at the population, group or item level. Here, p-values, based on a chi-square distance, and predicted distributions of scores are considered (see Fox, 2004, 2010; Azevedo et al., 2012). The chi-square posterior predictive checking is defined to evaluate the predicted score distribution with the observed score distribution. The discrepancy considered here is a slight modification of that one presented in Sinharay (2006), respectively, at item, group and population level, that are $D_i\left(\mathbf{y}\right) = \sum_l \sum_k \frac{(n_{ilk} - E(N_{ilk}))^2}{E(N_{ilk})}$, $D_k\left(\mathbf{y}\right) = \sum_l \sum_i \frac{(n_{ilk} - E(N_{ilk}))^2}{E(N_{ilk})}$, $D\left(\mathbf{y}\right) = \sum_l \sum_k \sum_i \frac{(n_{ilk} - E(N_{ilk}))^2}{E(N_{ilk})}$, where N_{ilk} is the number of subjects with a score l at group k that answer correctly the item i, and $E(.)$ stands for the respective expectation (which is calculated using the posterior predictive distribution). The posterior predictive checking is evaluated using MCMC output. Naturally, when an item is not presented to a given group and/or a given subject, the related quantities are skipped in the calculations above through the matrix $\mathbf{I}$, as defined in page 16.

The predicted score distribution is easily calculated using the MCMC output. In each iteration, a sample of the score distribution is obtained. This is accomplished by generating response data from the sampled parameters according to the model. Subsequently, the number of subjects can be calculated for each possible score at each group. For each possible score, the median and 95% equi-tailed credibility intervals are calculated to evaluate the score distribution. Concerning the proportion of correct response for each item, a similar procedure is used and then, the number of subjects that correctly answer the item, at each possible score, is then calculated.

Another based predictive checking technique is proposed here to determine the test dimensionality. It is known that the matrix of the tetrachoric correlation can be helpful to determine the test dimensionality, see Reckase (2009). The idea is to compare the eigenvalues associated to the matrix of the tetrachoric correlation of the replicated data with those obtained with the observed data. For example, if we have three competing MIRT models, that is, a uni-, two- and three-dimensional models, the model that produces predicted eigenvalues more similar to those obtained with the observed data, should be the most appropriate model. In other words, the test dimensionality should correspond to the model with the predicted eigenvalues closest to the observed eigenvalues. The underlying idea is similar to those presented above, that is, in each MCMC iteration we calculated the eigenvalues associated with the matrix of the tetrachoric correlation of the replicated data. Then, we have the posterior distribution of the eigenvalues, which can be used to calculate the posterior mean or median and the respective HPD intervals.

3.3. Statistics of model comparison

In this work, we considered the model comparison statistics presented in Spiegelhalter et al. (2002) and successfully used by Santos et al. (2013), Azevedo et al. (2011) and Bazan et al. (2006), which are based on the concept of Bayesian deviance, see Dempster et al. (1977). The Deviance, let us say $D(\boldsymbol{\vartheta})$, is given by $D(\boldsymbol{\vartheta}) = -2\ln\left[L(\boldsymbol{\theta}_{...},\boldsymbol{\zeta})p(\boldsymbol{\theta}|\boldsymbol{\eta}_{\boldsymbol{\theta}})\right]$, where $L(\boldsymbol{\theta}_{...},\boldsymbol{\zeta}) = \prod_{k=1}^{K}\prod_{j=1}^{n_k}\prod_{i|I_{ijk}=1} P_{ijk}^{y_{ijk}}(1-P_{ijk})^{1-y_{ijk}}$, is the original likelihood, $p(\boldsymbol{\theta}|\boldsymbol{\eta}_{\boldsymbol{\theta}}) = \prod_{k=1}^{K}\prod_{j=1}^{n_k} p(\boldsymbol{\theta}_{.jk}|\boldsymbol{\eta}_{\boldsymbol{\theta}_k})$ and $p(\boldsymbol{\theta}_{.jk}|\boldsymbol{\eta}_{\boldsymbol{\theta}_k})$ is as in (5). Also, let $\boldsymbol{\vartheta}^{(m)}(m = 1,...,M)$ the m-th value of the valid simulated MCMC sample.

The ρ_D statistic (also named effective sample size) is given by $\rho_D = \overline{D(\boldsymbol{\vartheta})} - D\left(\overline{\boldsymbol{\vartheta}}\right)$, where $\overline{\boldsymbol{\vartheta}}$ is the vector with the posterior expectation of each parameter, based on the valid MCMC sample and $\overline{D(\boldsymbol{\vartheta})} = \frac{1}{M}\sum_{m=1}^{M} D\left(\boldsymbol{\vartheta}^{(m)}\right)$. In addition, the deviance information criterion (DIC) is given by $DIC = D\left(\overline{\boldsymbol{\vartheta}}\right) + 2\rho_D$. The posterior expectation of the Akaike information (EAIC) criterion and of the Bayesian information criterion are, respectively, defined as, $EAIC = \overline{D(\boldsymbol{\vartheta})} + 2p$, $EBIC = \overline{D(\boldsymbol{\vartheta})} + p\ln(N)$, where p is the number of parameters and N is the number of observations, that is, $N = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\sum_{i=1}^{I} I_{ijk}$. Here we follow the suggestions of Spiegelhalter et al. (2002) by taking $p = \rho_D$.

4. Real data analysis

We analyzed a part of the 2013 first stage of the admission exam of the University of Campinas (see *https : //www.comvest.unicamp.br/*, in Portuguese). The test is composed of 48 multiple-choice items. Item 43 was excluded, since it presented a negative biserial correlation, but the original numeration was preserved. We selected a sample of 3,000 candidates (among those that answered all items on the test) spread uniformly over six

Table 1: Comvest dataset. Questions skills of the exam.

Item	Skill	a_{i1}	a_{i2}	Item	Skill	a_{i1}	a_{i2}
Q1	Philosophy	.512	.138	Q25	Chemistry	.877	.941
Q2	History	.554	.226	Q26	Chemistry	.728	.605
Q3	History	.772	.258	Q27	Chemistry	1.032	1.056
Q4	History	.996	.407	Q28	Chemistry	1.019	1.011
Q5	History	.797	.332	Q29	Chemistry	.986	1.240
Q6	History	1.258	.546	Q30	Chemistry	1.684	1.452
Q7	History	.956	.519	Q31	Physics	.715	.939
Q8	History	1.217	.491	Q32	Physics	.690	.914
Q9	History	.910	.463	Q33	Physics	.937	1.159
Q10	Geography	1.205	1.198	Q34	Physics	.716	1.058
Q11	Geography	.159	.104	Q35	Physics	.918	.772
Q12	Geography	.590	.274	Q36	Physics	1.132	1.633
Q13	Geography	.479	.301	Q37	Mathematics	.409	.586
Q14	Geography	.608	.430	Q38	Mathematics	.972	1.398
Q15	Geography	.513	.433	Q39	Mathematics	.484	.815
Q16	Geography	.959	.620	Q40	Mathematics	.861	1.164
Q17	Sociology	.707	.201	Q41	Mathematics	1.468	1.901
Q18	Geography	.926	1.086	Q42	Mathematics	1.312	1.783
Q19	Biology	1.476	1.061	Q44	Mathematics	1.040	1.623
Q20	Biology	.604	.352	Q45	Mathematics	.643	1.374
Q21	Biology	.812	.438	Q46	Mathematics	.936	1.252
Q22	Biology	1.463	1.296	Q47	Mathematics	.827	1.013
Q23	Biology	.544	.382	Q48	Mathematics	.968	1.541
Q24	Biology	.162	.038				

areas (Arts, Biological and Health Sciences, Exact and Earth Sciences, Humanities, Medicine and Technological), respectively groups 1, 2, 3, 4, 5 and 6. Therefore, we have six groups with $n_k = 500, k = 1, ..., 6$, submitted to the same 47 item-test in a complete design (every subject answered all items of the test). Each item on the test is supposed to be related, by design, to, at least, one of the following fields: Biology, Chemistry, Geography, Geology, History, Mathematics, Physics and Sociology. In Table 1, we present a classification of each item concerning these fields. Table 1 also presents the posterior estimates for the parameters $\boldsymbol{a}_i$ for each item. Further details regarding the estimates of parameter vector $\boldsymbol{\zeta}_i$ are presented in Figure 6.

