Journal of Statistical Computation and Simulation
Vol. 80, No. 11, November 2010, 1237–1252

Bayesian estimation of the multidimensional graded response model with nonignorable missing data

Zhi-Hui Fu[a,b], Jian Tao[c]* and Ning-Zhong Shi[c]

[a] *Research Institute of Mathematics, Jilin University, Changchun, Jilin 130012, People's Republic of China;* [b] *Department of Statistics, School of Mathematics and System Science, Shenyang Normal University, Shenyang, Liaoning 110034, People's Republic of China;* [c] *Key Laboratory of Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, People's Republic of China*

(*Received 7 November 2008; final version received 8 May 2009*)

A Bayesian approach is developed for analysing item response models with nonignorable missing data. The relevant model for the observed data is estimated concurrently in conjunction with the item response model for the missing-data process. Since the approach is fully Bayesian, it can be easily generalized to more complicated and realistic models, such as those models with covariates. Furthermore, the proposed approach is illustrated with item response data modelled as the multidimensional graded response models. Finally, a simulation study is conducted to assess the extent to which the bias caused by ignoring the missing-data mechanism can be reduced.

Keywords: Bayesian estimation; data augmentation; Gibbs sampling; item response theory; nonignorable missing data

AMS Classification: Primary: 62F15; Secondary: 65C60, 62M05

1. Introduction

In educational measurement, it often happens that item nonresponses are nonignorable missing data. An example is a test with a time-limit condition, where examinees of lower ability do not reach the items at the end. Thus, the pattern of missingness in this case depends on the ability that is measured and hence the missing data are not generally ignorable. However, if the given data set contains missing observations, the mechanism causing this missingness can be characterized by its variety of randomness [1] as missing at random (MAR) and missing completely at random.

Suppose $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are the parameters of the observed data and the missing-data process, respectively, and $\mathbf{D}$ is the missing data indicator matrix with elements $d_{ik} = 1$ if a realization x_{ik} was observed, and $d_{ik} = 0$ if x_{ik} was missing for a person i and an item k. Following Rubin's definition, missing data are said to be MAR if the probability of $\mathbf{D}$ given the observed data x_{obs}, missing data x_{mis}, some parameter $\boldsymbol{\zeta}$ and observed covariates y does not depend on the missing data x_{mis},

*Corresponding author. Email: taoj@nenu.edu.cn

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650903029276
http://www.informaworld.com

that is, if

$$P(\mathbf{D}|x_{\text{obs}}, x_{\text{mis}}, \boldsymbol{\zeta}, y) = P(\mathbf{D}|x_{\text{obs}}, \boldsymbol{\zeta}, y).$$

Furthermore, the parameters $\boldsymbol{\zeta}$ and $\boldsymbol{\theta}$ are distinct if there are no functional dependencies, that is, restrictions on the parameter space (frequentist version) or if the prior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are independent (Bayesian case). If these two components (MAR and distinctness) are satisfied, then the missing data are said to be ignorable, otherwise the missing data are nonignorable.

If the missing data cannot be ignored, a concurrent probability model must be defined for the observed and missing data, and inferences are made averaging over the missing data. In this article, item response theory (IRT) models on the missing data indicators will be considered for taking nonignorable missing data mechanisms into account. The model is closely related to the models proposed by Moustaki & O'Muircheartaigh [2] (see also [3–6]). Based on the marginal maximum likelihood (MML) estimation approach, Holman and Glas obtained concurrent parameter estimation for both the observed dichotomous or polytomous responses model and the missing data indicators model. However, a disadvantage of the MML procedure is the limited number of latent variables that can be analysed due to the multiple integrals [7, p. 214].

In this article, we propose a novel Bayesian approach for analysing multidimensional IRT model with nonignorable missing data. In this Gibbs sampling approach, it is easy to handle different kinds of prior information, and the convergence is fast. In the IRT setting: Albert [8] proposed the Gibbs sampler for normal ogive models; Ghosh *et al.* [9] examined integrability of the associated posterior distributions; Patz and Junker [10] developed a Markov chain Monte Carlo (MCMC) estimation method under the MAR assumption with an ignorable missing mechanism. Here, we aim to analyse nonignorable missing data. Based on an efficient data augmentation scheme (DAGS), Sahu [11] fitted the unidimensional three parameter normal ogive (3PNO) model and provided a much faster implementation of the Gibbs sampler. In this article, we extend Sahu's DAGS to the multidimensional and polytomous model. We fit the multidimensional graded response model (MGRM) [12] for the observed responses and the multidimensional two parameter logistic (2PL) model for the missing data. Below, the Gibbs sampling algorithm is used for sampling parameter values for the item parameters and the ability parameters from their posterior distributions. Using the method of data augmentation, realizations from a complicated posterior density can be obtained by augmenting the variables of interest by one or more additional variables such that sampling from the full conditional distributions (FCDs) is easy. We develop an auxiliary variable method for graded response models (GRMs) that handle different prior distributions in a flexible way. The augmented data are defined in such a way that each FCD becomes an indicator function with bounds specified by the other parameter values. As a result, the sampling of the parameters is easy to implement.

This article consists of five sections and is organized in the following manner. After this introduction, the general IRT models for both the observed data and missing data are presented. The following section describes the Bayesian estimation procedure. In Section 4, a simulation study is implemented to evaluate the performance of the proposed method. We conclude with some concluding remarks in Section 5. Finally, some computational details are presented in Appendix.

2. Modelling the missing process

2.1. *A general IRT model for missing-data process and observed responses*

Let $\mathbf{X}$ be a two-dimensional data matrix with (i, k) element X_{ik}, where persons are indexed as $i = 1, \ldots, N$ and items are indexed as $k = 1, \ldots, K$. If a combination of i and k has been observed, then the entry X_{ik} is the observation x_{ik}, otherwise it is equal to some arbitrary constant, which is different from all the observations of X_{ik}. That is, we let $p(X_{ik} = x_{ik}|d_{ik} = 0, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) = 1$ for all x_{ik} or an arbitrary value of x_{ik} that indicates a missing observation, i.e. the conditional distribution of X_{ik} given $d_{ik} = 0$ degenerates, here the $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ denote the discrimination and difficulty parameter, respectively. We define a design matrix $\mathbf{D}$ of the same dimension as $\mathbf{X}$ with element $d_{ik} = 1$ if X_{ik} was observed, otherwise $d_{ik} = 0$. Using the elements of $\mathbf{X}$ and $\mathbf{D}$, one of our objectives is to make inference on the individual person parameter $\boldsymbol{\theta}_i$, $i = 1, \ldots, N$, which are potentially influenced by a latent person variable $\boldsymbol{\zeta}$ representing the missing-data process.

For the observed responses X_{ik}, we consider items with dichotomous and polytomous responses and they will be analysed in general using the MGRM. The probability that person i scores in category g on item k is modelled by the MGRM as

$$\Phi_{kg}(\boldsymbol{\theta}_i) = p(X_{ik} = g|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) = \Phi^*_{kg}(\boldsymbol{\theta}_i) - \Phi^*_{k,g+1}(\boldsymbol{\theta}_i), \tag{1}$$

where $i = 1, \ldots, N$, $g = 0, 1, \ldots, m_k$, and $k = 1, \ldots, K$. Here we use $\Phi^*_{kg}(\boldsymbol{\theta}_i)$ to denote the boundary probability for examinee i having a score larger or equal to g on item k, and the boundary curve is given by

$$\Phi^*_{kg}(\boldsymbol{\theta}_i) = \frac{\exp(\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{kg})}{1 + \exp(\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{kg})}, \tag{2}$$

where $g = 1, \ldots, m_k$, $\Phi^*_{k0}(\boldsymbol{\theta}_i) = 1$ and $\Phi^*_{k,m_k+1}(\boldsymbol{\theta}_i) = 0$. Furthermore, the boundaries between the response categories are represented by an ordered vector of thresholds

$$\beta_{k0} < \beta_{k1} < \beta_{k2} < \cdots < \beta_{kg} < \cdots < \beta_{k,m_k+1} \tag{3}$$

with $\beta_{k0} = -\infty$, $\beta_{k,m_k+1} = +\infty$. Consequently, there are a total of m_k threshold parameters and one discrimination parameter for each item.

