Bayesian Multidimensional Item Response Theory Modeling Using Working Variables

Alvaro Montenegro[1] and Luisa Parra[1]

Departamento de Estadistica, Universidad Nacional de Colombia, Sede Bogotá
(E-mail: ammontenegrod@unal.edu.co, lfparraar@unal.edu.co)

Abstract. We propose a hybrid Metropolis-Hastings within Gibbs type algorithm with independent proposal distributions to the latent traits and the item parameters in order to fit classical Multidimensional Item Response Theory Models. The independent proposals are all multivariate normal distributions based on the working variables approach applied to the latent traits and the item parameters. The covariance matrix of the latent traits is estimated using an inverse Wishart distribution. The results show that the algorithm is very efficient, effective, and yields to high acceptation rates. The algorithm is applied to real data from a large-scale test applied in Universidad Nacional de Colombia.
Keywords: Multidimensional Item Response Theory, working variables, Bayesian Modeling, Large scale tests.

1 Introduction

The multidimensional item response theory (MIRT) models are based on the assumption that people require more than one basic ability to response correctly to an item on a test. There are two major types of MIRT models, the compensatory model, Reckase [?], [?] and the non-compensatory or partial compensatory model. In this paper we only refer to the compensatory MIRT model, which it will be called simply MIRT model.

Bayesian solutions based on MCMC algorithms are due to Albert [?], Patz [?], and other authors. Kim and Bolt [?] showed how to implement 2pl Bayesian UIRT models using *WinBUGS* which is a free software to implement general Bayesian models.

Full Bayesian inference methods for MIRT models have been proposed by several authors. Patz[?] proposed a framework of Bayesian inference for IRT models. After that work Glas and Beguin [?], Patz [?], Montenegro et al. [?], and other authors have proposed Bayesian implementations of a different kind of IRT models. Classical prior distributions are proposed in those implementations, which led to Gibbs, Metropolis-Hastings, and hybrid algorithms. None of them use independent priors. Specifically for the MIRT models, [?] implemented the three-parameter normal ogive (3PNO) model, by using a Gibbs sampler algorithm based on augmented variables. They included the cases of multiple groups and incomplete data. Patz [?] proposed a classical M-H algorithm for the 3pl MIRT model of simple structure. In that kind of models only one of the slopes in each vector of slopes is different from zero. Bold [?] compared some implementation of Bayesian MIRT models and showed how to carry out them using *WinBUGS*.

In this work, we propose a hybrid Metropolis-Hastings(M-H) within Gibbs algorithm, based on the technique of working variables, Gamerman [**?**], Gutierrez [**?**]. At each iteration an iterative weighted least squares (IWLS) step is included to move the chains toward the mode of the full conditional posterior densities. The IWLS step is based on the algorithm proposed by Nelder and Wedderburn [**?**] in the framework of the generalized linear (GLM) models. Bayesian algorithms based on that technique are M-H within Gibbs hybrid algorithms with independent proposals. Those algortihms usually have high acceptance rates and does not require tunning parameters.

The paper is organized as follows. In section **??** we introduce the 2PL MIRT model and in section **??** we discuss the problem of identifiability of the model and introduce a special parameterization to have an identifiable model. Section **??** presents the complete and full conditioned posterior distributions. The gradient vectors and the information matrices of MIRT models are derived in section **??**. Section **??** presents the details of the proposed Bayesian algorithm. In section **??** we exhibit the results of fitting a set of real data from an admission test. Finally, section **??** is the discussion of the paper.

2 Specification of the 2PL MIRT Model

In the dichotomous multidimensional item response theory (MIRT) models, the data is organized in a $N \times p$ matrix $\boldsymbol{y}$. The ijth element represents the response of person i to item j. It is assumed that the value y_{ij} is the realization of a binary random variable $Y_{ij} \sim Ber(\pi_{ij})$. The value 1 is assigned to a correct response and 0 otherwise. The latent trait of person i is denoted $\boldsymbol{\theta}_i$ and it is assumed to be an independent realization of a random d-vector $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ has some multivariate distribution as $N_m(\mathbf{0}, \boldsymbol{\Sigma})$. The probability π_{ij} is given byIn the dichotomous multidimensional item response theory (MIRT) models, the data is organized in a $N \times p$ matrix $\boldsymbol{y}$. The ijth element represents the response of the person i to the item j. It is assumed that the value y_{ij} is the realization of a binary random variable $Y_{ij} \sim Ber(\pi_{ij})$. The value 1 is assigned to a correct response and 0 otherwise. The latent trait of person i is denoted $\boldsymbol{\theta}_i$ and it is assumed to be an independent realization of a random d-vector $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ has some multivariate distribution as $N_m(\mathbf{0}, \boldsymbol{\Sigma})$. The probability π_{ij} is given by