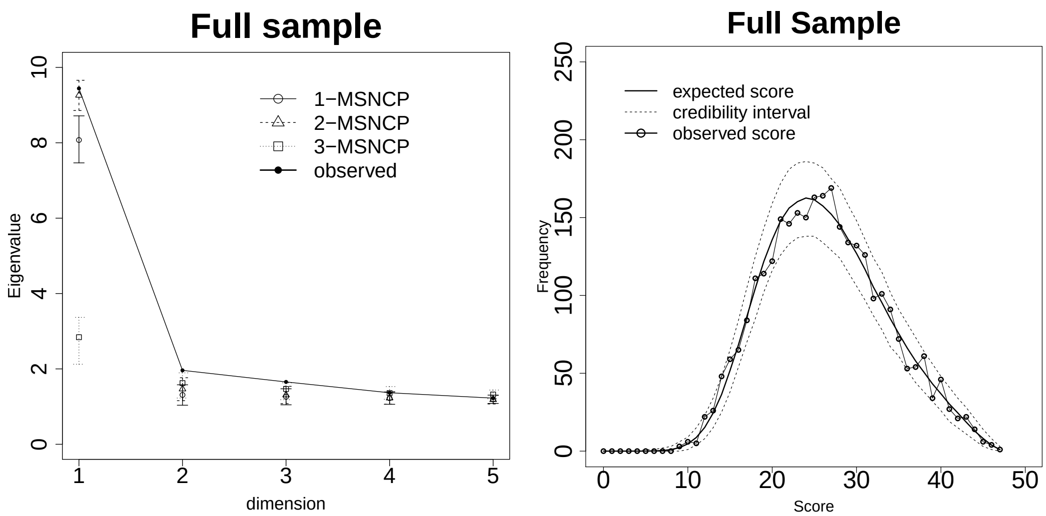

Figure 2: Comvest dataset. Full sample tetrachoric eigenvalues and predictive scores.

The goal is to analyze the items and the subjects, identifying the test dimensionality and the differences among the groups, and select the most suitable group-specific multivariate latent traits distribution (either multivariate symmetric normal or MSNCP). The reference group here considered is the second one.

We fitted a total of six models, varying according the latent trait distribution (normal or skew normal under the centered parameterization) and the number of dimensions (1, 2 or 3) of the test. The values of the hyperparameters, when pertinent, were the same for all models as well as the values of the parameters related to the kernel densities. We considered $\boldsymbol{\mu}_{\boldsymbol{\zeta i}} = (0.5, \ldots, 0.5, 0)^\top$ and $\boldsymbol{\Psi}_{\boldsymbol{\zeta i}} = diag(0.5, \ldots, 0.5, 2)$; $\kappa_1 = 100$ and $\kappa_2 = 300$; $\alpha_{\delta_1} = 0.5$ and $\alpha_{\delta_2} = 0.5$; $\tau \to 0$ and $|\boldsymbol{\Psi}_{\boldsymbol{\Psi}}| \to 0$; $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Psi}_{\boldsymbol{\mu}} = (5, \ldots, 5)^\top$. We observed that the two dimensional model with multivariate centered skew normal latent traits (2-MSNCP) model is selected by EDIC and EAIC whereas the statistic EBIC selected the one dimensional model with univariate centered skew normal latent traits (1-MSNCP). While the asymmetry of at least one of the latent trait distribution is detected, the test dimensionality varies between one and two. Figures 2 and 3 display the eigenvalues of the tetrachoric correlation matrices for the observed response

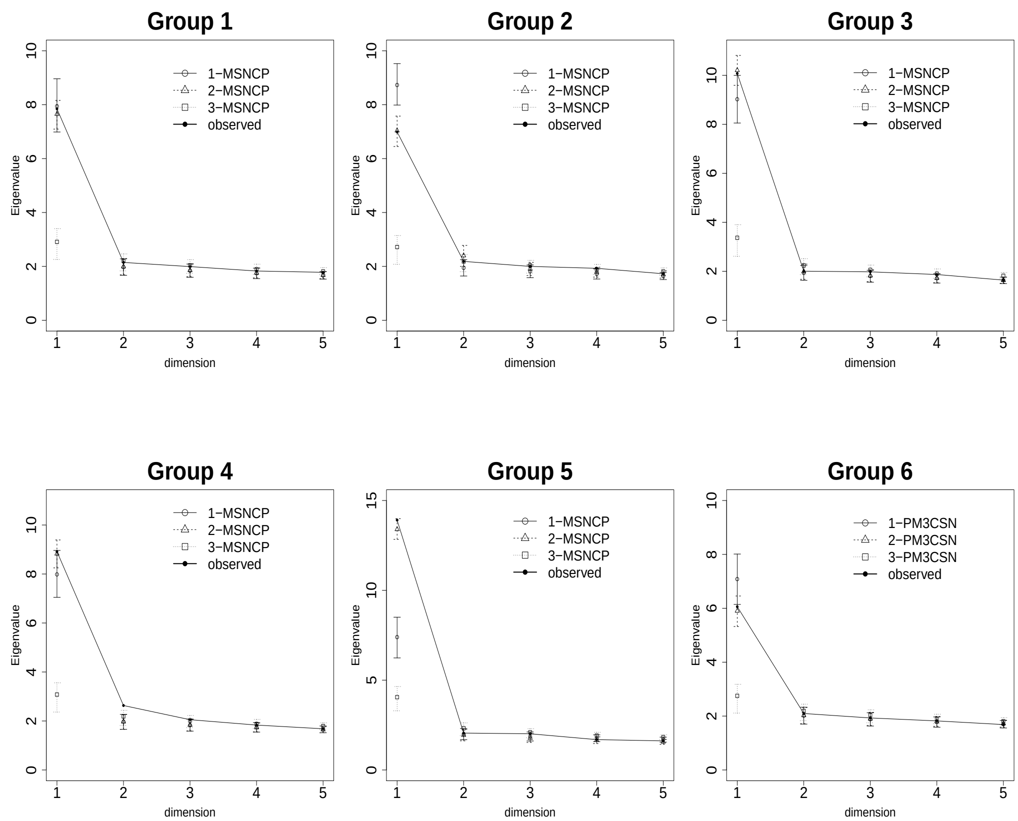

Figure 3: Comvest dataset. Eigenvalues tetrachoric correlation matrix.

matrices associated to the whole sample and to each group, respectively. These figures also display the eigenvalues of the tetrachoric correlation matrices for response matrices simulated according to the fitted asymmetric models of dimensions 1, 2 and 3.