To model the missing-data process, we use a Q_2-dimensional IRT model proposed by Reckase [13] and Ackerman [14]. This model (corresponding to model 2 in Holman and Glas [6]), which is in logistic form, has the probability of an observation given by

$$\phi_{ik} = p(d_{ik} = 1|\boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k) = \frac{\exp(\sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k)}{1 + \exp(\sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k)}, \tag{4}$$

where γ_{kq} and δ_k are the item parameters (discrimination and difficulty of the missing data indicator, which we will also refer to as the missing-data process). Model (4) is a specific form of model (1) when $m_k = 1$, further it reduces to the multidimensional (specifically Q_2-dimensional) Rasch model upon setting $\boldsymbol{\gamma}_k = \mathbf{1}$. The multidimensional Rasch model specializes further to the famous Rasch model for dichotomous items by setting $Q_2 = 1$. For more discussion of the relationship among these models, refer to Holman and Glas [6] and Verhelst and Glas [15].

2.2. *Combined IRT models for missing data and observed data*

Suppose $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are the person's latent variables related to the observed data and the missing data, with densities $g_1(\boldsymbol{\theta})$ and $g_2(\boldsymbol{\zeta})$, respectively. Let $p(x_{ik}|d_{ik}, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$ be the measurement model for the observed data. It is the probability of the response (observed) variable conditioned on the latent variable of the observed data, the design variable (missing data indicator) and item parameters. Let $p(d_{ik}|\boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k)$ be the measurement model for the missing data indicator. It is the probability of the design variable conditioned on the latent variable and item parameters for the missing-data process. We will use two models in our estimation procedure. The first model (corresponding to model 3 in Holman and Glas [6]), which we call the MAR model, is given in likelihood form as

$$G_1 = \prod_{i,k} p(x_{ik}|d_{ik}, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) p(d_{ik}|\boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k) g_1(\boldsymbol{\theta}_i) g_2(\boldsymbol{\zeta}_i). \tag{5}$$

It is the model that ignores the missing-data process, and we ignore the model for the missing-data process $p(d_{ik} = 1|\boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k)$ in the estimation process. The latent variables for the observed data and the missing-data process are not related in the MAR model. The second model, which we call the nonignorable model (NONMAR), is the model where the missing-data process is included in the estimation process. In this model, the latent variables for both the observed and missing-data process $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are correlated by $\boldsymbol{\Sigma}$. This model (corresponding to model 4 in Holman and Glas [6]) is written in likelihood form as

$$G_2 = \prod_{i,k} p(x_{ik}|d_{ik}, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) p(d_{ik}|\boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k) g(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{6}$$

where $g(\cdot)$ is the density of $\boldsymbol{\theta}_i$ and $\boldsymbol{\zeta}_i$. It is assumed to follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance–covariance $\boldsymbol{\Sigma}$, which can be used to index the extent to which ignorability holds. Expressions (5) and (6) will be used in our Bayesian procedure to make inference on the estimation of the model parameters. Béguin and Glas [16] give conditions for the identification of the model. From their conclusions, it follows that the basis of the two-dimensional latent space can always be transformed in such a way that both the model for the observations and the model for the missing data indicators depend on the same two latent variables. Therefore, the latent parameters of the two models are not distinct. In other words, within the framework of the model they are functionally dependent.

3. MCMC estimation procedure for a nonignorable IRT model

In this section, an MCMC procedure will be used to sample the posterior distributions of interest. The needed chains will be constructed using the Gibbs sampler [17]. To implement the Gibbs sampler, the parameter vector is divided into a number of components, and each successive component is sampled from its conditional distribution given sampled values for all other components. This sampling scheme is repeated until the sampled values form stable estimates of the posterior distributions. Sahu applies Gibbs sampling to estimate the parameters of the well-known unidimensional 3PNO model. In this section, the procedure will be generalized to the mixture of the multidimensional GRM and the multidimensional 2PL model. Let $\boldsymbol{\lambda} = (\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the collection of the unknown parameters. Given the formulation of above IRT models, proportional to a constant, the posterior $p(\boldsymbol{\lambda}|\mathbf{X}, \mathbf{D})$ can be written as follows:

$$\prod_{i,k} p(x_{ik}|d_{ik}, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) p(d_{ik}|\boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k) g(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{7}$$

The simultaneous posterior distribution of all model parameters is quite complicated. However, based on an efficient data augmentation scheme, the complete set of parameters can be split up into a number of subsets in such a way that the conditional posterior distribution of every subset given all other parameters has a tractable form and can be easily sampled.

To implement the Gibbs sampler for Equation (7), we augment the data by introducing independent random variables U_{ik} and V_{ik}, each having a Uniform $(0, 1)$ distribution. It is assumed that $X_{ik} = g$ if $\Phi^*_{k,g+1}(\boldsymbol{\theta}_i) \le U_{ik} \le \Phi^*_{kg}(\boldsymbol{\theta}_i)$, $g = 0, 1, \ldots, m$ (to keep the presentation simple, we assume all the items have m categories, so we dropped index k from m_k). For the missing indicator d_{ik}, we define $V_{ik} \leq \phi_{ik}$ if $d_{ik} = 1$ and $V_{ik} > \phi_{ik}$ otherwise. The item parameters $\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\beta}$ collected in a vector $\boldsymbol{\xi}$ have a prior $\prod_{k=1}^{K}[\prod_{q=1}^{Q_1} I(\alpha_{kq} > 0) \prod_{q=1}^{Q_2} I(\gamma_{kq} > 0)]$, which insures that the discrimination parameters are positive (where $I(\cdot)$ is an indicator variable taking the value one if its argument is true). Let $\boldsymbol{\eta}_i = (\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i)$, to model the relation between the latent variables, it will be assumed that the latent variables of person i, $\boldsymbol{\eta}_i$ have a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, that is, $\boldsymbol{\eta}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. And a noninformative prior for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is assumed, that is, $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto 1$.

The aim of the procedure is to simulate samples from the joint posterior distribution of $\boldsymbol{\lambda}$, given by

$$\begin{aligned} p(\boldsymbol{\lambda}, \mathbf{U}, \mathbf{V}|\mathbf{X}, \mathbf{D}) &\propto p(\mathbf{U}|\mathbf{X}, \mathbf{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) p(\mathbf{V}|\mathbf{D}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\zeta}) \phi(\boldsymbol{\theta}, \boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\xi}) \\ &= \prod_{i=1}^{N} \left\{ \prod_{k=1}^{K} p(U_{ik}|x_{ik}, d_{ik}, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) p(V_{ik}|d_{ik}, \boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k) \right\} \phi(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\quad \times \prod_{k=1}^{K} \left[\prod_{q=1}^{Q_1} I(\alpha_{kq} > 0) \prod_{q=1}^{Q_2} I(\gamma_{kq} > 0) \right], \end{aligned}$$

where $\mathbf{U} = (U_{ik})_{N \times K}$, $\mathbf{V} = (V_{ik})_{N \times K}$ and $\phi(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate normal density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. From the definition of U_{ik} and V_{ik}, it follows that $p(\boldsymbol{\lambda}, \mathbf{U}, \mathbf{V}|\mathbf{X}, \mathbf{D})$ is proportional to