$$\pi_{ij} = \frac{1}{1 + e^{-\eta_{ij}}}, \tag{1}$$

where $\eta_{ij} = \boldsymbol{a}_j^t \boldsymbol{\theta}_i + d_j = a_{j1}\theta_{i1} + \cdots + a_{jm}\theta_{im} + d_j, \quad a_{jk}, \theta_{ik} \in \Re$. It is common to assume that $a_{jk} > 0$ to guarantee that π_{ij} is a monotonic function of $\boldsymbol{\theta}$. Nevertheless, that assumption is not strictly necessary. In this work, we assume that $-\infty < a_{jk} < \infty$, for all j, k. However, in practical applications where it is required that $a_{jk} > 0$, a logarithmic reparameterization can be introduced to these parameters such that $\log a_{jk} = a^*_{jk}$ and $-\infty < a^*_{jk} < \infty$.

The set of item parameters of the test is denoted by $\boldsymbol{\beta}$. Thus, $\boldsymbol{\beta}$ is a $p \times (m + 1)$ matrix. The set of latent traits of the sample is represented by $\boldsymbol{\theta}$. Then $\boldsymbol{\theta}$ is a $N \times m$ matrix. The following notation is adopted. Let $\boldsymbol{\beta}_j = (\boldsymbol{a}_j^t, d_j)^t$. $\boldsymbol{\beta}_j$ will be called the latent regression parameter of the item j.

Let $\alpha_j = (a_{j1}^2 + \ldots + a_{jm})^{1/2}$ be the Euclidean norm of the vector $\boldsymbol{a}_j$. The value α_j is called the multidimensional item discrimination (MDISC) of item j. The vector $\boldsymbol{\nu}_j = \boldsymbol{a}_j/\alpha_j$ is the direction of the item j and $b_j = -d_j/\alpha_j$ is called the multidimensional item difficulty (MDIFF) of item j. For details see Reckase [?], [?].

3 Identifiability of the 2PL MIRT Model

It is known that the MIRT models are non-identifiable. In fact, if $\boldsymbol{B}$ is a non singular matrix, then $\boldsymbol{a}_j^t\boldsymbol{\theta}_i = (\boldsymbol{B}^t\boldsymbol{a}_j)^t(\boldsymbol{B}^{-1}\boldsymbol{\theta}_i)$. Our solution to have an identifiable model is partially based on the parameterization proposed by [?] in the framework of the nonlinear factor analysis models and Montenegro [?] and Montenegro et al. [?] in the framework of MIRT models.

1. The random d-vector $\boldsymbol{\Theta}$ has a multivariate normal distribution $N_m(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a hyperparameter that must be estimated.
2. We fix the parameters of m items. Yalcin and Amemiya [?] proposed to fix these parameters in a way such that if the items are reorder, the matrix of the selected $\boldsymbol{a}_j$'s is the identity matrix $\boldsymbol{I}_m$ and the corresponding intercept parameters conform the m-vector $\boldsymbol{0}_m$.

In the unidimensional case ($m = 1$), it is common to assume that $\boldsymbol{\Theta} \sim N(0, 1)$. This solution is not sufficient if $m > 1$. If we assume that $\boldsymbol{\Theta}$ has standard multivariate normal distribution $N_m(\boldsymbol{0}, \boldsymbol{I}_m)$ and $\boldsymbol{B}$ is an orthogonal matrix, thus $\boldsymbol{B}^{-1}\boldsymbol{\Theta} \sim N_m(\boldsymbol{0}, \boldsymbol{I}_m)$. This is an indeterminacy of the model by rotations and reflexions. The proposed parameterization overcomes this problem.

4 Posterior density and full conditional marginals

To build the likelihood it is necessary to state the following assumptions.

1. The response patterns among the people are independent.
2. Given the latent trait $\boldsymbol{\theta}_i$, the responses $Y_{ij}, j = 1, \ldots, p$ are independent (local independence).