From Figures 2 and 3, we can see that the eigenvalues of the tetrachoric correlation matrix, predicted from model 2-MSNCP, are the closest obtained from the observed (tetrachoric) correlation matrix, compared with the other models. Therefore we choose the model 2-MSNCP. Figures 2 and 4 represent the observed, predicted and the 95% credibility intervals, related to the score distributions at population level and group level, respectively. We can see that the model, in both cases, is well fitted to the data. The results of the

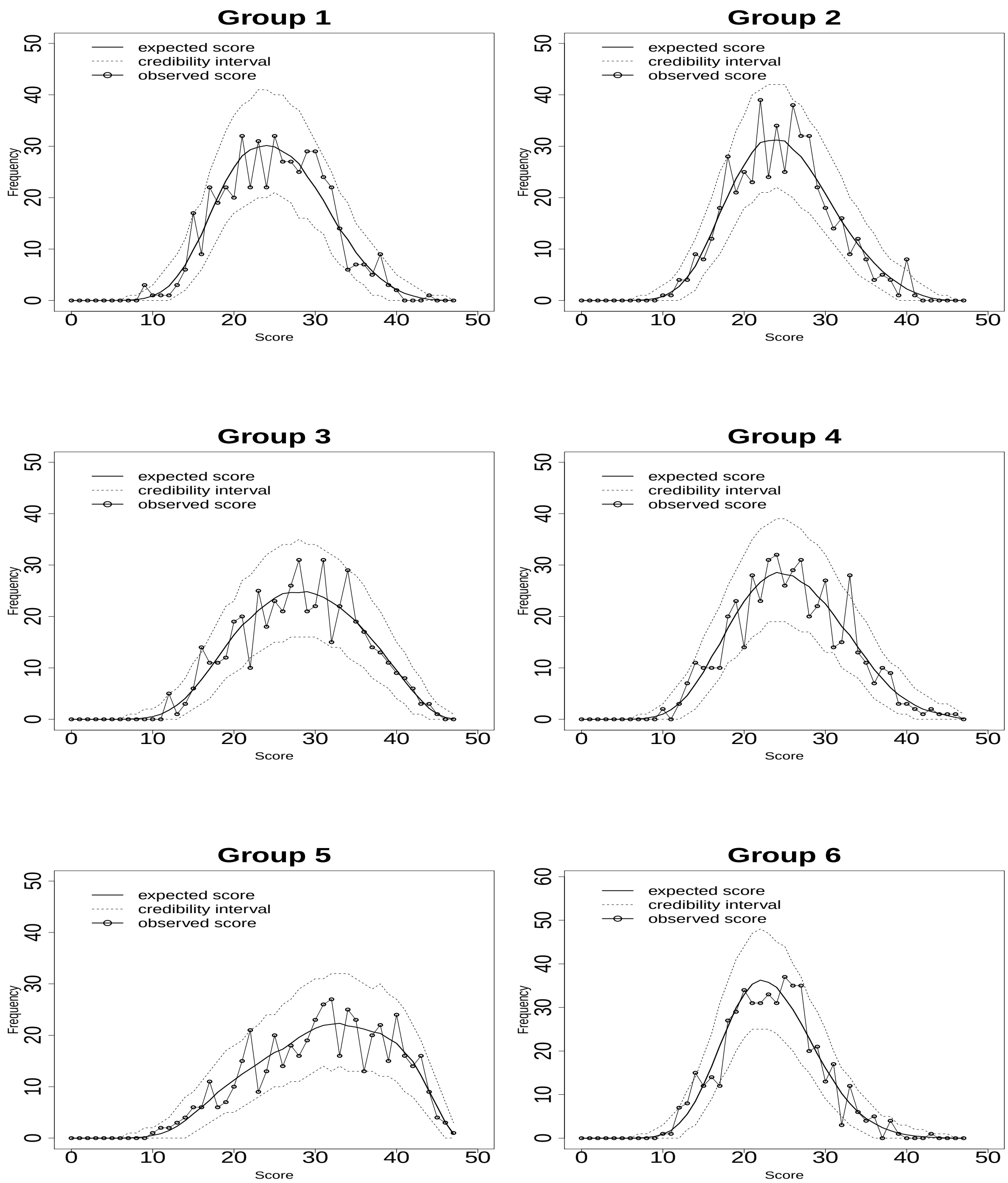

Figure 4: Comvest dataset. Scores distribution per group.

item-fit analysis, based on the p-value for the chi-square distance, are shown in Figure 10. We also compute the p-values at group level, which ranged from 0.12 and 0.95 indicating that the model fitted well all almost of the items. Figure 5 presents the item fit plots for some selected item according they were or not well fitted by the model. In conclusion, we can say that the model is properly fitted to almost all of the items.

In Tables 2 and 3, we present the estimatives for the population parameters, there we can notice that group 5 presents the largest population means (for the two dimensions), that the variances (for the two dimensions) are similar among the groups and that for groups 1, 2 and 5 the latent traits distributions, for one dimension at least, are skewed. In Figures 7 and 8, we compare the empirical densities of the estimated $\boldsymbol{\theta}_{j.}$ (for both the symmetric and skew model two-dimensional models) with their respective theoretical densities, that is, with either the multivariate normal density or the MSNCP density with parameters equal to those presented in Tables 2 and 3. From Figures 7 and 8, we can see how the empirical densities are very similar to the theoretical ones. In Figure 9, we present the Box-plots of the posterior distribution for the asymmetry parameters for each group. There we can notice how the asymmetry for some groups can be thought to be different from 0.

From Figure 6, we can conclude that most of the items present a reasonable discrimination power and that they are difficult for most of the groups (see the estimates of the population means from Table 2). Also, except for Item 22, all items present guessing estimates compatible with a pattern of random choice (for subjects with low latent trait level) for an item with four or five alternatives. Also, from Table 1, we can notice that the factor loadings have no pattern concerning the fields, since the magnitude of them are, in general, very similar within each field. Even after rotating them, no pattern was observed. Therefore, it is not easy to provide interpretations for the two retained factors. Figure 6 presents the posterior expectations and the equi-tailed 95% credibility intervals for the multidimensional discrimination (MULDISC $= \sqrt{\sum_{d=1}^{D} a_{id}^2}$) and the multidimensional difficulty (MULTDIF $= \frac{b_i}{\sum_{d=1}^{D} a_{id}^2}$), which have similar interpretations to those presented by their unidimensional versions and in Figure 10 we plot these estimates. See Reckase (2009) for further details.

