

2012

A Gibbs Sampler for the Multidimensional Item Response Model

Yanyan Sheng

Southern Illinois University Carbondale, ysheng@siu.edu

Todd C. Headrick

Southern Illinois University Carbondale, headrick@siu.edu

Follow this and additional works at: http://opensiuc.lib.siu.edu/epse_pubs

Published in *ISRN Applied Mathematics*, Vol. 2012 Article ID 269385 at [doi:10.5402/2012/269385](https://doi.org/10.5402/2012/269385)

Recommended Citation

Sheng, Yanyan and Headrick, Todd C. "A Gibbs Sampler for the Multidimensional Item Response Model." (Jan 2012).

This Article is brought to you for free and open access by the Educational Psychology and Special Education at OpenSIUC. It has been accepted for inclusion in Publications by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

Research Article

A Gibbs Sampler for the Multidimensional Item Response Model

Yanyan Sheng and Todd C. Headrick

*Section on Statistics and Measurement, Department of EPSE, Southern Illinois University Carbondale,
Wham 223, MailCode 4618, Carbondale, IL 62901-4618, USA*

Correspondence should be addressed to Yanyan Sheng, ysheng@siu.edu

Received 2 March 2012; Accepted 26 March 2012

Academic Editors: S. He and X. Xue

Copyright © 2012 Y. Sheng and T. C. Headrick. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Current procedures for estimating compensatory multidimensional item response theory (MIRT) models using Markov chain Monte Carlo (MCMC) techniques are inadequate in that they do not directly model the interrelationship between latent traits. This limits the implementation of the model in various applications and further prevents the development of other types of IRT models that offer advantages not realized in existing models. In view of this, an MCMC algorithm is proposed for MIRT models so that the actual latent structure is directly modeled. It is demonstrated that the algorithm performs well in modeling parameters as well as intertrait correlations and that the MIRT model can be used to explore the relative importance of a latent trait in answering each test item.

1. Introduction

Item response theory (IRT) is a popular approach used for describing probabilistic relationships between correct responses on a set of test items and continuous latent traits (see [1–4]). IRT models have also been used in other areas of applied mathematics and statistical research. Some examples include US Supreme Court decision-making processes [5], alcohol disorder analysis [6–9], nicotine dependency [10–12], multiple-recapture population estimation [13], psychiatric epidemiology [14–16], longitudinal data analysis [17, 18], latent regression models [19, 20], and missing data analysis [21].

IRT has the advantage of allowing the inference of what the items and persons have on the responses to be modeled by distinct sets of parameters. As a result, a primary concern associated with IRT research has been on parameter estimation, which offers the basis for the theoretical advantages of IRT. Specifically, of concern are the statistical complexities that can often arise when item and person parameters are simultaneously estimated (see [1, 22–24]). More recent attention has focused on fully Bayesian estimation where Markov chain Monte

Carlo (MCMC) simulation techniques are used (e.g., [25, 26]). Over the past decade, MCMC has been implemented in the context of IRT models where one latent trait is assumed (e.g., [3, 27–29]) as well as to models where multiple traits are considered (e.g., [30–36]), for a thorough review on the historical and current developments of MCMC in terms of IRT, see [37].

The compensatory multidimensional IRT (MIRT; [38]) model assumes that each item measures multiple latent traits. It differs from some other dichotomous models insofar as it has an additional source of model indeterminacy that creates difficulties when using MCMC. Some techniques have been developed to approach this problem by imposing a special structure that constrains the item slope parameters [30, 36, 39]. However, these approaches do not directly model the actual interrelation between the distinct latent traits and, thus, are limited in certain applications. In view of the above, the present aim is to derive an efficient MCMC algorithm via Gibbs sampling [40] that (a) obviates the additional source of model indeterminacy associated with the MIRT model and (b) directly models the underlying latent trait structure. The MIRT model considered herein is presented in normal ogive form as more complicated MCMC procedures would have to be adopted for the logistic form (e.g., [3, 28, 35, 36]). Further, given that parametric probability functions of correct responses are usually modeled by a normal ogive or a logistic function and noting that the logistic and normal ogive forms of the IRT models are essentially indistinguishable in terms of model fit or parameter estimates (given proper scaling, see [41]), MCMC procedures for logistic models are not considered.