The probability function of each variable Y_{ij} is given by $f(y_{ij}|\pi_{ij}) = \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}}$. Let us suppose that $p(\boldsymbol{\theta}_i)$ is a prior density for the latent trait of person i and $p(\boldsymbol{\beta}_j)$ is a prior density for the parameter of the item j. Let $\boldsymbol{y}_{i.} = (y_{i1}, \ldots, y_{ip})^t$ be the response pattern of individual i and $\boldsymbol{y}_{.j} = (y_{1j}, \ldots, y_{Nj})^t$ the complete vector of responses to item j. Thus the posterior joint density is specified by

$$p(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y}) \propto \prod_{i=1}^{N}\prod_{j=1}^{p} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}} p(\boldsymbol{\theta}_i)p(\boldsymbol{\beta}_j). \tag{2}$$

The full conditional marginals are specified by

$$p(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y}) \propto \prod_{i=1}^{N}\prod_{j=1}^{p} \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}} p(\boldsymbol{\theta}_i)p(\boldsymbol{\beta}_j). \tag{3}$$

The full conditional marginals are specified by

$$p(\boldsymbol{\theta}_i|\boldsymbol{y}_{i.}, \boldsymbol{\beta}) \propto \prod_{j=1}^{p} \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}} p(\boldsymbol{\theta}_i) \tag{4}$$

$$p(\boldsymbol{\beta}_j|\boldsymbol{y}_{.j}, \boldsymbol{\theta}) \propto \prod_{i=1}^{N} \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}} p(\boldsymbol{\beta}_j). \tag{5}$$

Since all the parameters are real numbers we propose priors based on multivariate normal distributions. For the latent trait vector $\boldsymbol{\theta}_i$ we propose priors $N_m(\boldsymbol{\mu}_{i0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ and for the item vector parameter $\boldsymbol{\beta}_j$ we propose priors $N_{(m+1)}(\boldsymbol{\beta}_{j0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$. In the next section, we derive normal approximations of the full conditional marginals. Such approximations will be used as independent proposals in the proposed algorithm.

5 Gradient and Information Matrices

It is easy to see that

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\theta}_i} = \frac{\partial l_{ij}}{\partial \pi_{ij}}\frac{\partial \pi_{ij}}{\partial \eta_{ij}}\frac{\partial \eta_{ij}}{\partial \boldsymbol{\theta}_i} = (y_{ij}-\pi_{ij})\boldsymbol{a}_j, \tag{6}$$

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} = \frac{\partial l_{ij}}{\partial \pi_{ij}}\frac{\partial \pi_{ij}}{\partial \eta_{ij}}\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}_j} = (y_{ij}-\pi_{ij})\begin{bmatrix}\boldsymbol{\theta}_i\\ 1\end{bmatrix}. \tag{7}$$

We have that $\nabla_{\boldsymbol{\theta}_i} = \frac{\partial L}{\partial \boldsymbol{\theta}_i} = \sum_{j=1}^{P} \frac{\partial l_{ij}}{\partial \boldsymbol{\theta}_i}$ and $\nabla_{\boldsymbol{\beta}_j} = \frac{\partial L}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{N} \frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_i}$.

5.1 Information Matrices

Tra matrices information are given by

$$\mathfrak{I}_{\boldsymbol{\theta}_i} = E\left[\frac{\partial L}{\partial \boldsymbol{\theta}_i}\frac{\partial L}{\partial \boldsymbol{\theta}_i^t}\right] = \sum_{j=1}^{p} \pi_{ij}(1-\pi_{ij})\boldsymbol{a}_j\boldsymbol{a}_j^t, \tag{8}$$

$$\mathfrak{I}_{\boldsymbol{\beta}_j} = E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial \boldsymbol{\beta}_j^t}\right] = \sum_{i=1}^{N} \pi_{ij}(1-\pi_{ij})\begin{bmatrix}\boldsymbol{\theta}_i\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}_i\\ 1\end{bmatrix}^t. \tag{9}$$

Let $w_{ij} = \pi_{ij}(1-\pi_{ij})$, $\boldsymbol{W}_{i.} = \text{diag}(w_{ij}, \ldots, wip)$, $\boldsymbol{W}_{.j} = \text{diag}(w_{1j}, \ldots, w_{Nj})$, $i = 1, \ldots, N$, $j = 1, \ldots, p$. Thus we have that,

$$\Im_{\boldsymbol{\theta}_i} = \boldsymbol{A}^t \boldsymbol{W}_{i.} \boldsymbol{A} \tag{10}$$
$$\Im_{\boldsymbol{\beta}_j} = [\boldsymbol{\theta}|\mathbf{1}_N]^t \boldsymbol{W}_{.j} [\boldsymbol{\theta}|\mathbf{1}_N]. \tag{11}$$

For the unidimensional case the matrix $\boldsymbol{A}$ reduces to the p-vector, whose jth component is a_j and $\boldsymbol{\Theta}$ is the N-vector whose ith component is θ_i.