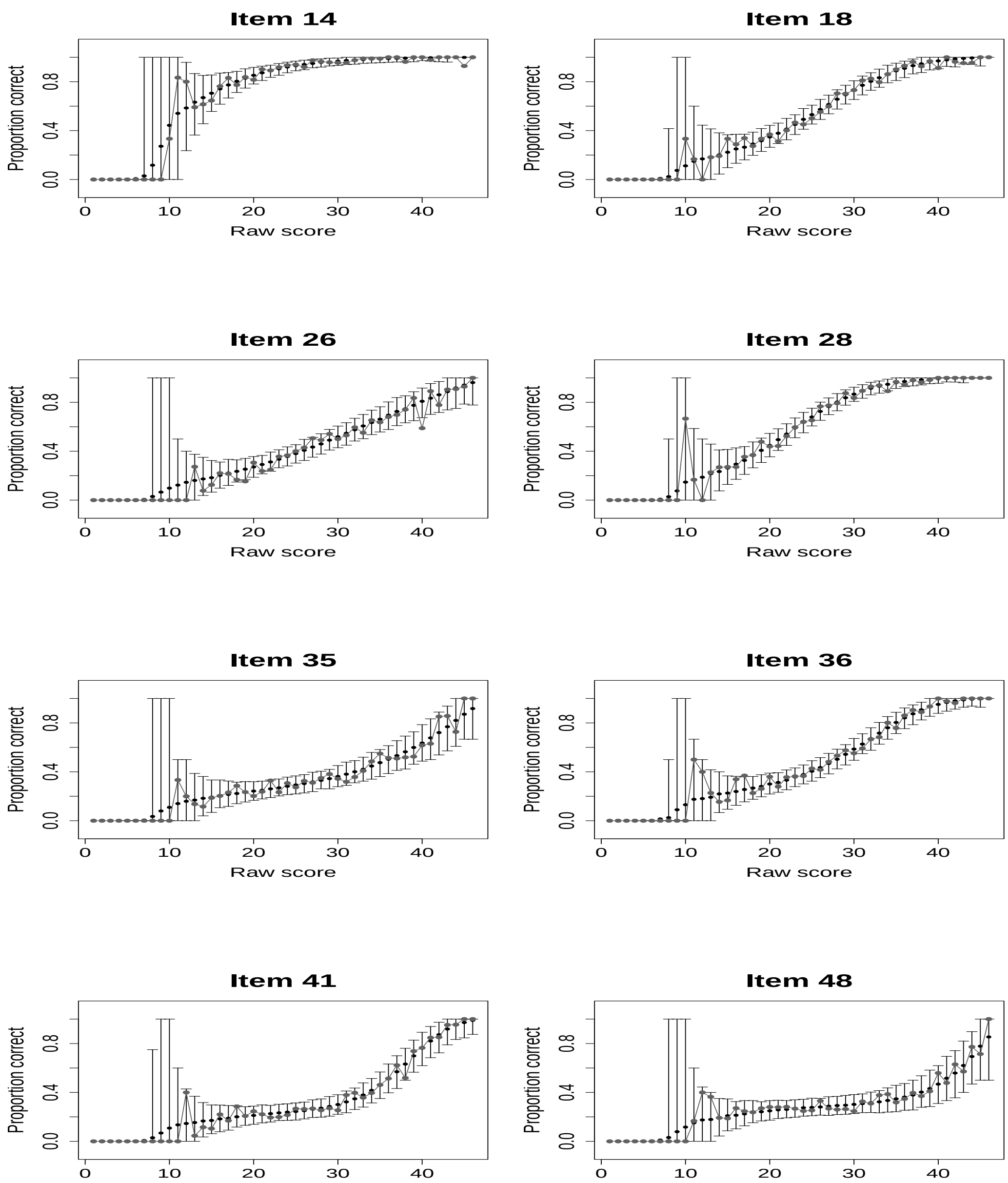

Figure 5: Comvest dataset. Item fit plots.

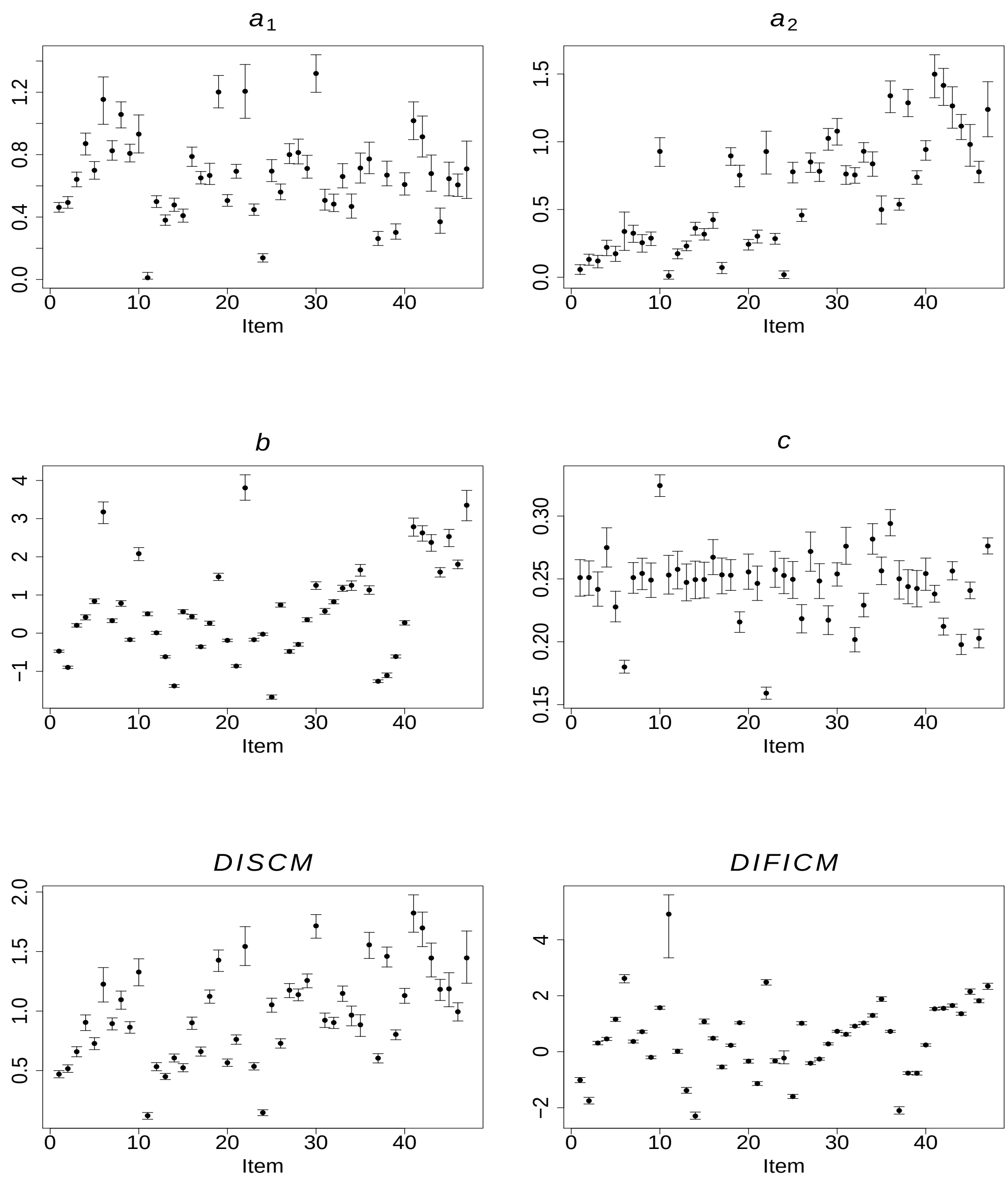

Figure 6: Comvest dataset. Posterior expectations and equi-tailed 95% credibility intervals of the item parameters.

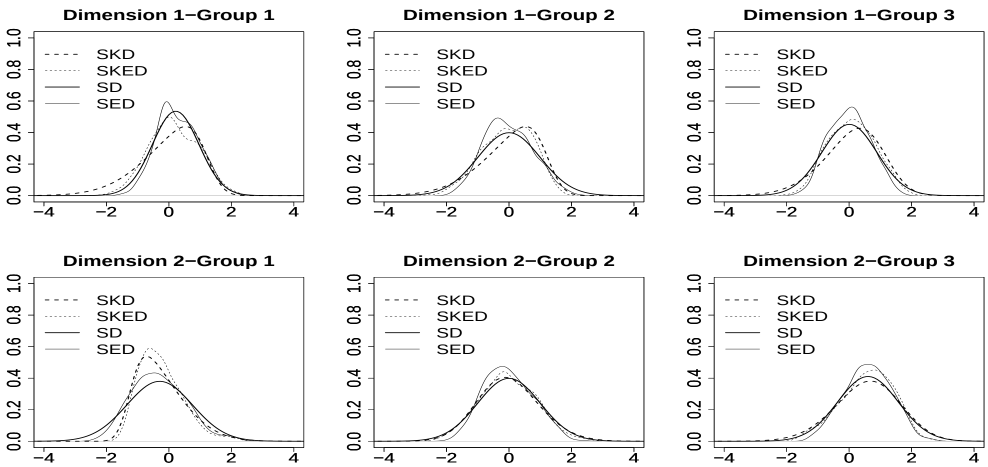

Figure 7: Comvest dataset. Empirical and theoretical posterior densities: groups 1, 2 and 3. Where SKD, SKED, SD and SED means skewed density, skewed empirical density, symmetrical density and symmetric empirical density, respectively.