The remainder of this paper is organized as follows. In Section 2, the two-parameter normal ogive (2PNO) MIRT model is outlined. In Section 3, the Gibbs sampler is derived, and the prior specifications for the model parameters are described. Section 4 gives examples of implementing the Gibbs sampling algorithm in the context of simulated and real data to demonstrate the proposed methodology.

2. Preliminaries

The MIRT model is introduced by considering a test that consists of k dichotomous items with each measuring m latent traits. Let $\mathbf{y} = [\mathbf{y}_{ij}]_{n \times k}$ denote a matrix of n responses to the k items where $y_{ij} = 1$ ($y_{ij} = 0$) if the i th person answers the j th item correctly (incorrectly) for $i = 1, \dots, n$ and $j = 1, \dots, k$.

Definition 2.1. The probability of the i th person obtaining a correct response on the j th item is defined for the 2PNO MIRT model as

$$P(y_{ij} = 1 \mid \boldsymbol{\theta}_i, \boldsymbol{\alpha}_j, \beta_j) = \Phi(\boldsymbol{\alpha}_j \boldsymbol{\theta}_i - \beta_j) = \Phi\left(\sum_{l=1}^m \alpha_{lj} \theta_{li} - \beta_j\right) = \Phi(\eta_{ij}). \quad (2.1)$$

The vector $\boldsymbol{\theta}_i = (\theta_{1i}, \dots, \theta_{li}, \dots, \theta_{mi})'$ denotes latent trait parameters associated with the i th person, and the vector $\boldsymbol{\alpha}_j = (\alpha_{1j}, \dots, \alpha_{lj}, \dots, \alpha_{mj})$ denotes nonnegative slope parameters where larger values of α_{lj} have more influence on determining a success on the j th item. The intercept parameter β_j denotes the location in the latent space where the j th item is maximally informative, and Φ denotes the unit normal cdf. The model in (2.1) is also referred to as a compensatory MIRT model [38] because a low level of θ_{li} in one dimension can be compensated by a high level of θ_{li} in another dimension.

Remark 2.2. If the vector of slope parameters in (2.1) is such that $\alpha_j = (0, \dots, 0, \alpha_{lj}, 0, \dots, 0)$, then the MIRT model reduces to the 2PNO multi-unidimensional model as

$$P(y_{lij} = 1 \mid \theta_{li}, \alpha_{lj}, \beta_{lj}) = \Phi(\alpha_{lj}\theta_{li} - \beta_{lj}), \quad (2.2)$$

where the test involves multiple parameters of θ_{li} and where each item measures one of these latent variables (see [31, 32]). The difference between (2.1) and (2.2) is analogous to the distinction made between factor analysis and that with a rotation to achieve a simple structure [42]. As such, (2.2) can be viewed as a special case of (2.1) where each item measures only one of the several latent traits. Further, the two models differ in that (2.1) is exploratory whereas (2.2) is confirmatory in nature.

The unidimensional IRT model, which has a systematic component form of $\alpha_j\theta_i - \beta_j$, has a well-known identification problem in terms of location and scale invariance (e.g., [43]). Common practices of resolving this problem are to impose some constraint on the item parameters, that is, $\prod \alpha_j = 1$ and $\sum \beta_j = 0$, or select some specific values for the location and scale parameters for the prior normal distribution of θ_i , for example, $\theta_i \sim N(0, 1)$ (see, e.g., [3, 27–29, 43]). Further, Bafumi et al. [5] proposed using a parameter transformation to approach the identification problem in the context of unidimensional IRT models. More specifically, the model parameters are transformed using a normalization procedure after estimation is completed. Bafumi et al. [5] noted that this transformation procedure obviates the problem of elusive convergence that results from highly correlated samples.

In terms of the multi-unidimensional IRT model in (2.2), Lee [31] extended Tsutakawa's [43] approach by adopting a constrained covariance matrix for the latent traits and modeling the constrained covariance matrix indirectly. Lee's [31] method not only solves the model indeterminacy problem, but also appropriately estimates the interrelationship between multiple latent traits (see also [32, 44]).

The more general MIRT model, as defined in (2.1), involves a new source of model indeterminacy called rotational invariance and is statistically more complicated than the unidimensional or multi-unidimensional models. As such, a Gibbs sampler is subsequently derived based on the ideas suggested in [5, 31] to address the general MIRT model identification problems and to model the latent structure directly.