$$\begin{aligned} &\prod_{i=1}^{N} \left\{ \prod_{k=1}^{K} \left[\sum_{g=0}^{m} I(X_{ik} = g) I_{(\Phi^*_{k,g+1}, \Phi^*_{kg})}(U_{ik}) \right]^{d_{ik}} [I_{(0,\phi_{ik})}(V_{ik})^{d_{ik}} I_{(\phi_{ik},1)}(V_{ik})^{1-d_{ik}}] \right\} \\ &\quad \times \phi(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{k=1}^{K} \left[\prod_{q=1}^{Q_1} I(\alpha_{kq} > 0) \prod_{q=1}^{Q_2} I(\gamma_{kq} > 0) \right], \end{aligned}$$

further, by inserting expressions (2) and (4) into the above formula, it can be rewritten as follows:

$$\begin{aligned} &\prod_{i=1}^{N} \left\{ \prod_{k=1}^{K} \left[\sum_{g=0}^{m} I(X_{ik} = g) \cdot I \left(\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{k,g+1} \leq \log\left(\frac{U_{ik}}{1 - U_{ik}}\right) \leq \sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{kg} \right) \right]^{d_{ik}} \right. \\ &\quad \left. \times\ I \left[\sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k > \log\left(\frac{V_{ik}}{1 - V_{ik}}\right) \right]^{d_{ik}} I \left[\sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k < \log\left(\frac{V_{ik}}{1 - V_{ik}}\right) \right]^{1-d_{ik}} \right\} \\ &\quad \times \phi(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{k=1}^{K} \left[\prod_{q=1}^{Q_1} I(\alpha_{kq} > 0) \prod_{q=1}^{Q_2} I(\gamma_{kq} > 0) \right]. \end{aligned} \tag{8}$$

Although the distribution given by Equation (8) has an intractable form, the conditional distributions of the components $\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$, $\mathbf{U}$ and $\mathbf{V}$ are each tractable and easy to sample from. A draw from the conditional distribution (8) can be obtained in the following steps.

Step 1. Sampling the latent data of the observed response and missing indicator: $\mathbf{U}$ and $\mathbf{V}$.

Given the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, the latent variables U_{ik} are independent, and

$$U_{ik} \mid d_{ik} = 1, \quad X_{ik} = g, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k \sim \text{Uniform}(\Phi^*_{k,g+1}(\boldsymbol{\theta}_i), \Phi^*_{kg}(\boldsymbol{\theta}_i)), \tag{9}$$

similarly,

$$V_{ik} \mid d_{ik}, \boldsymbol{\zeta}_i, \boldsymbol{\gamma}_k, \delta_k \sim \begin{cases} \text{Uniform}(0, \phi_{ik}), & \text{if } d_{ik} = 1, \\ \text{Uniform}(\phi_{ik}, 1), & \text{if } d_{ik} = 0. \end{cases} \tag{10}$$

Step 2. Sampling $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

From Equation (8), we have the condition that

$$\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{kg} \le \log\left(\frac{U_{ik}}{1-U_{ik}}\right) \le \sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{k,g-1} \quad \text{if } X_{ik} = g-1,$$

and

$$\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{k,g+1} \le \log\left(\frac{U_{ik}}{1-U_{ik}}\right) \le \sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \beta_{kg} \quad \text{if } X_{ik} = g,$$

or equivalently

$$\beta_{kg} \ge \sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \log\left(\frac{U_{ik}}{1-U_{ik}}\right) \quad \text{if } X_{ik} = g-1,$$

$$\beta_{kg} \le \sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \log\left(\frac{U_{ik}}{1-U_{ik}}\right) \quad \text{if } X_{ik} = g,$$

also $\beta_{k,g-1} < \beta_{kg} < \beta_{k,g+1}$, so that

$$\beta^{(\mathrm{L})}_{kg} \le \beta_{kg} \le \beta^{(\mathrm{R})}_{kg},$$

where

$$\beta^{(\mathrm{L})}_{kg} = \max_{i \in B_{k,g-1}} \left\{ \max\left[\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \log\left(\frac{U_{ik}}{1-U_{ik}}\right), \beta_{k,g-1} \right] \right\},$$

$$\beta^{(\mathrm{R})}_{kg} = \min_{i \in B_{kg}} \left\{ \min\left[\sum_{q=1}^{Q_1} \alpha_{kq}\theta_{iq} - \log\left(\frac{U_{ik}}{1-U_{ik}}\right), \beta_{k,g+1} \right] \right\},$$

in which

$$B_{k,g-1} = \{i : X_{ik} = g-1\}, \quad B_{kg} = \{i : X_{ik} = g\},$$

then the FCD of $\beta_{kg},\ k = 1, 2, \ldots, K,\ g = 1, 2, \ldots, m$, given all the other parameters and $\mathbf{U}$, $\mathbf{X}$, $\mathbf{D}$, is

$$\beta_{kg} \mid \cdot \sim \text{Uniform}(\beta^{(\mathrm{L})}_{kg}, \beta^{(\mathrm{R})}_{kg}). \tag{11}$$

Similarly, from Equation (8), we have

$$\sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k \ge \log\left(\frac{V_{ik}}{1-V_{ik}}\right) \quad \text{if } d_{ik} = 1,$$

then the FCD of each $\delta_k,\ k = 1, 2, \dots, K$ is

$$\delta_k \,|\cdot \sim \text{Uniform}(\delta_k^{(L)}, \delta_k^{(R)}), \tag{12}$$

where

$$\delta_k^{(L)} = \max_{i \in C_k^1} \left\{ \sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k - \log\left(\frac{V_{ik}}{1 - V_{ik}}\right) \right\},$$

$$\delta_k^{(R)} = \min_{i \in C_k^2} \left\{ \sum_{q=1}^{Q_2} \gamma_{kq}\zeta_{iq} - \delta_k - \log\left(\frac{V_{ik}}{1 - V_{ik}}\right) \right\},$$

in which C_k^1 and C_k^2 are the following sets

$$C_k^1 = \{i : d_{ik} = 0\} \quad \text{and} \quad C_k^2 = \{i : d_{ik} = 1\}.$$

Step 3. Sampling $\boldsymbol{\eta}$.

Here $\boldsymbol{\eta}_i = (\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i) = (\theta_{i1}, \dots, \theta_{iq}, \dots, \theta_{iQ_1}, \zeta_{i1}, \dots, \zeta_{iq}, \dots, \zeta_{iQ_2})$, let $W = Q_1 + Q_2$ for simplification, similar to the sampling of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, the FCD of each $\eta_{iw},\ i = 1, 2, \dots, N,\ w = 1, 2, \dots, W$, is

$$\eta_{iw} | \boldsymbol{\eta}_i(w), \boldsymbol{\xi}, \mathbf{U}, \mathbf{V}, \mathbf{X}, \mathbf{D} \sim N(\nu_w, \tau_w^2) I(\eta_{iw}^{(L)} \le \eta_{iw} \le \eta_{iw}^{(R)}), \tag{13}$$

where $\boldsymbol{\eta}_i(w)$ is the vector $\boldsymbol{\eta}_i$ with η_{iw} deleted, that is,

$$\boldsymbol{\eta}_i(w) = \begin{cases} (\eta_{i2}, \eta_{i3}, \dots, \eta_{iW})^{\mathrm{T}}, & \text{if } w = 1, \\ (\eta_{i1}, \dots, \eta_{i,w-1}, \eta_{i,w+1}, \dots, \eta_{iW})^{\mathrm{T}}, & \text{if } w \ne 1, \end{cases}$$

$N(\nu_w, \tau_w^2)$ is the conditional distribution of η_{iw} given $\boldsymbol{\eta}_i(w)$ with

$$\nu_w = E(\eta_{iw} | \boldsymbol{\eta}_i(w)) \quad \text{and} \quad \tau_w^2 = \text{Var}(\eta_{iw} | \boldsymbol{\eta}_i(w))$$