6 Proposed Bayesian Algorithm

The algorithm proposed in this section is based directly on the Bayesian IWLS algorithm. At each iteration a IWLS step is introduced to improve the approximation, by searching a point closer to the mode of the full conditional distribution than the current position.

6.1 Prior distributions

For the $\boldsymbol{\beta}_j$ vector parameters we propose a prior distribution given by the normal $N(\widetilde{\boldsymbol{\pi}}_\beta, \widetilde{\boldsymbol{\Sigma}}_\beta)$, where $\widetilde{\boldsymbol{\pi}}_\beta = \left(\frac{1}{\sqrt{m}}\mathbf{1}_m^t, 0\right)^t$ and $\widetilde{\boldsymbol{\Sigma}}_\beta = \boldsymbol{I}_{m+1}$. For the latent traits we propose the normal distribution $N(\widetilde{\boldsymbol{\pi}}_\theta, \widetilde{\boldsymbol{\Sigma}}_\theta)$. where $\widetilde{\boldsymbol{\pi}}_\theta = \mathbf{0}_m$, and $\widetilde{\boldsymbol{\Sigma}}_\theta$ is a correlation matrix, with 1s on the diagonal and correlation σ_{st} between $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_t$, $s \neq t$. To model $\widetilde{\boldsymbol{\Sigma}}_\theta$ it was introduced an unconstraint covariance matrix $\boldsymbol{R}$, where $\boldsymbol{R} = [\rho_{st}]$ and such that the constrained covariance matrix $\widetilde{\boldsymbol{\Sigma}}_\theta$ can be obtained from $\boldsymbol{R}$ using

$$\sigma_{st} = \frac{\rho_{st}}{\sqrt{\rho_{ss}\rho_{tt}}}, \quad s \neq t. \tag{12}$$

A non informative prior that can be assumed for $\boldsymbol{R}$ is the Jeffreys' prior given by $p(\theta) \propto |I(\theta)|^{\frac{1}{2}}$, where $I(w)$ is the expected Fisher information matrix of w. In this work, it was used the Jeffreys' prior given by $p(\boldsymbol{R}) \propto |\boldsymbol{R}|^{-(m+1)/2}$. According to [?], the posterior distribution for $\boldsymbol{R}$ is given by the inverse Wishart distribution $IW_m(S, N)$, where $S = \sum_{i=1}^N \boldsymbol{\theta}_i \boldsymbol{\theta}_i^t$.

6.2 The Algorithm

In this section we use the following notation. For a matrix $\boldsymbol{S}$, $\boldsymbol{S}_{i.}$ represents the ith row of $\boldsymbol{S}$, and $\boldsymbol{S}_{.j}$ is its jth column. The symbol $\dot{\boldsymbol{S}}_{i.}$ represents the diagonal matrix whose diagonal $\boldsymbol{S}_{i.}$, and $\dot{\boldsymbol{S}}_{.j}$ is the diagonal matrix whose diagonal is $\boldsymbol{S}_{.j}$. The symbol $\odot$ represents the Hadamard product between matrices or vectors, $\mathbf{1}_k$ is the k vector of ones, and $\boldsymbol{J}_{Np}$ is the $N \times p$ matrix whose elements are all equal to 1. The proposed Bayesian IWLS algorithm for 2PL-MIRT models is as follows.

1. Step 1. Obtain Initial values for both, the item parameters $\boldsymbol{\beta}^*$ and the latent traits $\boldsymbol{\theta}^*$. Set $\boldsymbol{\beta}_j^{(1)} = \boldsymbol{\beta}_j^*$, $j = 1, \ldots, p$ and $\boldsymbol{\theta}_i^{(1)} = \boldsymbol{\theta}_i^*$, $i = 1, \ldots, N$. Let $\widetilde{\boldsymbol{\Sigma}}_\theta^{(1)} = \boldsymbol{I}_m$.