It is noted that in an effort to develop computer software, Sheng [45] has shown that the approaches based on [5, 31] are useful for the 2PNO additive MIRT model, whose systematic component for modeling y_{lij} takes the form $\alpha_{0lj}\theta_{0i} + \alpha_{lj}\theta_{li} - \beta_{lj}$. The model assumes that each item measures two latent traits: θ_{0i} , a common latent trait that all items measure, and θ_{li} , a latent trait that is specific for items in the l th subtest. The difference between the model in [45] and the general MIRT model presented herein is comparable to that between a bifactor model (see [46]) and a general factor analysis model. The two models assume different latent structures, and hence the approaches for resolving their model indeterminacies are not the same.

3. The Gibbs Sampler

The derivation of the Gibbs sampler associated with the MIRT model defined in (2.1) begins by considering a multivariate distribution for θ_i and a linear transformation on it, which will be based on the following definitions.

Definition 3.1. Let $\theta_i \sim N_m(\mathbf{0}, \mathbf{P})$, where $\mathbf{P}$ is a constrained covariance matrix or a correlation matrix, with 1's on the diagonal and with correlations ρ_{st} (between θ_{si} and θ_{ti}) on the off-diagonal.

Definition 3.2. Let $\theta_i^* \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = [\sigma_{ll'}]_{m \times m}$ and $\boldsymbol{\Sigma} = D\mathbf{P}D$, where D is an $m \times m$ diagonal matrix. Note that this variance-correlation decomposition of $\boldsymbol{\Sigma}$ [47] makes the interpretation easier [48] and is essential for modeling the correlation matrix indirectly while solving the model indeterminacy in the context of the MIRT model.

From Definitions 3.1 and 3.2, it can be shown that

$$\theta_i^* - \boldsymbol{\mu} = D\theta_i, \quad (3.1)$$

where $\mathbf{P}$ can be transformed from $\boldsymbol{\Sigma}$ using

$$\rho_{st} = \frac{\sigma_{st}}{\sqrt{\sigma_{ss}\sigma_{tt}}} \quad (3.2)$$

for $s \neq t$. To obviate the identification problem associated with the unconstrained parameters, let θ_i^* be related with the item parameters (α_j^* and β_j^*) so that the likelihoods are preserved given

$$\alpha_j^* \theta_i^* - \beta_j^* = \alpha_j \theta_i - \beta_j, \quad (3.3)$$

where the item parameters (α_j^* and β_j^*) will have to be constrained such that $\prod_j \alpha_{lj}^* = 1$ and $\sum_j \beta_j^* = 0$. This leads us to the following proposition.

Proposition 3.3. *If α_j^* are constrained such that $\prod_j \alpha_{lj}^* = 1$, then*

$$D = \text{diag} \left(\left(\prod_j \alpha_{1j} \right)^{1/k}, \dots, \left(\prod_j \alpha_{mj} \right)^{1/k} \right). \quad (3.4)$$

Proof. It follows from (3.1) that $\theta_i^* = D\theta_i + \boldsymbol{\mu}$, and thus, substituting $D\theta_i + \boldsymbol{\mu}$ into (3.3) gives

$$\alpha_j^* (D\theta_i + \boldsymbol{\mu}) - \beta_j^* = \alpha_j \theta_i - \beta_j. \quad (3.5)$$

Using (3.5), we can subsequently derive

$$\alpha_j^* D = \alpha_j. \quad (3.6)$$

Setting $D = \text{diag}(d_1, \dots, d_m)$ in (3.6) and subsequently multiplying the left-hand side yields

$$\alpha_{lj}^* d_l = \alpha_{lj}, \quad (3.7)$$

which leads to

$$\prod_j \alpha_{lj}^* d_l = \prod_j \alpha_{lj}, \quad (3.8)$$

for $l = 1, \dots, m$. Hence, given the constraint that $\prod_j \alpha_{lj}^* = 1$, each nonzero element in D is $d_l = (\prod_j \alpha_{lj})^{1/k}$. $\square$