(the details for calculating the ν_w and τ_w^2 are given in appendix), and

$$\eta_{iw}^{(L)} = \begin{cases} \max\limits_{g \in M} \left\{ \max\limits_{k \in A_{ig}} \dfrac{[\log(U_{ik}/(1 - U_{ik})) + \beta_{kg} - \sum_{s \ne q} \alpha_{ks}\theta_{is}]}{\alpha_{kq}} \right\}, & \text{if } w = 1, 2, \dots, Q_1, \\ \max\limits_{\{k: d_{ik} = 1\}} \left\{ \dfrac{[\log(V_{ik}/(1 - V_{ik})) - \sum_{s \ne q} \gamma_{ks}\zeta_{is} + \delta_k]}{\gamma_{kq}} \right\}, & \text{if } w = Q_1 + 1, \dots, W, \end{cases}$$

$$\eta_{iw}^{(R)} = \begin{cases} \min\limits_{g \in M} \left\{ \min\limits_{k \in A_{ig}} \dfrac{[\log(U_{ik}/(1 - U_{ik})) + \beta_{k,g+1} - \sum_{s \ne q} \alpha_{ks}\theta_{is}]}{\alpha_{kq}} \right\}, & \text{if } w = 1, 2, \dots, Q_1, \\ \min\limits_{\{k: d_{ik} = 0\}} \left\{ \dfrac{[\log(V_{ik}/(1 - V_{ik})) - \sum_{s \ne q} \gamma_{ks}\zeta_{is} + \delta_k]}{\gamma_{kq}} \right\}, & \text{if } w = Q_1 + 1, \dots, W, \end{cases}$$

where $A_{ig},\ M$ are the sets

$$A_{ig} = \{k : X_{ik} = g\}, \quad M = \{0, 1, 2, \dots, m\}.$$

Step 4. Sampling $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

Similarly, from Equation (8) we have that (assuming $\theta_{iq} \neq 0$)

$$\frac{[\log(U_{ik}/(1-U_{ik})) + \beta_{kg} - \sum_{s\neq q} \alpha_{ks}\theta_{is}]}{\theta_{iq}} \leq \alpha_{kq} \leq \frac{[\log(U_{ik}/(1-U_{ik})) + \beta_{k,g+1} - \sum_{s\neq q} \alpha_{ks}\theta_{is}]}{\theta_{iq}}$$

for all i for which $X_{ik} = g$ and $\theta_{iq} > 0$, and

$$\frac{[\log(U_{ik}/(1-U_{ik})) + \beta_{k,g+1} - \sum_{s\neq q} \alpha_{ks}\theta_{is}]}{\theta_{iq}} \leq \alpha_{kq} \leq \frac{[\log(U_{ik}/(1-U_{ik})) + \beta_{kg} - \sum_{s\neq q} \alpha_{ks}\theta_{is}]}{\theta_{iq}}$$

for all i for which $X_{ik} = g$ and $\theta_{iq} < 0$, combining with the prior $I(0, +\infty)$, we have

$$\alpha_{kq} \mid \cdot \sim \text{Uniform}(\alpha_{kq}^{(L)} \leq \alpha_{kq} \leq \alpha_{kq}^{(R)}) I(\alpha_{kq} > 0), \quad k = 1, 2, \ldots, K, \quad q = 1, 2, \ldots, Q_1, \tag{14}$$

if we let

$$\tau_{ikq}^{g} = \frac{[\log(U_{ik}/(1-U_{ik})) + \beta_{kg} - \sum_{s\neq q} \alpha_{ks}\theta_{is}]}{\theta_{iq}},$$

$$\tau_{ikq}^{g+1} = \frac{[\log(U_{ik}/(1-U_{ik})) + \beta_{k,g+1} - \sum_{s\neq q} \alpha_{ks}\theta_{is}]}{\theta_{iq}},$$

then

$$\alpha_{kq}^{(L)} = \max_{g\in M} \left\{ \max_{i\in\Lambda_{kg}} \{\min(\tau_{ikq}^{g}, \tau_{ikq}^{g+1})\} \right\},$$

$$\alpha_{kq}^{(R)} = \min_{g\in M} \left\{ \min_{i\in\Lambda_{kg}} \{\max(\tau_{ikq}^{g}, \tau_{ikq}^{g+1})\} \right\},$$

and $\Lambda_{kg} = \{i : X_{ik} = g\}$.

Similarly, let κ_{kq}^1 and κ_{kq}^2 be the sets

$$\kappa_{kq}^1 = \{i : (d_{ik} = 1, \zeta_{iq} > 0) \cup (d_{ik} = 0, \zeta_{iq} < 0)\}$$

and

$$\kappa_{kq}^2 = \{i : (d_{ik} = 1, \zeta_{iq} < 0) \cup (d_{ik} = 0, \zeta_{iq} > 0)\}.$$

Then, the FCD of γ_{kq}, $k = 1, 2, \ldots, K$, $q = 1, 2, \ldots, Q_2$, given all the other parameters and $\mathbf{V}$, $\mathbf{D}$ is

$$\pi(\gamma_{kq} \mid \cdot) \propto \text{Uniform}(\gamma_{kq}^{(L)}, \gamma_{kq}^{(R)}) I(\gamma_{kq} > 0), \tag{15}$$

where

$$\gamma_{kq}^{(L)} = \max_{i\in\kappa_{kq}^1} \left\{ \frac{1}{\zeta_{iq}} \left[\log\left(\frac{V_{ik}}{1-V_{ik}}\right) + \delta_k - \sum_{s\neq q} \gamma_{ks}\zeta_{is} \right] \right\},$$

$$\gamma_{kq}^{(R)} = \min_{i\in\kappa_{kq}^2} \left\{ \frac{1}{\zeta_{iq}} \left[\log\left(\frac{V_{ik}}{1-V_{ik}}\right) + \delta_k - \sum_{s\neq q} \gamma_{ks}\zeta_{is} \right] \right\}.$$

Step 5. Sampling $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

With the assumption of noninformative prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto 1$, the corresponding posterior distribution can be written as

$$\boldsymbol{\Sigma}|\boldsymbol{\mu}, \boldsymbol{\eta} \sim \text{Inverse-Wishart}(\boldsymbol{\Delta}^{-1}, N), \ \boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{\eta} \sim N_W(\bar{\boldsymbol{\eta}}, \boldsymbol{\Sigma}/N), \tag{16}$$

where $\bar{\boldsymbol{\eta}} = \sum_{i=1}^{N} \boldsymbol{\eta}_i/N$ and $\boldsymbol{\Delta} = \sum_{i=1}^{N}(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}})^{\mathrm{T}}$.

So the FCDs of all the parameters can be obtained, the key advantage of our algorithm is that it is easy to sample from all FCDs. In particular, all of them are truncated distributions except for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The FCDs for the latent data are truncated uniform distributions, and the FCDs for the ability and item parameters are truncated priors. Note that the samplers of the observed models are all conditional on the missing data indicator matrix $\mathbf{D}$. This means that, for example, the sampling of $\boldsymbol{\alpha}_k$ in Step 4, $U_{ik}, \boldsymbol{\theta}_i$ are only used if $d_{ik} = 1$. With initial values $\boldsymbol{\eta}^{(0)}$, $\boldsymbol{\xi}^{(0)}$, $\boldsymbol{\mu}^{(0)}$, $\boldsymbol{\Sigma}^{(0)}$, the Gibbs sampler iteratively samples $\mathbf{U}, \mathbf{V}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$, from the distributions (9)–(16). The components are updated in the order given by Steps 1–5 above. A computer programme written in MATLAB language is available upon request.