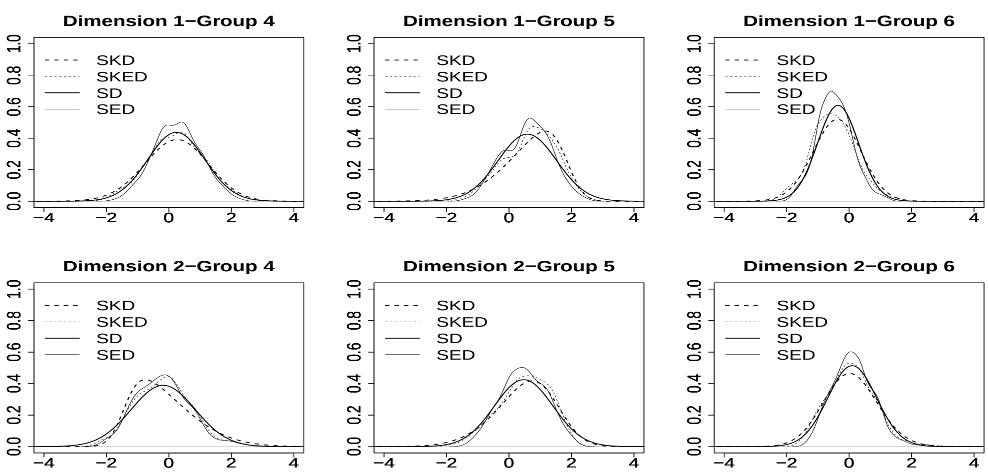

Figure 8: Comvest dataset. Empirical and theoretical posterior densities: groups 4, 5 and 6. Where SKD, SKED, SD and SED means skewed density, skewed empirical density, symmetrical density and symmetric empirical density, respectively.

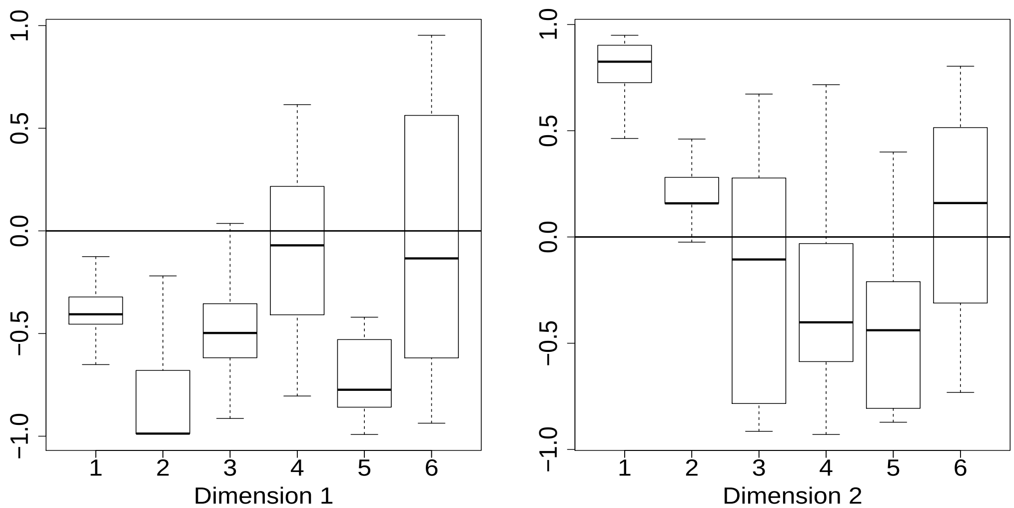

Figure 9: Comvest dataset. Box-plots of the posterior distributions of $\boldsymbol{\delta}_{\theta k}$.

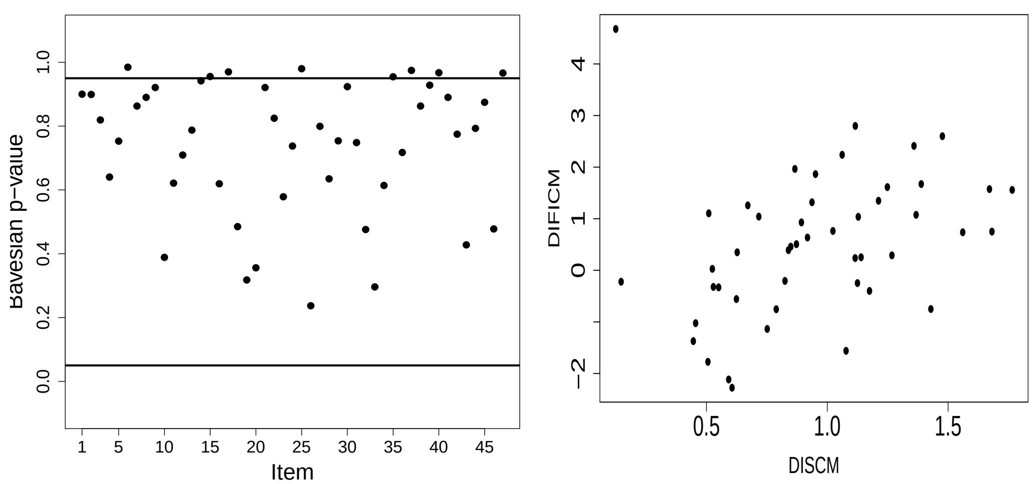

Figure 10: Comvest dataset. Bayesian p-value per item; multidimensional difficulty vs multidimensional discrimination

Table 2: Comvest dataset. Posterior expectation and equi-tailed 95% credibility intervals.

		Model					
		Skew			Symmetric		
Group	Parameter	Mean	SD	CI(95%)	Mean	SD	CI(95%)
1	μ_{θ_1}	.15	.08	$[.00, .33]$	.22	.07	$[.09, .35]$
	μ_{θ_2}	-0.25	.12	$[-0.53, -0.06]$	-0.29	.09	$[-0.49, -0.13]$
	$\sigma^2_{\theta_1}$	.87	.00	$[.87, .87]$	.56	.03	$[.53, .60]$
	$\sigma^2_{\theta_2}$	.67	.01	$[.67, .67]$	1.10	.15	$[.88, 1.27]$
	ρ_θ	-0.03	.00	$[-0.03, -0.03]$	-0.22	.09	$[-0.37, -0.05]$
	δ_{θ_1}	-0.37	.19	$[-0.72, .18]$	-	-	-
	δ_{θ_2}	.79	.12	$[.49, .94]$	-	-	-
2	ρ_θ	-0.20	.05	$[-0.31, -0.07]$	-0.42	.10	$[-0.59, -0.23]$
	δ_{θ_1}	-0.76	.37	$[-0.99, -0.17]$	-	-	-
	δ_{θ_2}	.24	.20	$[-0.13, .70]$	-	-	-
3	μ_{θ_1}	.08	.09	$[-0.11, .26]$	.01	.10	$[-0.20, .20]$
	μ_{θ_2}	.57	.09	$[.39, .73]$	.60	.09	$[.43, .76]$
	$\sigma^2_{\theta_1}$	.95	.05	$[.87, .99]$	.78	.05	$[.69, .81]$
	$\sigma^2_{\theta_2}$	1.12	.15	$[.89, 1.27]$	.94	.11	$[.78, 1.17]$
	ρ_θ	.02	.06	$[-0.05, .11]$	-0.13	.12	$[-0.39, .00]$
	δ_{θ_1}	-0.47	.21	$[-0.80, .00]$	-	-	-
	δ_{θ_2}	-0.20	.53	$[-0.89, .59]$	-	-	-
4	μ_{θ_1}	.22	.08	$[.06, .41]$	.24	.08	$[.09, .39]$
	μ_{θ_2}	-0.16	.09	$[-0.35, .02]$	-0.18	.10	$[-0.40, .02]$
	$\sigma^2_{\theta_1}$	1.05	.14	$[.78, 1.45]$	.83	.08	$[.73, .95]$
	$\sigma^2_{\theta_2}$	1.08	.13	$[.89, 1.24]$	1.05	.15	$[.86, 1.34]$
	ρ_θ	-0.08	.07	$[-0.18, .01]$	-0.17	.06	$[-0.23, -0.06]$
	δ_{θ_1}	-0.08	.36	$[-0.68, .51]$	-	-	-
	δ_{θ_2}	-0.30	.39	$[-0.86, .57]$	-	-	-
5	μ_{θ_1}	.67	.09	$[.50, .85]$	.59	.10	$[.37, .77]$
	μ_{θ_2}	.49	.12	$[.23, .71]$	.47	.14	$[.21, .73]$
	$\sigma^2_{\theta_1}$	.96	.11	$[.83, 1.08]$	.88	.06	$[.79, .95]$
	$\sigma^2_{\theta_2}$	.99	.09	$[.81, 1.05]$	.88	.20	$[.61, 1.05]$
	ρ_θ	.38	.06	$[.32, .47]$	.14	.06	$[.10, .26]$
	δ_{θ_1}	-0.73	.16	$[-0.98, -0.49]$	-	-	-
	δ_{θ_2}	-0.44	.32	$[-0.86, .09]$	-	-	-

Table 3: Cont. of Table 2.