To implement Gibbs sampling for the MIRT model in (2.1), a latent variable Z is introduced such that $Z_{ij} \sim N(\eta_{ij}, 1)$ (see, e.g., [27, 49]). Further, from Definition 3.1, we assume that $\theta_i \sim N_m(\mathbf{0}, \mathbf{P})$ to ensure unique scaling for θ , which precludes the identification problem associated with such models (see [45]). Furthermore, for the unconstrained covariance matrix Σ , we assume that $p(\Sigma) = |\Sigma|^{-(m+2)/2}$. Thus, if $\xi_j = (\alpha_j, \beta_j)'$ with assumed prior distributions, then the joint posterior distribution of (θ, ξ, Z, Σ) is

$$p(\theta, \xi, Z, \Sigma) \propto f(\mathbf{y} | Z) p(Z | \theta, \xi) p(\xi) p(\theta | \mathbf{P}) p(\Sigma), \quad (3.9)$$

where $f(\mathbf{y} | Z) = \prod_{i=1}^n \prod_{j=1}^k p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}}$ is the likelihood function, with p_{ij} being the model probability function as defined in (2.1).

The proposed Gibbs sampler involves the following five steps:

- (1) sampling of the augmented parameters from

$$Z_{ij} | \bullet \sim \begin{cases} N_{(0,\infty)}(\eta_{ij}, 1) & \text{if } y_{ij} = 1, \\ N_{(-\infty,0)}(\eta_{ij}, 1) & \text{if } y_{ij} = 0, \end{cases} \quad (3.10)$$

- (2) sampling of the latent variable (person) parameters θ_i from

$$\theta_i | \bullet \sim N_m\left((\mathbf{A}'\mathbf{A} + \mathbf{P})^{-1} \mathbf{A}'\mathbf{B}, (\mathbf{A}'\mathbf{A} + \mathbf{P})^{-1}\right), \quad (3.11)$$

$$\text{where } \mathbf{A} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}_{k \times m} \text{ and } \mathbf{B} = \begin{bmatrix} Z_{i1} + \beta_1 \\ \vdots \\ Z_{ik} + \beta_k \end{bmatrix}_{k \times 1},$$

- (3) sampling of the item parameters ξ_j from

$$\xi_j | \bullet \sim N_m\left((\mathbf{x}'\mathbf{x} + \mathbf{I})^{-1} \mathbf{x}'\mathbf{Z}_j, (\mathbf{x}'\mathbf{x} + \mathbf{I})^{-1}\right) I(\alpha_{lj} > 0), \quad (3.12)$$

where $\mathbf{x} = [\theta, -1]$, assuming uniform priors $\alpha_{lj} > 0$ and $p(\beta_j) \propto 1$, or from

$$\xi_j | \bullet \sim N_m\left((\mathbf{x}'\mathbf{x} + \Sigma_\xi^{-1})^{-1} (\mathbf{x}'\mathbf{Z}_j + \Sigma_\xi^{-1} \boldsymbol{\mu}_\xi), (\mathbf{x}'\mathbf{x} + \Sigma_\xi^{-1})^{-1}\right) I(\alpha_{lj} > 0), \quad (3.13)$$

where $\boldsymbol{\mu}_\xi = (\mu_{\alpha_1}, \dots, \mu_{\alpha_m}, \mu_\beta)'$ and $\Sigma_\xi = \text{diag}(\sigma_{\alpha_1}^2, \dots, \sigma_{\alpha_m}^2, \sigma_\beta^2)$, assuming conjugate normal priors $\alpha_{lj} \sim N_{(0,\infty)}(\mu_{\alpha_l}, \sigma_{\alpha_l}^2)$, $\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$,

(4) sampling of the unconstrained covariance matrix Σ from

$$\Sigma \mid \bullet \sim W^{-1}(\mathbf{S}^{-1}, n), \quad (3.14)$$

where W^{-1} is an inverse Wishart distribution, $\mathbf{S} = \sum_{i=1}^n (D\theta_i)(D\theta_i)'$, and where D is derived from (3.4),

(5) a transformation from Σ to $\mathbf{P}$.

In view of the additional model indeterminacy that results from the additive nature of θ_{lj} , the parameters are further normalized after each Markov transition step is completed [5, 45]. More specifically, θ_{li} , α_{lj} , and β_j are transformed (t) to the following normalized parameters: $\theta_{li}^t = (\theta_{li} - \bar{\theta}_l) / s_{\theta_l}$, $\alpha_{lj}^t = \alpha_{lj} s_{\theta_l}$ and $\beta_j^t = \beta_j - \sum_l \alpha_{lj} \bar{\theta}_l$, where $\bar{\theta}_l$ and s_{θ_l} represent the mean and standard deviation of θ_{lj} . This rescaling preserves the likelihood because $\sum_l \alpha_{lj}^t \theta_{li}^t - \beta_j^t = \sum_l \alpha_{lj} \theta_{li} - \beta_j$, while allowing the computation to proceed more efficiently [50]. Further, the transformation also assists in terms of speeding up the convergence of the Markov chains by reducing the posterior correlation in the posterior probability densities [51].