4. A simulation study

A simulation study was undertaken to assess the effect of a missing-data process as described in Equations (5) and (6) on the estimates of item parameters. The simulation study consisted of two parts. The first part considers the dichotomous case, where the models for the observed response and missing data indicators are the Rasch model (RM). The second part considers the multidimensional and polytomous case, where we use a four dimensional GRM to model the observed response and two dimensional RM to model the missing data indicators.

In this study, data were simulated using the NONMAR model (6) and analysed using the MAR model (5) and NONMAR model (6), respectively. Priors for item parameters are taken as $\pi(\boldsymbol{\beta}_k, \delta_k) = 1$, $k = 1, 2, \ldots, K$. We assume that latent trait parameters $(\boldsymbol{\theta}_i, \boldsymbol{\zeta}_i) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, 2, \ldots, N$. As for the hyper-parameters μ and $\boldsymbol{\Sigma}$, conventionally, we work with $\boldsymbol{\mu} = \mathbf{0}$ for identification purpose [16] and $\boldsymbol{\Sigma} = (\sigma_{st})_{Q\times Q}$. Here, the correlation between the latent trait variables $\boldsymbol{\theta}_i$ and $\boldsymbol{\zeta}_i$, $\rho(\boldsymbol{\theta}, \boldsymbol{\zeta})$ is a key quantity, which serves to index the ignorability: the more the correlation differs from zero, the more ignorability is violated.

- For the dichotomous case in the simulation, to assess the bias imposed when ignorability is used unjustifiably, latent person parameters were drawn from a bivariate normal distribution. We take $\sigma_1^2 = \mathrm{Var}(\theta_i) = 1.0$, $\sigma_2^2 = \mathrm{Var}(\zeta_i) = 1.0$ and consider ten different values of $\rho(\theta_i, \zeta_i)$, i.e. $\rho = 0.0, 0.1, \ldots, 0.9$, respectively. For each case, latent trait values (θ_i, ζ_i) were drawn with sample sizes $N = 500$ and 1000, respectively. The test consisted of $K = 10$, 20, 30 dichotomously scored items. The observations x_{ik} and the missing-data indicators d_{ik} were both generated using the Rasch model, with item parameters β_k and δ_k, respectively. The true values were $\beta_k = \delta_k = 0$, for all k. The starting values were $\beta_k = \delta_k = 0.5$ for all items, and latent traits $\theta_i, \ \zeta_i, \ i = 1, 2, \ldots, N$ were all drawn from $N(0, 1)$.
- For the polytomous case in the simulation, latent person parameters were drawn from a six-variate normal distribution. The sample size was $N = 1000$. The variances of the latent variables was always equal to one. The correlation between the latent trait variables $\boldsymbol{\theta}_i$ and $\boldsymbol{\zeta}_i$, $\rho(\boldsymbol{\theta}; \boldsymbol{\zeta})$, varied as 0.0, 0.2, 0.4, 0.6, 0.8. Further, we fix the correlations between the components of the latent variables $\boldsymbol{\theta}$, i.e. $\rho(\theta_s, \theta_t)$, and the two dimensions of the missing-data process $\rho(\zeta_1, \zeta_2)$ at 0.5, where $s, t = 1, 2, 3, 4$ and $s \neq t$. The items were polytomously scored. We use a four dimensional GRM to model the observed response. The test consisted of $K = 10$ items and items with three response categories were used in the simulation. The true values were $\alpha_{kq_1} = 1$, $\delta_k = 0$

and $\beta_{k1} = -1$, $\beta_{k2} = 1$. The starting values were $\alpha_{kq_1} = 2$, $\beta_{k1} = -0.5$, $\beta_{k2} = \delta_k = 0.5$ for all items and latent traits θ_{iq_1}, ζ_{iq_2} were all drawn from $N(0, 1)$, where $k = 1, 2, \ldots, K$; $i = 1, 2, \ldots, N$; $q_1 = 1, 2, 3, 4$; $q_2 = 1, 2$.

Based on the above settings and the generated data set, our proposed procedure based on the Gibbs sampler is used to obtain the Bayesian estimates $\hat{\beta}_k$, from MAR model G_1, and estimates $\hat{\beta}'_k$ and $\hat{\delta}_k$ from NONMAR model G_2. We have used the Gelman–Rubin method [18,19] to test the convergence of our Bayesian method. We have considered the mean of the elements of the chain and implemented this method using 10 chains. Figure 1 presents the trace plots for the posterior mean of an arbitrarily selected item difficulty parameter β_1 (of MAR model (5)) under 10 starting points for 9 correlations $\rho = 0.1, \ldots, 0.9$ for the dichotomous case. It appears that the sequences are close to converging, since they are all approaching the same values. In addition, the plots of EPSR values (estimated potential scale reduction [19]) of some item parameters (of NONMAR model (6)) against iterations are displayed in Figure 2. This seems to confirm that the chains are getting close to converging. Our final values of EPSR at the last iteration of the chains are all around one. From the Gelman–Rubin method, to be conservative, we take a burn-in period of $n_0 = 4000$ and run for 20,000 iterations. And observations are collected after 4000 iterations in producing the simulation results for the 20 replications.

Then the values of item parameter estimates over replications, say $\hat{\vartheta}^{(r)}$, $r = 1, \ldots, R$ with $R = 20$, were compared with the values of the parameters used to generate the data based on the following three criteria: the mean squared error (MSE), mean absolute error (MAE) and bias. There is no index k since all item parameters were equal. The MSE for item parameters is

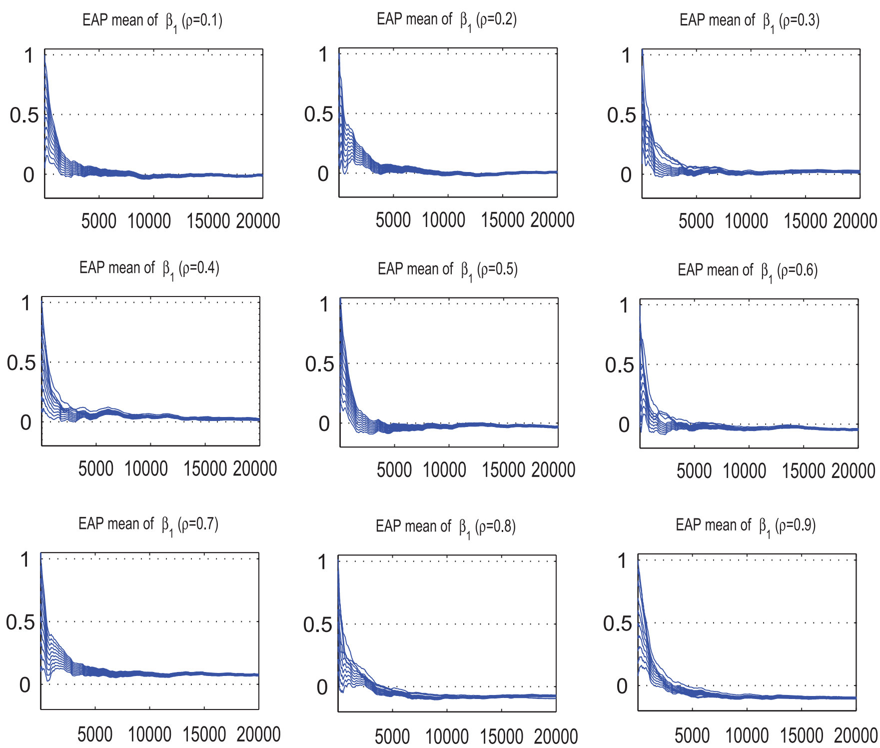

Figure 1. The parallel trace plots of EAP mean of β_1 under 10 different initial values for different ρ.