		Model					
		Skew			Symmetric		
Group	Parameter	Mean	SD	CI(95%)	Mean	SD	CI(95%)
6	μ_{θ_1}	-0.39	.08	$[-0.53, -0.24]$	-0.36	.07	$[-0.50, -0.21]$
	μ_{θ_2}	.06	.09	$[-0.10, .22]$	.10	.08	$[-0.06, .26]$
	$\sigma^2_{\theta_1}$	.59	.03	$[.56, .62]$	.43	.03	$[.41, .45]$
	$\sigma^2_{\theta_2}$	.74	.03	$[.67, .77]$	.60	.06	$[.52, .70]$
	ρ_θ	-0.06	.01	$[-0.07, -0.05]$	-0.11	.04	$[-0.18, -0.07]$
	δ_{θ_1}	-0.05	.61	$[-0.88, .86]$	-	-	-
	δ_{θ_2}	.10	.45	$[-0.64, .76]$	-	-	-

5. Conclusions and Comments

In this work, a Multidimensional Multiple Group IRT model with a multivariate skew normal latent trait distribution under the centered parameterization was presented. Bayesian inference for parameter estimation, model comparison and model fit assessment were developed through MCMC algorithms. Simulation studies indicate that the proposed model presented more accurate results compared to the symmetric one, when the underlying latent trait distributions corresponds to the MSNCP. Moreover, the model/MCMC algorithm recovered all parameters properly (simulation study not shown here). The developed tools were illustrated through an analysis of a real data set related to the first stage of the University of Campinas Admission Exam. In this example, a two dimensional MIRT model and skew normal distribution for the latent traits was selected, which indicated that three, of the six groups, presented asymmetric behavior. Also, this model fitted to the data quite properly.

As future research we suggest to consider other distributions for the latent traits and/or other link functions (item response functions) as the skew-t uni and multivariate distributions. Also, other estimation methods such as CADEM, see Azevedo et al. (2012b) and Metropolis-Hastings Robbins-Monro algorithm see Cai (2010), can be considered. Other techniques for determination of the test dimensionality, as RJMCMC algorithms could be explored. In addition, other tools for model fit assessment, as residual analysis, can be developed.

6. Acknowledgments

The authors are thankful to the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) for the financial support.

7. Bibliography

Andrade, D.F. and Tavares, H.R. Item response theory for longitudinal data: population parameter estimation. *Journal of Multivariate Analysis*, 95(1):1–22, 2005.

Arellano-Valle, R.B. and Azzalini, A. The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99(7):1362–1382, 2008.

Azevedo, C.L.N., Andrade, D.F. and Fox, J.-P. A Bayesian generalized multiple group IRT model with model-fit assessment tools. *Computational statistics & Data Analysis*, 56(12):4399–4412, 2012.

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

Azevedo, C.L.N., Bolfarine, H. and Andrade D.F. Parameter recovery for a skew-normal IRT model under a bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size. *Journal of Statistical Computation and Simulation*, 82(11):1679–1699, 2012.

Azevedo, C. L. N., Andrade, D. F. and Fox,J.-P. CADEM: A conditional augmented data EM algorithm for fitting one parameter probit models. *Brazilian Journal of Probability and Statistics*, 27, 245–262, 2012.

Azevedo, C.L.N., Fox, J.-P. and Andrade, D.F. Longitudinal multiple-group irt modelling: covariance pattern selection using MCMC and RJMCMC. *International Journal of Quantitative Research in Education*, 2(3-4):213–243, 2015.

Azevedo, C.L.N, Fox, J.-P. and Andrade, D.F. Bayesian longitudinal item response modeling with restricted covariance pattern structures. *Statistics and Computing*, 26(1-2):443–460, 2016.

Azzalini, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985.

Azzalini, A. and Capitanio, A. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):579–602, 1999.

Bartolucci, F. A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika*, 72(2):141–157, 2007.

Bazán, J.L., Branco, M. and Bolfarine. A skew item response model. *Bayesian Analysis*, 1(4):861–892, 2006.

Béguin, A. and Glas, C.A.W. Mcmc estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66(4):541–561, 2001.

Bock, R.D. and Zimowski, M.F. Multiple group IRT. In *Handbook of modern item response theory*, pages 433–448. Springer, 1997.

Bolt, D.M. and Lall, V.F. Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, 27(6):395–414, 2003.

Cai, L. Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35(3):307–335, 2010.

De Jong, M.G. and Steenkamp, J.-B. Finite mixture multilevel multidimensional ordinal IRT models for large scale cross-cultural research. *Psychometrika*, 75(1):3–32, 2010.

Torre, J. and Patz, R.J. Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, 30(3):295–311, 2005.

Dempster, A.P, Laird, N.M. and Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B, 1–38, 1977.

Fox, J.-P. Multilevel IRT model assessment. *New developments in categorical data analysis for the social and behavioral sciences*, pages 227–252, 2005.

Fox, J.-P. *Bayesian item response modeling: Theory and applications*. Springer Science & Business Media, 2010.

Fox, J.-P. and Glas, C.A.W. Bayesian modification indices for IRT models. *Statistica Neerlandica*, 59(1):95–106, 2005.

Fragoso, T.M. and Cúri, M. Improving psychometric assessment of the beck depression inventory using multidimensional item response theory. *Biometrical Journal*, 55(4):527–540, 2013.

Fu, Z.H, Tao, J. and Shi, N.-Z. Bayesian estimation in the multidimensional three-parameter logistic model. *Journal of Statistical Computation and Simulation*, 79(6):819–835, 2009.

Gamerman, D. and Lopes, H.F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

Genton, M.G. *Skew-elliptical distributions and their applications: a journey beyond normality*. CRC Press, 2004.

Horn, R.A. and Johnson, C.R. *Matrix analysis*. Cambridge university press, 2012.

Lachos, V.H. *Skew linear mixed models*. PhD thesis, 2004.

Leon-Gonzalez, R. Data augmentation in the Bayesian multivariate probit model. 2004. `http://eprints.whiterose.ac.uk/9887/1/SERP2004001.pdf`.

Levy, R. and Mislevy, R.J. and Sinharay, S. Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement*, 33(7):519–537, 2009.

Matos, G.S. *Multidimensional IRT models with skew distributions for the latent traits*. PhD thesis.

Montgomery, D.C. *Design and analysis of experiments*. John Wiley & Sons, 2008.

Padilla, J.L. Multidimensional multiple group skew item response theory models for dichotomous responses under a Bayesian approach. Master's thesis (In Portuguese), 2014.

Padilla, J.L., Azevedo, C.L.N. and Lachos, V.H. Parameter recovery for a skew multidimensional item response model: a comparison of MCMC algorithms and measurement of some effects of interest, *manuscript under preparation.*

Patz, R.J. and Junker, B.W. Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of educational and behavioral statistics*, 24(4):342–366, 1999.

Pewsey, A. Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27(7):859–870, 2000.

Reckase, M. *Multidimensional item response theory*, volume 150. Springer, 2009.

Rivers, D. Identification of multidimensional spatial voting models. *Typescript. Stanford University*, 2003.

Sahu, S.K. Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, 72(3):217–232, 2002.

Santos, J.R., Azevedo, C.L.N. and Bolfarine, H. A multiple group item response theory model with centered skew-normal latent trait distributions under Bayesian framework. *Journal of Applied Statistics*, 40(10):2129–2149, 2013.