Thus, with initial starting values of $\theta^{(0)}$, $\xi^{(0)}$, and $\mathbf{P}^{(0)}$, the observations (i.e., $\mathbf{Z}^{(\ell)}$, $\theta^{(\ell)}$, $\xi^{(\ell)}$, $\Sigma^{(\ell)}$, and $\mathbf{P}^{(\ell)}$) can be drawn or transformed iteratively from (3.10), (3.11), (3.12), (3.14), and (3.2) (or (3.13) in lieu of (3.12)), respectively. This iterative process continues for a sufficient number of samples after the posterior distributions reach stationarity (i.e., a phase commonly referred to as burn-in). The posterior means of all the samples collected after the burn-in stage are considered to be estimates of the model parameters (θ , ξ) and the hyperparameter ($\mathbf{P}$).

4. Numerical Examples

To demonstrate the methodology presented above, the proposed Gibbs sampler was implemented using both simulated and real data. In terms of simulated data, tests that measure two latent traits were considered. In particular, three 1000×18 (i.e., $m = 2$, $n = 1000$, and $k = 18$) dichotomous data matrices were simulated from the 2PNO MIRT model where the population correlation between the two latent traits was set to $\rho_{\theta_{1i}, \theta_{2i}} = 0.2, 0.4, 0.6$, respectively. The item parameters were generated randomly from uniform distributions so that $\alpha_{lj} \sim U(0, 2)$, $\beta_j \sim U(-2, 2)$. Gibbs sampling was subsequently implemented to recover the model parameters assuming informative normal (i.e., $\mu_{\alpha_1} = \mu_{\alpha_2} = \mu_{\beta} = 0$ and $\sigma_{\alpha_1}^2 = \sigma_{\alpha_2}^2 = \sigma_{\beta}^2 = 1$) or uniform priors for ξ_j . Convergence was evaluated using the Gelman and Rubin [52] R statistic for each item parameter. While the usual practice is to use multiple Markov chains from different starting points, a single chain can also be divided into subchains so that convergence is assessed by comparing the between and within subchain variances (see [53]). In view of the fact that a single chain is more economical in the number of iterations needed, the latter approach was adopted. The posterior estimates of item parameters ($\alpha_1, \alpha_2, \beta$), the intertrait correlation hyperparameter, and the associated Gelman-Rubin R statistics were obtained and are listed in Tables 1, 2, and 3 (note that $\rho_{\theta_{1i}, \theta_{2i}}$ is denoted as ρ_{12} in these tables).

The Gelman-Rubin R statistic provides a numerical measure for assessing convergence for each item parameter. With values close to 1, it is determined that in the implementation of the Gibbs sampler, Markov chains reached stationarity with a run length of 10,000 iterations and a burn-in period of 5,000 iterations. The posterior estimates of the item parameters as

Table 1: Posterior estimates and Gelman-Rubin R statistics for α_1 , α_2 , β , and ρ_{12} when the specified intertrait correlation is 0.2 (chain length = 10,000, burn-in = 5,000).