2. Step $k+1$, $k \geq 1$. Block $\boldsymbol{\beta}$. IWLS step.
   (a) $\boldsymbol{\eta}^{(k)} = [\boldsymbol{\theta}^{\overline{(k)}}|\mathbf{1}_N]\boldsymbol{\beta}^{t(k)}$
   (b) $\boldsymbol{\pi}^{(k)} = h^{-1}(\boldsymbol{\eta}^{(k)})$
   (c) $\boldsymbol{z}^{(k)} = \boldsymbol{\eta}^{(k)} + h'(\boldsymbol{\pi}^{(k)}) \odot (\boldsymbol{y} - \boldsymbol{\pi}^{(k)})$
   (d) $\boldsymbol{W}^{(k)} = \boldsymbol{\pi}^{(k)} \odot (\boldsymbol{J}_{Np} - \boldsymbol{\pi}^{(k)}))$
   (e) $\widetilde{\boldsymbol{\beta}}_j^{(k)} = \left([\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]^t \dot{\boldsymbol{W}}_{.j}[\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]\right)^{-1} [\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]^t \dot{\boldsymbol{W}}_{.j}\boldsymbol{z}^{(k)}$, $j = 1,\ldots,p$; (except the fixed item parameters). Let $\widetilde{\boldsymbol{\beta}}^{(k)}$ be the matrix whose rows are $\widetilde{\boldsymbol{\beta}}_j$.
3. Step $k+1$, $k \geq 1$. Block $\boldsymbol{\beta}$. Sampling step
   (a) $\boldsymbol{\eta}^{(k)} = [\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]\widetilde{\boldsymbol{\beta}}^{t(k)}$
   (b) $\boldsymbol{\pi}^{(k)} = h^{-1}(\boldsymbol{\eta}^{(k)})$
   (c) $\boldsymbol{z}^{(k)} = \boldsymbol{\eta}^{(k)} + h'(\boldsymbol{\pi}^{(k)}) \odot (\boldsymbol{y} - \boldsymbol{\pi}^{(k)})$
   (d) $\boldsymbol{W}^{(k)} = \boldsymbol{\pi}^{(k)} \odot (\boldsymbol{J}_{Np} - \boldsymbol{\pi}^{(k)}))$
   (e) For $j = 1,\ldots,p$; (except the fixed item parameters) do
      e.1 $\Im_{\beta_j}^{(k)} = [\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]^t \dot{\boldsymbol{W}}_{.j}^{(k)}[\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]$
      e.2 $\boldsymbol{\Sigma}_{\beta_j}^{(k)} = \left[\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^{-1} + \Im_{\beta_j}^{(k)}\right]^{-1}$
      e.3 $\boldsymbol{\pi}_{\beta_j}^{(k)} = \boldsymbol{\Sigma}_{\beta_j}^{(k)} \left[\widetilde{\boldsymbol{\Sigma}}_{\beta}^{-1}\widetilde{\boldsymbol{\pi}}_{\beta} + [\boldsymbol{\theta}^{(k)}|\mathbf{1}_N]^t \dot{\boldsymbol{W}}_{.j}^{(k)}\boldsymbol{z}_{.j}^{(k)}\right]$
      e.4 Obtain a candidate sample $\boldsymbol{\beta}_j^*$ from $N(\boldsymbol{\pi}_{\beta_j}^{(k)}, \boldsymbol{\Sigma}_{\beta_j}^{(k)})$.
      e.5 Acceptation rule. Let $q_{\beta_j}(\boldsymbol{\beta}_j|\cdot)$ denote the kernel of the density of $N(\boldsymbol{\pi}_{\beta_j}^{(k)}, \boldsymbol{\Sigma}_{\beta_j}^{(k)})$, and $p(\boldsymbol{\beta}_j|\cdot)$ the posterior density of $\boldsymbol{\beta}_j$. Define

$$\alpha_{\beta_j} = \frac{p(\boldsymbol{\beta}_j^*|\boldsymbol{y}_{.j}, \boldsymbol{\theta}^{(k)})/q_{\beta_j}(\boldsymbol{\beta}_j^*|\boldsymbol{\pi}_{\beta_j}^{(k)}, \boldsymbol{\Sigma}_{\beta_j}^{(k)})}{p(\boldsymbol{\beta}_j^{(k)}|\boldsymbol{y}_{.j}, \boldsymbol{\theta}^{(k)})/q_{\beta_j}(\boldsymbol{\beta}_j^{(k)}|\boldsymbol{\pi}_{\beta_j}^{(k)}, \boldsymbol{\Sigma}_{\beta_j}^{(k)})} \tag{13}$$