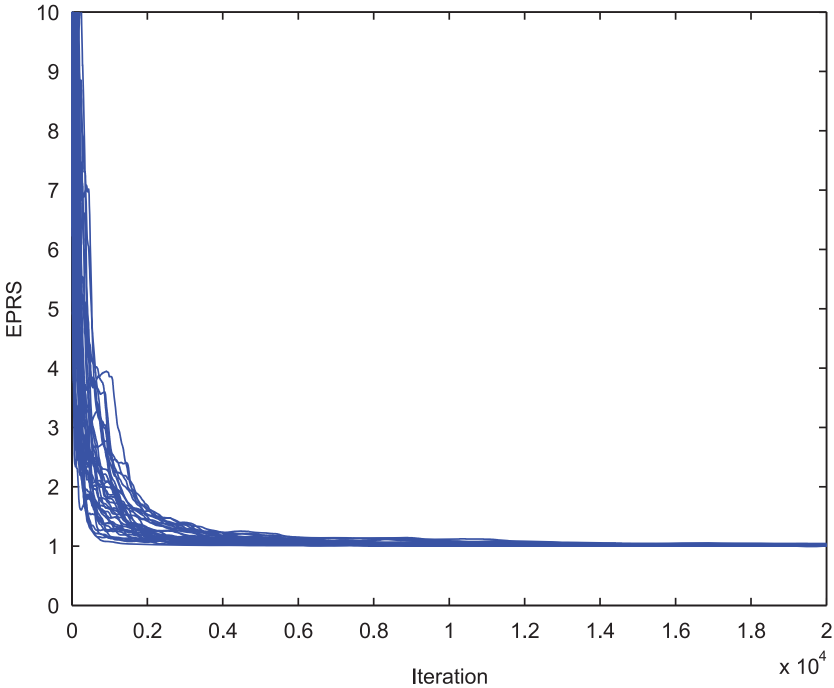

Figure 2. The EPSR values of some parameters against iteration numbers.

defined as

$$\mathrm{MSE}(\vartheta) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\vartheta}^{(r)} - \vartheta)^2, \tag{17}$$

where $r = 1, 2, \ldots, R$. The MAE for ϑ is defined as

$$\mathrm{MAE}(\vartheta) = \frac{1}{R}\sum_{r=1}^{R} \mid \hat{\vartheta}^{(r)} - \vartheta \mid . \tag{18}$$

The bias for ϑ is defined as

$$\mathrm{Bias}(\vartheta) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\vartheta}^{(r)} - \vartheta). \tag{19}$$

The MAE, MSE and bias results of the item parameter estimates for the two parts are summarized in Figures 3–5 and Table 1. In Figures 3–5, the notation ρ refers to the correlation of the latent variables of the observed data θ and the missing-data process ζ. The β refers to the item difficulty parameter for the MAR model, β' the item difficulty parameter for the NONMAR model and δ the item difficulty for the missing process. It is apparent that, for $\rho = 0.0$, the ignorability holds, so the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$ are equal and served as the baseline. Further, the results showed that when the correlation increases, the MAE, MSE and bias of the estimates under the assumption of MAR increase considerably. For instance, if we look specifically at Figure 3, the second subplot (labelled 10 items 1000 respondents), where ρ increases from 0.0 to 0.9 with an increment of 0.1, the MAE of β for the MAR model had values of 0.078, 0.104, 0.109, 0.126, 0.126, 0.142, 0.178, 0.190, 0.225 and 0.233. These values of the MAE were inflated as expected since the missing-data process was excluded. It was also true for the MSE and bias of β for the MAR model. The MAE,

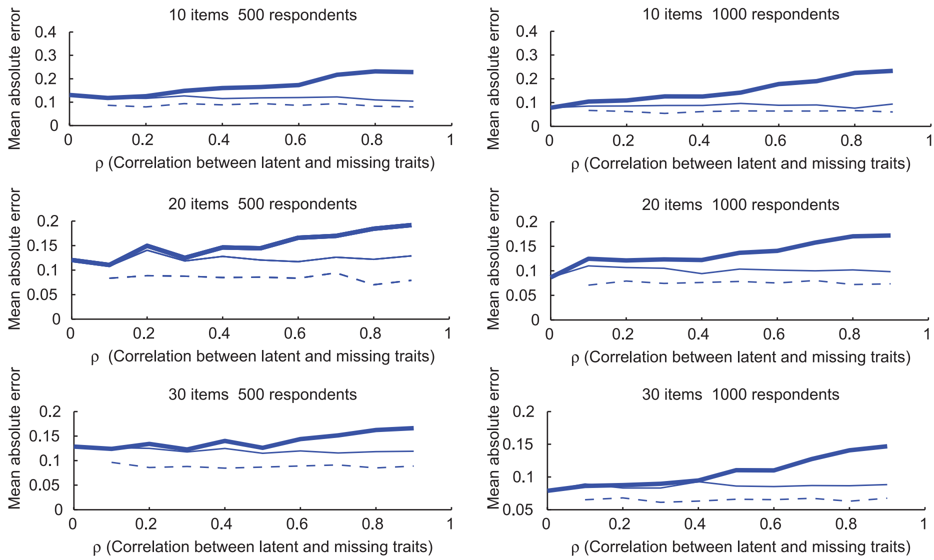

Figure 3. Comparing the values of MAE(β') (fine lines) and MAE(β) (thick lines) and the dashed line denotes the MAE(δ).

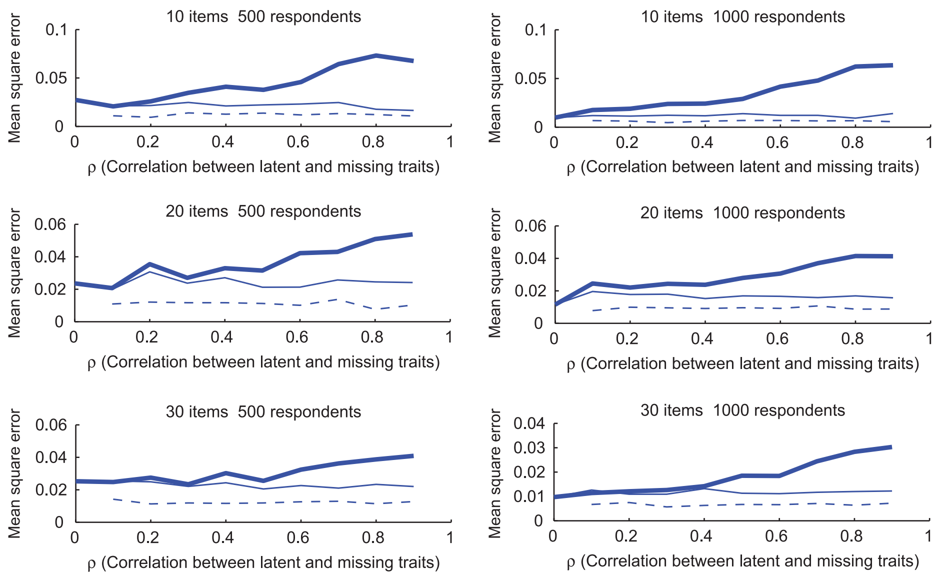

Figure 4. Comparing the values of MSE(β') (fine lines) and MSE(β) (thick lines) and the dashed line denotes the MSE(δ).

MSE and bias values of β' for the NONMAR model only showed random fluctuation. Looking again in Figure 3, the second subplot, the MAE of β' for the NONMAR model had values 0.078, 0.085, 0.086, 0.088, 0.088, 0.096, 0.089, 0.090, 0.077 and 0.094. Similarly, the MAE of δ, the item difficulty for the missing-data process had values, 0.067, 0.063, 0.055, 0.062, 0.065, 0.064, 0.065, 0.066 and 0.060, which remain stable apart from random fluctuations.