True	Uniform priors Estimate (R)	Normal priors Estimate (R)	True	Uniform priors Estimate (R)	Normal priors Estimate (R)
α_1			β		
0.905	0.962 (1.036)	0.947 (1.223)	-1.433	-1.593 (1.017)	-1.553 (1.219)
0.310	0.229 (1.044)	0.237 (1.007)	-1.552	-1.610 (1.027)	-1.606 (1.010)
0.605	0.579 (1.068)	0.581 (1.007)	-0.656	-0.627 (1.005)	-0.626 (1.014)
0.117	0.082 (1.034)	0.091 (1.015)	-0.146	-0.099 (1.011)	-0.099 (1.010)
0.197	0.064 (1.008)	0.062 (1.003)	-1.206	-1.176 (0.999)	-1.169 (1.001)
0.498	0.517 (1.032)	0.518 (1.012)	-1.457	-1.728 (1.100)	-1.622 (1.021)
0.442	0.443 (1.043)	0.467 (1.085)	1.904	1.892 (1.025)	1.831 (1.057)
1.006	1.045 (1.008)	1.029 (1.117)	0.029	-0.046 (1.005)	-0.045 (1.003)
1.881	1.903 (1.191)	1.837 (1.116)	0.805	0.758 (1.037)	0.737 (1.007)
0.733	0.838 (1.024)	0.828 (1.047)	-1.267	-1.326 (1.037)	-1.296 (1.049)
0.687	0.500 (1.012)	0.506 (1.042)	1.145	1.053 (1.021)	1.056 (1.024)
1.728	1.760 (1.090)	1.663 (1.148)	-0.828	-0.898 (1.048)	-0.861 (1.107)
1.429	1.582 (1.175)	1.450 (1.085)	0.786	0.948 (1.203)	0.876 (1.033)
0.888	0.895 (1.023)	0.893 (1.052)	1.116	1.137 (1.033)	1.135 (1.017)
1.223	1.346 (1.047)	1.301 (1.117)	-0.197	-0.196 (1.005)	-0.185 (1.010)
1.155	1.348 (1.028)	1.296 (1.049)	1.265	1.222 (1.260)	1.187 (1.052)
1.582	1.728 (1.085)	1.673 (1.170)	-0.708	-0.856 (1.440)	-0.838 (1.058)
1.427	1.097 (1.019)	1.084 (1.061)	0.513	0.426 (1.021)	0.424 (1.003)
α_2			ρ_{12}		
1.504	1.722 (1.049)	1.666 (1.079)	0.2	0.183 (1.199)	0.163 (1.038)
0.886	1.047 (1.029)	1.042 (1.022)			
1.829	1.792 (1.089)	1.792 (1.033)			
1.578	1.589 (1.036)	1.558 (1.023)			
0.131	0.088 (1.018)	0.089 (1.007)			
1.562	1.888 (1.095)	1.731 (1.019)			
1.833	1.616 (1.215)	1.531 (1.008)			
0.210	0.190 (1.115)	0.197 (1.023)			
1.533	1.635 (1.175)	1.573 (1.003)			
1.436	1.309 (1.090)	1.263 (1.057)			
0.961	0.879 (1.091)	0.888 (1.039)			
0.678	0.807 (1.130)	0.783 (1.106)			
1.888	2.341 (1.281)	2.134 (1.055)			
0.618	0.744 (1.106)	0.751 (1.013)			
1.967	1.921 (1.076)	1.837 (1.034)			
1.862	1.934 (1.142)	1.862 (1.060)			
0.062	0.135 (1.169)	0.143 (1.060)			
1.389	1.223 (1.137)	1.215 (1.041)			

Table 2: Posterior estimates and Gelman-Rubin R statistics for α_1 , α_2 , β , and ρ_{12} when the specified intertrait correlation is 0.4 (chain length = 10,000, burn-in = 5,000).