      If $\alpha_{\beta_j} \geq 1$ accept $\boldsymbol{\beta}_j$. If $\alpha_{\beta_j} < 1$ accept $\boldsymbol{\beta}_j$ with probability α_{β_j}. That is, generate a random value $u \sim U(0,1)$. If $\alpha_{\beta_j} > u$ accept $\boldsymbol{\beta}_j$. If accept $\boldsymbol{\beta}_j$, set $\boldsymbol{\beta}_j^{(k+1)} = \boldsymbol{\beta}_j^*$. Otherwise $\boldsymbol{\beta}_j^{(k+1)} = \boldsymbol{\beta}_j^{(k)}$.
4. Step $k+1, k \geq 1$. Block $\boldsymbol{\Theta}$. IWLS step.
   (a) $\boldsymbol{\eta}^{(k)} = [\boldsymbol{\theta}^{\overline{(k)}}|\mathbf{1}_N]\boldsymbol{\beta}^{t(k+1)}$
   (b) $\boldsymbol{\pi}^{(k)} = h^{-1}(\boldsymbol{\eta}^{(k)})$
   (c) $\boldsymbol{z}^{(k)} = \boldsymbol{\eta}^{(k)} + h'(\boldsymbol{\pi}^{(k)}) \odot (\boldsymbol{y} - \boldsymbol{\pi}^{(k)})$
   (d) $\boldsymbol{W}^{(k)} = \boldsymbol{\pi}^{(k)} \odot (\boldsymbol{J}_{Np} - \boldsymbol{\pi}^{(k)}))$
   (e) $\widetilde{\boldsymbol{\theta}}_i^{(k)} = \left(\boldsymbol{A}^{t(k)}\dot{\boldsymbol{W}}_{i.}^{(k)}\boldsymbol{A}^{(k)}\right)^{-1}\boldsymbol{A}^{t(k)}\dot{\boldsymbol{W}}_{i.}^{(k)}\boldsymbol{z}^{(k)}$, $i = 1,\ldots,N$. Let $\widetilde{\boldsymbol{\theta}}^{(k)}$ be the matrix whose rows are $\widetilde{\boldsymbol{\theta}}_i$.
5. Step $k+1, k \geq 1$. Block $\boldsymbol{\Theta}$. Sampling step.
   (a) $\boldsymbol{\eta}^{(k)} = [\widetilde{\boldsymbol{\theta}}^{(k)}|\mathbf{1}_N]\boldsymbol{\beta}^{t(k+1)}$
   (b) $\boldsymbol{\pi}^{(k)} = h^{-1}(\boldsymbol{\eta}^{(k)})$
   (c) $\boldsymbol{z}^{(k)} = \boldsymbol{\eta}^{(k)} + h'(\boldsymbol{\pi}^{(k)}) \odot (\boldsymbol{y} - \boldsymbol{\pi}^{(k)})$
   (d) $\boldsymbol{W}^{(k)} = \boldsymbol{\pi}^{(k)} \odot (\boldsymbol{J}_{Np} - \boldsymbol{\pi}^{(k)}))$
   (e) For $i = 1,\ldots,N$; do
      e.1 $\Im_{\theta_i}^{(k)} = \boldsymbol{A}^{t(k+1)}\dot{\boldsymbol{W}}_{i.}^{(k)}\boldsymbol{A}^{(k+1)}$

e.2 $\boldsymbol{\Sigma}_{\theta_i}^{(k)} = \left[\left(\widetilde{\boldsymbol{\Sigma}}_{\theta}^{(k)} \right)^{-1} + \mathfrak{I}_{\theta_i}^{(k)} \right]^{-1}$

e.3 $\boldsymbol{\pi}_{\theta_i}^{(k)} = \boldsymbol{\Sigma}_{\theta_i}^{(k)} \left[\left(\widetilde{\boldsymbol{\Sigma}}_{\theta}^{(k)} \right)^{-1} \widetilde{\boldsymbol{\pi}}_{\theta} + \boldsymbol{A}^{t(k)} \dot{\boldsymbol{W}}_{i.}^{(k)} \boldsymbol{z}_{i.}^{(k)} \right]$

e.4 Obtain a candidate sample $\boldsymbol{\theta}_i^*$ from $N(\boldsymbol{\pi}_{\theta_i}^{(k)}, \boldsymbol{\Sigma}_{\theta_i}^{(k)})$.