Sheng, Y. and Wikle, C.K. Bayesian multidimensional IRT models with a hierarchical structure. *Educational and Psychological Measurement*, 68(3):413–430, 2008.

Sheng, Y. and Wikle, C.K. Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 67(6):899–919, 2007.

Sinharay, S. Bayesian item fit analysis for unidimensional item response theory models. *British Journal of Mathematical and Statistical Psychology*, 59(2):429–449, 2006.

Sinharay, S., Johnson, M.S. and Stern, H.S. Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, 30(4):298–321, 2006.

Spiegelhalter, D.J., Best, N.G. and Carlin, B.P. and Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

Stern, H.S. and Sinharay, S. Bayesian model checking and model diagnostics. *Handbook of Statistics*, 25:171–192, 2005.

8. Appendix

The algorithm is described in the following:

- *Step 1:* Simulate the augmented variables $U_{ijk}^{(t)}$ from

$$\begin{aligned} U_{ijk}|(.) &= 0 \times \mathbb{1}_{\{0\}}(y_{ijk}) + 1 \times \mathbb{1}_{\{1\}}(y_{ijk})\mathbb{1}_{(-\infty,0)}(z_{ijk}^{(t-1)}) \\ &+ \text{Bernoulli}(c_i^{(t-1)})\mathbb{1}_{\{1\}}(y_{ijk})\mathbb{1}_{(0,\infty)}(z_{ijk}^{(t-1)}), \end{aligned} \tag{18}$$

  for $i = 1, ..., I$, $j = 1, ..., N_k$ and $k = 1, ..., K$.

- *Step 2:* Simulate the augmented variables $Z_{ijk}^{(t)}$ from

$$\begin{aligned} Z_{ijk}|(.) &\sim N_{(-\infty,0)}(\boldsymbol{a}_i^{\top(t-1)}\boldsymbol{\theta}_{.jk}^{(t-1)} - b_i^{(t-1)}, 1)\mathbb{1}_{\{0\}}(y_{ijk}) \\ &+ N_{(0,\infty)}(\boldsymbol{a}_i^{\top(t-1)}\boldsymbol{\theta}_{.jk}^{(t-1)} - b_i^{(t-1)}, 1)\mathbb{1}_{\{1\}}(y_{ijk})\mathbb{1}_{\{0\}}(u_{ijk}^{(t)}) \\ &+ N(\boldsymbol{a}_i^{\top(t-1)}\boldsymbol{\theta}_{.jk}^{(t)} - b_i^{(t-1)}, 1)\mathbb{1}_{\{1\}}(y_{ijk})\mathbb{1}_{\{1\}}(u_{ijk}^{(t)}), \end{aligned} \tag{19}$$

  where $N_{[a,b]}(\mu, \psi)$ stands for a truncated normal distribution at the interval $[a, b]$ with nontruncated mean μ and non truncated variance ψ. This for $i = 1, ..., I$, $j = 1, ..., n_k$ and $k = 1, ..., K$.

- *Step 3:* Simulate the latent variables $T_{jk}^{(t)}$ from $T_{jk}|(.) \sim HN\left(\widehat{\psi}_{Tjk}^{(t-1)}\widehat{t}_{jk}^{(t-1)}, \widehat{\psi}_{Tjk}^{(t-1)}\right)$ for $j = 1, ..., n_k$, $k = 1, ..., K$, mutually independently, where

$$
\begin{aligned}
\widehat{\psi}_{Tjk}^{(t-1)} &= \left(1+\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)}(\boldsymbol{I}_D-\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})^{-1}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\right)^{-1}, \\
\widehat{t}_{jk}^{(t-1)} &= \boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)}(\boldsymbol{I}_D-\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})^{-1}\left(\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{1/2\top(t-1)}\right)^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\right)^{-1}\left(\boldsymbol{\theta}_{jk}^{(t-1)}-\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t-1)}\right) \\
&+ \boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)}(\boldsymbol{I}_D-\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})^{-1}\left(\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{1/2\top(t-1)}\right)^{-1}\boldsymbol{\varepsilon}_{\boldsymbol{\theta}k}^{(t-1)},
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\varepsilon}_{\boldsymbol{\theta}k}^{(t-1)} &= -\left(\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{1/2\top(t-1)}\right)^{-1}\boldsymbol{\mu}_{\boldsymbol{Z}k}, \\
\boldsymbol{\mu}_{\boldsymbol{Z}k} &= \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}, \\
\boldsymbol{\Sigma}_{\boldsymbol{Z}k} &= \boldsymbol{I}_D-\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}.
\end{aligned} \tag{20}
$$

- *Step 4:* Simulate $\boldsymbol{\theta}_{jk}^{(t)}|(.)$ from $N_D\left(\widehat{\boldsymbol{\Xi}}_{\boldsymbol{\theta}k}^{-1(t-1)}\widehat{\boldsymbol{\varrho}}_{\boldsymbol{\theta}k}^{(t-1)},\widehat{\boldsymbol{\Xi}}_{\boldsymbol{\theta}k}^{-1(t-1)}\right)$, for $j = 1,...,n_k$, $k=1,...,K$, mutually independently, where

$$
\begin{aligned}
\widehat{\boldsymbol{\Xi}}_{\boldsymbol{\theta}k}^{(t-1)} &= \boldsymbol{a}_{jk}^{\top(t-1)}\boldsymbol{a}_{jk}^{(t-1)} \\
&+ \left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}(\boldsymbol{I}_D-\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2(t-1)}\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2(t-1)}\right)^{-1}, \\
\widehat{\boldsymbol{\varrho}}_{\boldsymbol{\theta}k} &= \boldsymbol{a}_{jk}^{\top(t-1)}\boldsymbol{z}_{.jk}+\boldsymbol{a}_{jk}^{\top(t-1)}\boldsymbol{b}_{jk}^{(t-1)} \\
&+ \left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}(\boldsymbol{I}_D-\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\right)^{-1} \\
&\times \left(\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t-1)}+\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\varepsilon}_{\boldsymbol{\theta}k}^{(t-1)}+\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}t_{jk}^{(t)}\right),
\end{aligned} \tag{21}
$$

where $\boldsymbol{a}_{jk}$ stand for the $(I_k \times D)$ dimensional matrix containing the parameters $\boldsymbol{a}_{i.}$ of every item i answered by subject j of the group k and $\boldsymbol{b}_{jk}$ stand for the $(I_k \times 1)$ vector containing the parameters b_i of every item i answered by subject j of group k.

- *Step 5:* Simulate $\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t)}|(.)$ from $N_D\left(\widehat{\boldsymbol{\Xi}}_{\boldsymbol{\mu}k}^{-1(t-1)}\widehat{\boldsymbol{\varrho}}_{\boldsymbol{\mu}k}^{(t-1)}, \widehat{\boldsymbol{\Xi}}_{\boldsymbol{\mu}k}^{-1(t-1)}\right)$, for $k = 1, ..., K$, mutually independently, where

$$
\begin{aligned}
\widehat{\boldsymbol{\Xi}}_{\boldsymbol{\mu}k}^{(t-1)} &= \boldsymbol{\Psi}_{\boldsymbol{\mu}}^{-1} + n_k\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}(\boldsymbol{I}_D - \boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2(t-1)}\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2(t-1)}\right)^{-1}, \\
\widehat{\boldsymbol{\varrho}}_{\boldsymbol{\theta}k}^{(t-1)} &= \left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}(\boldsymbol{I}_D - \boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{\top(t-1)})\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2(t-1)}\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2(t-1)}\right)^{-1}, \\
&\times \sum_{j=1}^{n_k}\left(\boldsymbol{\theta}_{jk}^{(t)} - \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\varepsilon}_{\boldsymbol{\theta}k}^{(t-1)} + \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{1/2\top(t-1)}\boldsymbol{\Sigma}_{\boldsymbol{Z}k}^{-1/2\top(t-1)}\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}t_{jk}^{(t)}.\right)
\end{aligned} \tag{22}
$$

- *Step 6:* Simulate $\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t)}|(.)$ for $k = 1, ..., K$, mutually independently. A MH step is required

  1. Draw $\boldsymbol{\Psi}_{0\boldsymbol{\theta}k}^{(c)}$ from $q\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t-1)}\right)$.