True	Uniform priors Estimate (R)	Normal priors Estimate (R)	True	Uniform priors Estimate (R)	Normal priors Estimate (R)
α_1			β		
0.905	0.977 (1.177)	0.918 (1.018)	-1.433	-1.640 (1.035)	-1.583 (1.035)
0.310	0.311 (1.138)	0.297 (1.076)	-1.552	-1.790 (1.024)	-1.739 (1.035)
0.605	0.609 (1.170)	0.575 (1.057)	-0.656	-0.676 (1.039)	-0.664 (1.027)
0.117	0.116 (1.291)	0.096 (1.071)	-0.146	-0.303 (1.035)	-0.297 (1.004)
0.197	0.243 (1.005)	0.257 (1.006)	-1.206	-1.278 (1.004)	-1.280 (1.004)
0.498	0.583 (1.141)	0.551 (1.065)	-1.457	-1.567 (1.046)	-1.549 (1.021)
0.442	0.371 (1.152)	0.356 (1.073)	1.904	2.013 (1.345)	1.830 (1.519)
1.006	0.897 (1.020)	0.896 (1.021)	0.029	-0.004 (1.010)	-0.003 (1.000)
1.881	2.120 (1.007)	1.998 (1.361)	0.805	0.788 (1.040)	0.775 (1.197)
0.733	0.712 (1.126)	0.695 (1.113)	-1.267	-1.363 (1.029)	-1.353 (1.048)
0.687	0.710 (1.070)	0.692 (1.075)	1.145	1.100 (1.007)	1.104 (1.006)
1.728	2.277 (1.176)	2.026 (1.091)	-0.828	-1.034 (1.132)	-0.951 (1.131)
1.429	1.561 (1.081)	1.495 (1.075)	0.786	0.698 (1.045)	0.692 (1.012)
0.888	1.071 (1.006)	1.028 (1.068)	1.116	1.109 (1.005)	1.096 (1.004)
1.223	1.226 (1.230)	1.175 (1.036)	-0.197	-0.397 (1.021)	-0.386 (1.042)
1.155	1.185 (1.022)	1.115 (1.174)	1.265	1.017 (1.094)	1.005 (1.020)
1.582	1.450 (1.039)	1.451 (1.050)	-0.708	-0.679 (1.035)	-0.681 (1.005)
1.427	1.714 (1.062)	1.667 (1.110)	0.513	0.476 (1.069)	0.477 (1.048)
α_2			ρ_{12}		
1.504	1.417 (1.045)	1.375 (1.114)	0.4	0.434 (1.288)	0.425 (1.058)
0.886	1.011 (1.045)	0.968 (1.103)			
1.829	1.620 (1.013)	1.621 (1.127)			
1.578	1.660 (1.069)	1.628 (1.015)			
0.131	0.084 (1.012)	0.081 (1.011)			
1.562	1.430 (1.053)	1.419 (1.038)			
1.833	2.199 (1.342)	1.942 (1.335)			
0.210	0.114 (1.065)	0.126 (1.155)			
1.533	1.497 (1.097)	1.495 (1.093)			
1.436	1.372 (1.054)	1.371 (1.025)			
0.961	0.996 (1.025)	1.014 (1.132)			
0.678	0.456 (1.115)	0.486 (1.212)			
1.888	1.622 (1.152)	1.631 (1.071)			
0.618	0.437 (1.073)	0.461 (1.108)			
1.967	1.781 (1.129)	1.780 (1.290)			
1.862	1.502 (1.093)	1.517 (1.047)			
0.062	0.121 (1.083)	0.138 (1.237)			
1.389	1.320 (1.073)	1.348 (1.198)			

Table 3: Posterior estimates and Gelman-Rubin R statistics for α_1 , α_2 , β , and ρ_{12} when the specified intertrait correlation is 0.6 (chain length = 10,000, burn-in = 5,000).

True	Uniform priors Estimate (R)	Normal priors Estimate (R)	True	Uniform priors Estimate (R)	Normal priors Estimate (R)
α_1			β		
0.905	0.841 (1.159)	0.844 (1.007)	-1.433	-1.263 (1.030)	-1.251 (1.021)
0.310	0.201 (1.017)	0.216 (1.006)	-1.552	-1.561 (1.015)	-1.553 (1.046)
0.605	0.525 (1.064)	0.546 (1.098)	-0.656	-0.594 (1.071)	-0.581 (1.047)
0.117	0.105 (1.031)	0.132 (1.008)	-0.146	-0.034 (1.013)	-0.037 (1.017)
0.197	0.143 (1.003)	0.144 (1.005)	-1.206	-1.237 (1.004)	-1.237 (1.001)
0.498	0.750 (1.107)	0.723 (1.016)	-1.457	-1.426 (1.030)	-1.400 (1.060)
0.442	0.283 (1.149)	0.380 (1.199)	1.904	1.961 (1.122)	1.832 (1.119)
1.006	1.001 (1.212)	0.984 (1.054)	0.029	-0.003 (1.004)	-0.007 (1.001)
1.881	1.586 (1.125)	1.493 (1.077)	0.805	0.808 (1.039)	0.776 (1.025)
0.733	0.745 (1.040)	0.759 (1.049)	-1.267	-1.389 (1.022)	-1.387 (1.025)
0.687	0.741 (1.127)	0.726 (1.054)	1.145	1.183 (1.027)	1.179 (1.021)
1.728	1.696 (1.166)	1.652 (1.056)	-0.828	-0.675 (1.045)	-0.676 (1.009)
1.429	1.690 (1.406)	1.610 (1.087)	0.786	1.085 (1.239)	1.015 (1.196)
0.888	0.956 (1.054)	0.926 (1.045)	1.116	1.056 (1.083)	1.048 (1.015)
1.223	1.283 (1.114)	1.238 (1.117)	-0.197	-0.067 (1.077)	-0.070 (1.046)
1.155	1.376 (1.372)	1.323 (1.079)	1.265	1.352 (1.155)	1.325 (1.022)
1.582	1.439 (1.076)	1.409 (1.045)	-0.708	-0.676 (1.067)	-0.677 (1.052)
1.427	1.679 (1.234)	1.548 (1.237)	0.513	0.513 (1.048)	0.477 (1.021)
α_2			ρ_{12}		
1.504	1.459 (1.191)	1.419 (1.026)	0.6	0.646 (1.031)	0.608 (1.100)
0.886	0.968 (1.021)	0.958 (1.007)			
1.829	2.192 (1.129)	2.048 (1.041)			
1.578	1.637 (1.054)	1.592 (1.028)			
0.131	0.099 (1.014)	0.100 (1.010)			
1.562	1.247 (1.080)	1.230 (1.015)			
1.833	1.986 (1.070)	1.746 (1.028)			
0.210	0.287 (1.232)	0.302 (1.046)			
1.533	1.468 (1.161)	1.426 (1.169)			
1.436	1.611 (1.149)	1.591 (1.046)			
0.961	0.816 (1.154)	0.827 (1.025)			
0.678	0.606 (1.144)	0.640 (1.078)			
1.888	2.479 (1.113)	2.288 (1.054)			
0.618	0.560 (1.095)	0.572 (1.059)			
1.967	2.146 (1.324)	2.000 (1.036)			
1.862	1.727 (1.106)	1.695 (1.070)			
0.062	0.082 (1.091)	0.105 (1.209)			
1.389	1.394 (1.221)	1.355 (1.025)			