e.5 Acceptation rule. Let $q_{\theta_i}(\boldsymbol{\theta}_i|\cdot)$ denote the density of $N(\boldsymbol{\pi}_{\theta_i}^{(k)}, \boldsymbol{\Sigma}_{\theta_i}^{(k)})$, and $p(\boldsymbol{\theta}_i|\cdot)$ the posterior density of $\boldsymbol{\theta}_i$. Define

$$\alpha_{\theta_i} = \frac{p(\boldsymbol{\theta}_i^*|\boldsymbol{y}_{i.}, \boldsymbol{\beta}^{(k+1)})/q_{\theta_i}(\boldsymbol{\theta}_i^*|\boldsymbol{\pi}_{\theta_i}^{(k)}, \boldsymbol{\Sigma}_{\theta_i}^{(k)})}{p(\boldsymbol{\theta}_i^{(k)}|\boldsymbol{y}_{i.}, \boldsymbol{\beta}^{(k+1)})/q_{\theta_i}(\boldsymbol{\theta}_i^{(k)}|\boldsymbol{\pi}_{\theta_i}^{(k)}, \boldsymbol{\Sigma}_{\theta_i}^{(k)})} \tag{14}$$

If $\alpha_{\theta_i} \geq 1$ accept $\boldsymbol{\theta}_i$. If $\alpha_{\theta_i} < 1$ accept $\boldsymbol{\theta}_i$ with probability α_{θ_i}. That is, generate a random value $u \sim U(0,1)$. If $\alpha_{\theta_i} > u$ accept $\boldsymbol{\theta}_i$. If accept $\boldsymbol{\theta}_i$, set $\boldsymbol{\theta}_i^{(k+1)} = \boldsymbol{\theta}_i^*$. Otherwise $\boldsymbol{\theta}_i^{(k+1)} = \boldsymbol{\theta}_i^{(k)}$.

6. Step $k+1$. Block $\boldsymbol{\Sigma}$.
   (a) Obtain a sample $\boldsymbol{R}$ from $IW_m(S^{(k)}, N)$, where $S^{(k)} = \sum_{i=1}^N \boldsymbol{\theta}_i^{(k)}(\boldsymbol{\theta}_i^{(k)})^t$.
   (b) Obtain $\widetilde{\boldsymbol{\Sigma}}_{\theta}^{(k+1)}$ from $\boldsymbol{R}$ by using equation (**??**).

An implementation of the algorithm in R, with scripts for simulation data and real data can be obtianed by request to the authors.

7 Real data

The data are from the admission test at the Universidad Nacional de Colombia, applied in the second semester of 2009. We took a sample size of N=1449. The test was taken by more than 35,000 people. There were seven forms of the test, but the only difference between them was the order of the questions. The data correspond to a sample from one form. The test size was $p = 95$ with 5 subtests. The subtests were: mathematics (Math) with $p_1 = 20$ items, natural sciences (Science) with $p_2 = 25$ items, social sciences (Social) with $p_3 = 26$, textual analysis (Textual) with $p_4 = 15$ items, and image analysis (Image) with $p_5 = 9$ items. The items 10, 21, 55, 75, 95 were fixed. The results are based on 6000 iterations after a burn-in period of 2500 iterations. Table **??** shows a summary of the acceptance rates for both, the item parameters and the latent traits. The sampling chains passed the Heidelberger-Welch stationarity test for almost all the cases.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
latent traits	0.10	0.4678	0.5692	0.5478	0.6518	0.8352
item parameters	0.8662	0.9364	0.9517	0.9467	0.9615	0.9758

Table 1. Summary of the acceptation rates. Real data.

Table **??** shows a summary of the item parameters estimated by the Bayesian algorithm. Some of the discrimination parameters are negative, but they are small. We did not do any special treatment with those values. However, it is easy to include the logarithm transformation in the model to obtain only positive values.