  2. Let $\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)} = D^{-1/2}\boldsymbol{\Psi}_{0\boldsymbol{\theta}k}^{(c)}D^{-1/2}$ with D the diagonal matrix with elements equal to the principal diagonal of $\boldsymbol{\Psi}_{0\boldsymbol{\theta}k}^{(c)}$. Note that this step is only required on the case that $\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)}$ is a correlation matrix, otherwise $\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)} = \boldsymbol{\Psi}_{0\boldsymbol{\theta}k}^{(c)}$.

  3. Accept $\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)} = \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t)}$ with probability $\pi_j\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t-1)}\right) = min\{R_{\boldsymbol{\Psi}_k}, 1\}$ where

$$
R_{\boldsymbol{\Psi}_k} = \frac{\prod_{j=1}^{n_k} p\left(\boldsymbol{\theta}_{jk}^{(t)}|\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)}, \boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}, t_{jk}^{(t)}\right) p\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)}\right) q\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t-1)}|\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)}\right)}{\prod_{j=1}^{n_k} p\left(\boldsymbol{\theta}_{jk}^{(t)}|\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t-1)}, \boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t-1)}, t_{jk}^{(t)}\right) p\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t-1)}\right) q\left(\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(c)}|\boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t-1)}\right)}, \tag{23}
$$

  $p(.)$ and $q(.)$ stands for the prior and transition densities previously defined.

- *Step 7:* Simulate $\boldsymbol{\delta}_{\boldsymbol{\theta}k}^{(t)}|(.)$ for $k = 1, ..., K$, mutually independently. A MH step is required. For a general D, first draw $\delta_{1k}^{(t)}|\delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}$ then draw $\delta_{2k}^{(t)}|\delta_{1k}^{(t)}, ..., \delta_{Dk}^{(t-1)}$ and repeat this process for all the D elements.

1. Draw $\delta_{1k}^{(c)}$ from $q\left(\delta_{1k}^{(t-1)}, \delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}\right)$,

2. Accept $\delta_{1k}^{(c)} = \delta_{1k}^{(t)}$ with probability $\pi_{1k}\left(\delta_{1k}^{(c)}, \delta_{1k}^{(t-1)}\right) = min\{R_{\delta_{1k}}, 1\}$ where

$$R_{\delta_{1k}} = \frac{\prod_{j=1}^{n_k} p\left(\boldsymbol{\theta}_{jk}^{(t)}|\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t)}, \delta_{1k}^{(c)}, \delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}, t_{jk}^{(t)}\right)}{\prod_{j=1}^{n_k} p\left(\boldsymbol{\theta}_{jk}^{(t)}|\boldsymbol{\mu}_{\boldsymbol{\theta}k}^{(t)}, \boldsymbol{\Psi}_{\boldsymbol{\theta}k}^{(t)}, \delta_{1k}^{(t-1)}, \delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}, t_{jk}^{(t)}\right)}$$
$$\times \frac{p\left(\delta_{1k}^{(c)}|\delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}\right) q\left(\delta_{1k}^{(c)}|\delta_{1k}^{(t-1)}, \delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}\right)}{p\left(\delta_{1k}^{(t-1)}|\delta_{2k}^{(t-1)}, ..., \delta_{Dk}^{(t-1)}\right) q\left(\delta_{1k}^{(t-1)}|\delta_{1k}^{(c)}, \delta_{2k}^{(t-1)} ..., \delta_{Dk}^{(t-1)}\right)},$$

then repeat these steps for $\delta_{2k}^{(t)}, ..., \delta_{Dk}^{(t)}$, here $p(.)$ and $q(.)$ stands for the prior and transition densities previously defined.

- *Step 8:* Simulate the item parameter $\boldsymbol{\zeta}_i^{(t)}$ from $\boldsymbol{\zeta}_i|(.) \sim N_{D+1}\left(\widehat{\boldsymbol{\Lambda}}_i^{(t-1)}, \widehat{\boldsymbol{\zeta}}_j^{(t-1)}\right)$, for $i = 1, ..., I$ mutually independent, where

$$\begin{aligned} \widehat{\boldsymbol{\Lambda}}_i^{(t-1)} &= \left(\boldsymbol{\Psi}_{0a,b}^{-1} + \boldsymbol{H}_{i.}^{\top}\boldsymbol{H}i.\right)^{-1}, \\ \widehat{\boldsymbol{\zeta}}_j^{(t-1)} &= \boldsymbol{H}i.^{\top}\boldsymbol{z}_{i..}^{(t)} + \boldsymbol{\Psi}_{0a,b}^{-1}\boldsymbol{\mu}_{0a,b}, \\ \boldsymbol{H}i. &= [\boldsymbol{\theta}^{(t)}, -\mathbf{1}_N] \bullet I_i, \end{aligned} \quad (24)$$

where I_i is an $(n \times (D+1))$ matrix with elements, in each line, equal to 1 or 0, according to whether or not the item has been or not presented to the the corresponding subject and "$\bullet$" stands for the Hadamard's product, see Horn & Johnson (2012).

The implementation of the convergence acceleration algorithm of Gonzalez (2004) is achieved by:

Fix $(\overline{a_{i2}}^{(t)}, \ldots, \overline{a_{iD}}^{(t)}, \overline{b_i}^{(t)}) = (a_{i2}^{(t)}/a_{i1}^{(t)}, \ldots, a_{iD}^{(t)}/a_{i1}^{(t)}, b_i^{(t)}/a_{i1}^{(t)})$. Then, simulate ν from $f(\nu)$ where $f(\nu) \sim U[0,5;1,5]$ and fix

$$a_{i1}^{(t)} = \nu a_{i1}^{(t)},\ a_{i2}^{(t)} = \nu a_{i1}^{(t)}\overline{a}_{i2}^{(t)}, \ldots, a_{iD}^{(t)} = \nu a_{i1}^{(t)}\overline{a}_{iD}^{(t)},\ b_i^{(t)} = \nu a_{i1}^{(t)}\overline{b}_i^{(t)},$$

with probability

$$pr = \min\left\{ \frac{L(\boldsymbol{y}|\nu a_{i1}^{(t)}, \ldots, \nu b_i^{(t)})\pi(\nu a_{i1}^{(t)}, \ldots, \nu b_i^{(t)})f(1/\nu)}{L(\boldsymbol{y}|a_{i1}^{(t)}, \ldots, b_i^{(t)})\pi(a_{i1}^{(t)}, \ldots, b_i^{(t)})f(\nu)} |\nu^{D+1}|, 1 \right\},$$

and set

$$a_{i1}^{(t)} = a_{i1}^{(t)},\ a_{i2}^{(t)} = a_{i1}^{(t)}\overline{a}_{i2}^{(t)}, \ldots,\ a_{iD}^{(t)} = a_{i1}^{(t)}\overline{a}_{iD}^{(t)},\ b_i^{(t)} = a_{i1}^{(t)}\overline{b}_i^{(t)},$$

with probability $(1 - pr)$, where $L(\boldsymbol{y}|a_{i1}^{(t)}, \ldots, b_i^{(t)})$ is the original likelihood.

- *Step 9:* Draw the guessing parameters $c_i^{(t)}$, from:

$$c_i|(.) \sim \text{beta}\left(\kappa_1 + \sum_{k=1}^{K}\sum_{j=1}^{n_k} I_{ijk}u_{ijk}; \kappa_2 + \sum_{k=1}^{K}\sum_{j=1}^{n_k} I_{ijk}\left(1 - u_{ijk}\right)\right).$$