In Table 1, the MAE, MSE and bias of the item parameters issued from our Gibbs-sampling MCMC procedure were given for five correlations $\rho = 0.0, 0.2, 0.4, 0.6, 0.8$. Consider the column pertaining to $\rho = 0$, the baseline of ignorable missing data. Consider the next columns

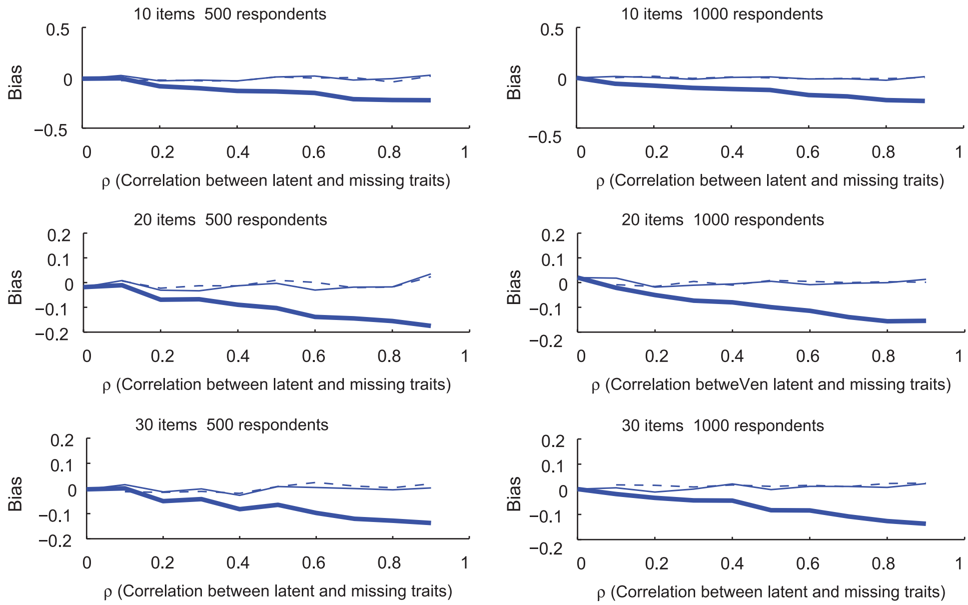

Figure 5. Comparing the values of bias(β') (fine lines) and bias(β) (thick lines) and the dashed line denotes the bias(δ).

Table 1. The MAE, MSE and bias of item parameter estimates under the MAR and NONMAR model (multidimensional and polytomous case); estimation model (observed data: 4D-GRM; missing data: 2D-RM); $N = 1000$, $K = 10$; $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$, $\beta_1 = -1$, $\beta_2 = 1$, $\delta = 0$; $\rho = \rho(\theta_{q1}, \zeta_{q2})$, $q_1 = 1, 2, 3, 4$, $q_2 = 1, 2$.

	$\rho = 0.0$	$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
	MAR	MAR	NONMAR	MAR	NONMAR	MAR	NONMAR	MAR	NONMAR
MAE									
α_1	0.11	0.12	0.10	0.13	0.10	0.14	0.09	0.17	0.09
α_2	0.12	0.11	0.11	0.13	0.11	0.15	0.09	0.18	0.10
α_3	0.12	0.11	0.11	0.13	0.09	0.14	0.10	0.17	0.11
α_4	0.12	0.12	0.12	0.12	0.09	0.14	0.10	0.16	0.11
β_1	0.11	0.13	0.11	0.13	0.10	0.16	0.11	0.17	0.09
β_2	0.10	0.10	0.09	0.12	0.09	0.13	0.10	0.16	0.10
δ			0.08		0.08		0.08		0.07
MSE									
α_1	0.02	0.02	0.02	0.02	0.01	0.03	0.01	0.04	0.02
α_2	0.02	0.02	0.02	0.02	0.02	0.03	0.01	0.04	0.02
α_3	0.02	0.02	0.02	0.03	0.01	0.03	0.02	0.04	0.02
α_4	0.02	0.02	0.02	0.02	0.01	0.03	0.02	0.04	0.02
β_1	0.02	0.02	0.02	0.03	0.02	0.04	0.02	0.04	0.01
β_2	0.02	0.02	0.01	0.02	0.02	0.03	0.02	0.04	0.02
δ			0.01		0.01		0.01		0.01
Bias									
α_1	−0.03	−0.08	−0.00	−0.09	0.01	−0.13	−0.01	−0.15	0.01
α_2	−0.03	−0.07	−0.01	−0.09	0.01	−0.14	0.01	−0.17	0.02
α_3	0.03	−0.08	−0.03	−0.07	−0.00	−0.12	−0.00	−0.16	−0.00
α_4	0.02	−0.08	−0.05	−0.06	−0.00	−0.11	−0.02	−0.14	0.00
β_1	0.00	−0.06	−0.00	−0.10	0.00	−0.15	−0.01	−0.16	−0.00
β_2	0.01	−0.03	0.01	−0.07	0.01	−0.11	−0.02	−0.15	−0.00
δ			−0.02		−0.01		−0.00		0.00

labelled $\rho = 0.2$ to $\rho = 0.8$, where the missingness was no longer ignorable, and the data was analysed using the MAR model and NONMAR model, respectively. The rows denoted by $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \delta$ are the estimated values of the MAE, MSE and bias of the item parameters for the observed data $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ (discrimination), β_1, β_2 (difficulty) and the missing-data process δ (difficulty). From Table 1 we can see, invoking the missing-data process leads to a reduction of estimation errors. And for the multidimensional and polytomous case, the relationship between ρ and the estimation errors are similar to those presented in the first part (Figures 3–5), it can be seen that increasing the correlation of the latent variables $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ is increasing the violation of ignorability.

5. Discussion and conclusion

Analysis of the missing data has received tremendous attention in statistics. This important topic motivates numerous important contributions in computational statistics, for example, the EM [20] and its related algorithms [21] for the ML approach and various MCMC methods for the Bayesian approach. In the framework of IRT, some recent attention has focussed on analysing missing data that are ignorable and MAR [10,16]. For the nonignorable missing data, the MML procedure was used [6]. However, a disadvantage of the MML procedure is, the higher the dimension of the latent variables, the more likely computational problems will arise during the estimation process. For example, Diggle *et al.* [7] indicated that numerical maximum likelihood methods are only feasible for models with a limited number of latent variables (five or fewer). To our knowledge, no Bayesian method has been used for the nonignorable missing data in the framework of IRT. In this article, having benefited much from the work of Holman & Glas [6], we developed a data augmentation method based on Gibbs sampling for estimating the parameters in multidimensional GRM model with nonignorable missing-data mechanism. Our simulation results show that ignoring the missing-data process results in considerable bias in the estimates of the item parameters. This bias increases as a function of the correlation between the proficiency to be measured and latent variable governing the missing-data process.

Further, it was shown that this bias can be reduced using the NONMAR model (6). This approach can be generalized by the inclusion of covariates in the missing data model and can also be applied to longitudinal data, testlet and multilevel response model.

Acknowledgements

This work is supported by NSFC (10431010, 10828102, and 10871037) and Training Fund of NENU'S Scientific Innovation Project (NENU-STC07002).