	$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	$\boldsymbol{a}_3$	$\boldsymbol{a}_4$	$\boldsymbol{a}_5$	d
Min.	-0.20658	-0.1728	-0.228664	-0.1487	-0.19353	-2.3188
1st Qu.	0.00973	0.1352	-0.002583	0.1383	0.02684	-0.9227
Median	0.16772	0.2357	0.132120	0.2248	0.14929	-0.3533
Mean	0.20694	0.2504	0.161899	0.2882	0.18663	-0.1765
3rd Qu.	0.34229	0.3513	0.256368	0.4058	0.27698	0.4400
Max.	2.22373	1.0544	1.700400	1.3903	1.51406	2.1505

Table 2. Summary of the item parameters. Real data.

Finally, table **??** shows the main statistics of the latent traits estimated by the Bayesian algorithm.

	$\boldsymbol{\theta}_1$	$\boldsymbol{\theta}_2$	$\boldsymbol{\theta}_3$	$\boldsymbol{\theta}_4$	$\boldsymbol{\theta}_5$
Min.	-1.88994	-2.084656	-1.96438	-2.45430	-2.09705
1st Qu.	-0.55483	-0.423623	-0.52520	-0.46252	-0.47362
Median	-0.17325	-0.005061	-0.08354	0.01508	0.02279
Mean	-0.04458	0.006222	-0.02150	0.01428	0.01376
3rd Qu.	0.38927	0.417533	0.48706	0.52226	0.53826
Max.	2.73310	2.358814	2.21566	2.10364	2.02972

Table 3. Summary of the latent traits. Real data.

8 Discussion

We have introduced a IWLS Bayesian algorithm to fit a 2PL MIRT model. The proposed algorithm is a hybrid MH within Gibbs sampler with independent proposals for the latent traits and the item parameters. The algorithm has several characteristics that promise a very good performance in other IRT models. The first important issue is that the proposal densities are normal distributions derived from a linear approximation of the link function at the observations and around their expect values. This is the basis of the working variables technique. Even though the MIRT models can be considered mixed models with repeated observations, we did not use that fact. Instead we treated both, the latent traits and the item parameters in a symmetrical way. In the β block we fixed the latent traits and obtained the normal approximation of the posterior densities of the item parameters, introducing the corresponding working variables. In the θ block we fixed the item parameters and obtained the normal approximation of the posterior densities of latent traits introducing the corresponding working variables and considering the covariance matrix obtained in the previouos step for the prior densities.

References

1. J. Albert. Bayesian Estimation of Normal Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*, 17,3,251-269, 1992
2. D. Bold and V.F. Lall. Estimation of Compensatory and non-compensatory multidimensional Item Response Theory Models Using Markov Chain Monte Carlo. *Applied Psychological Measurement*, 27,6, 395-414, 2003.
3. D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing, 7, 57-68, 1997.
4. A. Gutierrez, The Use of Working Variables in the Bayesian Modeling of Mean and Dispersion Parameters in Generalized Nonlinear Models with Random Effects, *Communications in Statistics-Simulation and Computation*, 44, 1, 168-195, 2015.
5. C.A. Glas and A. Béguin. MCMC Estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66, 4, 21., 2001.
6. J. Kim, D. and Bolt. Estimating Item Response Theory Models Using Markov Chain Monte Carlo Methods. *Educational Measurement: Issues and Practice*,26, 38-51, 2007.
7. A. Montenegro. Multidimensional Item Response Theory Models where the Ability has a Linear Latent Structure, PhD dissertation, Univesidad Nacional e Colombia, 2010.
8. A. Montenegro, L. Parra, E. Berdugo. A MIRT Model for Tests with Multiple Subtests. *IOSR-JRME*, 6,6 7-20,2016.
9. J. Nelder, and R, Wedderburn. Generalized Linear Models. *Journal of the Royal statistical society: Series A*, 135, 370-384,1972.
10. A Straighforward Approach to Markov Chain Monte Carlo Methods for item Response Models. *Journal of Educational and Behavioral Statistics*, 24, 2 , 146-178, 1999.
11. J. De la Torre and R. Patz. 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
12. M. Reckase. The difficulty of test items that measure more than one ability. *Applied Psycological Measurement*, 9, 9, 401-412, 1985.
13. M. Reckase, *Multidimensional Item Response Theory*, Springer, 2009.
14. I. Yalcin and Y. Amemiya, Nonlinear Factor Analysis as a Statistical Method. *Statistical Science*, 16,3,275-294, 2001.