References

[1] D.B. Rubin, *Inference and missing data*, Biometrika 63 (1976), pp. 581–592.
[2] I. Moustaki and C. O'Muircheartaigh, *A one dimensional latent trait model to infer attitude from nonresponse for nominal data*, Statistica 60(2) (2000), pp. 259–276.
[3] I. Moustaki and M. Knott, *Weighting for item non-response in attitude scales by using latent variable models with covariates*, J. R. Statist. Soc. Ser. A 163 (2000), pp. 445–459.
[4] C.A. Bernaards and K. Sijtsma, *Factor analysis of multidimensional polytomous item response data suffering from ignorable item nonresponse*, Multivariate Behav. Res. 34(3) (1999), pp. 277–313.
[5] C.A. Bernaards and K. Sijtsma, *Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable*, Multivariate Behav. Res. 35(3) (2000), pp. 321–364.
[6] R. Holman and C.A.W. Glas, *Modelling non-ignorable missing-data mechanism with item response theory models*, British J. Math. Statist. Psych. 58 (2005), pp. 1–18.
[7] P.J. Diggle, P. Heagerty, K.Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data*, Oxford, Clarendon, 2002.

[8] J.H. Albert, *Bayesian estimation of normal ogive item response functions using Gibbs sampling*, J. Educ. Statist. 17 (1992), pp. 251–269.
[9] M. Ghosh, A. Ghosh, M.H. Chen, and A. Agresti, *Noninformative priors for one-parameter item response models*, J. Statist. Plann. Inference 88 (2000), pp. 99–115.
[10] R.J. Patz and B.W. Junker, *Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses*, J. Educ. Behav. Stat. 24 (1999), pp. 146–178.
[11] S.K. Sahu, *Bayesian estimation and model choice in item response models*, J. Stat. Comput. Simul. 72 (2002), pp. 217–232.
[12] F. Samejima, *Estimation of latent ability using a response pattern of graded scores*, Psychometrika Monogr. Suppl. 34(4) (1969), pp. 100–114.
[13] M.D. Reckase, *A linear logistic multidimensional model for dichotomous item response data*, in *Handbook of Modern Item Response Theory*, W.J. van der Linden and R.K. Hambleton, eds., Springer, New York, 1997, pp. 271–286.
[14] T.A. Ackerman, *Developments in multidimensional item response theory*, Appl. Psychol. Meas. 20 (1996), pp. 309–310.
[15] N.D. Verhelst and C.A.W. Glas, *The one-parameter logistic model*, in *Rasch Models: Foundations, Recent Developments, and Applications*, G.H. Fischer and I.W. Molenaar, eds., Springer-Verlag, New York, 1995, pp. 215–237.
[16] A.A. Béguin and C.A.W. Glas, *MCMC estimation and some model-fit analysis of multidimensional IRT models*, Psychometrika 66 (2001), pp. 541–562.
[17] A.E. Gelfand and A.F.M. Smith, *Sampling based approaches to calculating marginal densities*. J. Amer. Statist. Assoc. 85 (1990), pp. 398–409.
[18] A. Gelman and D.B. Rubin, *Inference from iterative simulation using multiple sequences* (*with discussion*), Statist. Sci. 7 (1992), pp. 457–511.
[19] A. Gelman, *Inference and monitoring convergence*, in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.T. Spiegelhalter, eds., Chapman and Hall, London, 1996, pp. 131–143.
[20] A.P. Dempster, N.M. Laird, and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm* (*with discussion*), J. R. Statist. Soc. Ser. B 39 (1977), pp. 1–38.
[21] X.L. Meng and D. van Dyk, *The EM algorithm–an old folk-song sung to a fast new tune*, J. R. Statist. Soc. Ser. B 59 (1997), pp. 511–540.
[22] R.P. Gupta, *Multivariate Statistical Analysis.* North-Holland, New York, 1980.

Appendix. Details for computing the conditional mean and variance

To compute the conditional mean and variance of η_{iq} in Equation (13), we need to take a permutation on $\boldsymbol{\eta}_i$. Let

$$\boldsymbol{\eta}_i^*(w) = \begin{cases} \boldsymbol{\eta}_i, & \text{if } w = 1, \\ \mathbf{P}_{w1}\boldsymbol{\eta}_i, & \text{if } w \neq 1, \end{cases} \quad i = 1, 2, \dots, n, \quad w = 1, 2, \dots, W,$$

where $\mathbf{P}_{w1} = (p_{st})_{W\times W}$ is a permutation matrix with (s, t) element

$$p_{st} = \begin{cases} 1, & \text{if } s = t, \ \ s \neq w, 1; \ \text{ or } s = w, \ \ t = 1; \ \text{ or } s = 1, \ \ t = w, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\boldsymbol{\eta}_i \sim N_W(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ we have $\boldsymbol{\eta}^*(w) \sim N_W(\boldsymbol{\mu}^*(w), \boldsymbol{\Sigma}^*(w))$, where

$$\boldsymbol{\mu}^*(w) = \mathbf{P}_{w1}\boldsymbol{\mu}, \quad \boldsymbol{\Sigma}^*(w) = \mathbf{P}_{w1}\boldsymbol{\Sigma}\mathbf{P}_{w1}^{\mathrm{T}}.$$

In order to get the conditional distribution of η_{iq}, partitioning $\boldsymbol{\eta}^*(w)$ as

$$\boldsymbol{\eta}_i^*(w) = \begin{pmatrix} \lambda_w^{(1)} \\ \boldsymbol{\lambda}_w^{(2)} \end{pmatrix}, \quad \text{where } \lambda_w^{(1)} = \eta_{iw}, \ \ \boldsymbol{\lambda}_w^{(2)} = \boldsymbol{\eta}_i(w); \ \text{ so we can obtain}$$

$$\boldsymbol{\mu}^*(w) = \begin{pmatrix} \nu_w^{(1)} \\ \boldsymbol{\nu}_w^{(2)} \end{pmatrix}, \quad \boldsymbol{\Sigma}^*(w) = \begin{pmatrix} \Gamma_{11}(w) & \boldsymbol{\Gamma}_{12}(w) \\ \boldsymbol{\Gamma}_{21}(w) & \boldsymbol{\Gamma}_{22}(w) \end{pmatrix},$$

where

$$\nu_w^{(1)} = E(\lambda_w^{(1)}) = \mu_{iw}, \quad \boldsymbol{\nu}_w^{(2)} = E(\boldsymbol{\lambda}_w^{(2)}) = \boldsymbol{\mu}_i(w),$$

($\boldsymbol{\mu}_i(w)$ is defined as the mean of $\boldsymbol{\eta}_i(w)$, that is, $\boldsymbol{\mu}_i(w) = E(\boldsymbol{\eta}_i(w))$.

$$\Gamma_{11}(w) = \mathrm{Cov}(\eta_{iw}, \eta_{iw}), \quad \boldsymbol{\Gamma}_{12}(w) = \mathrm{Cov}(\eta_{iw}, \boldsymbol{\eta}_i(w)),$$
$$\boldsymbol{\Gamma}_{21}(w) = \mathrm{Cov}(\boldsymbol{\eta}_i(w), \eta_{iw}), \quad \boldsymbol{\Gamma}_{22}(w) = \mathrm{Cov}(\boldsymbol{\eta}_i(w), \boldsymbol{\eta}_i(w)).$$

So the conditional distribution of $\lambda_w^{(1)}$ given $\boldsymbol{\lambda}_w^{(2)}$ [22] is as follows:

$$\lambda_w^{(1)} | \boldsymbol{\lambda}_w^{(2)} \sim N(\nu_w, \tau_w^2),$$

where

$$\nu_w = \nu_w^{(1)} + \boldsymbol{\Gamma}_{12}(w)\boldsymbol{\Gamma}_{22}^{-1}(w)(\boldsymbol{\lambda}_w^{(2)} - \boldsymbol{\nu}_w^{(2)}), \quad \tau_w^2 = \Gamma_{11}(w) - \boldsymbol{\Gamma}_{12}(w)\boldsymbol{\Gamma}_{22}^{-1}(w)\boldsymbol{\Gamma}_{21}(w),$$

that is

$$\eta_{iw} | \boldsymbol{\eta}_i(w) \sim N(\nu_w, \tau_w^2).$$