Table 4: Posterior estimates and Gelman-Rubin R statistics for α_1 , α_2 , and β for the CBASE data, assuming uniform priors (chain length = 10,000, burn-in = 5,000).

j	$\hat{\alpha}_1$	R	$\hat{\alpha}_2$	R	$\hat{\beta}$	R
1	0.0807	1.0546	0.5486	1.0226	-0.5861	1.0020
2	0.1396	1.0313	0.3630	1.0089	-0.6092	1.0001
3	0.0758	1.0178	0.3234	1.0001	-1.0515	0.9996
4	0.0737	1.0579	0.4157	1.0471	-1.4012	1.0183
5	0.1608	1.0259	0.4287	1.0040	-1.2238	1.0025
6	0.3050	1.0558	0.6824	1.0106	-0.9232	1.0243
7	0.1306	1.0066	0.2844	1.0227	-0.4324	1.0002
8	0.2804	1.0080	0.3106	1.0307	-1.2559	1.0024
9	0.1905	1.0478	0.4238	1.0176	-0.1286	1.0000
10	0.2425	1.0145	0.4838	1.0052	-0.0581	1.0005
11	0.1250	1.0272	0.3620	1.0016	0.4158	1.0028
12	0.0466	1.0076	0.4509	1.0154	-0.8238	1.0046
13	0.1177	1.0351	0.4070	1.0061	-0.2584	1.0019
14	0.0514	1.0115	0.3165	1.0065	-0.0282	0.9998
15	0.1378	1.0354	0.4790	1.0029	-0.8686	1.0005
16	0.1698	1.0299	0.3259	1.0113	-0.0595	1.0017
17	0.1797	1.0300	0.3652	1.0037	-0.2178	1.0004
18	0.2006	1.0100	0.1696	1.0026	-0.2419	1.0001
19	0.3423	1.0486	0.4101	1.0374	0.2914	1.0035
20	0.3028	1.1008	0.8084	1.0204	-1.2920	1.0084
21	0.2741	1.0435	0.4980	1.0020	-0.2798	1.0018
22	0.3171	1.0810	0.3796	1.0272	-0.4662	1.0045
23	0.2517	1.0528	0.5849	1.0080	-0.9708	1.0087
24	0.3044	1.0269	0.3127	1.0388	-0.1639	0.9997
25	0.2126	1.0359	0.2268	1.0244	-0.3475	1.0008
26	0.1996	1.0497	0.4524	1.0142	-0.8906	1.0007
27	0.2709	1.0072	0.1593	1.0205	-0.9018	1.0025
28	0.1758	1.0547	0.4325	1.0450	-0.6202	1.0031
29	0.2979	1.0322	0.2235	1.0454	-0.2678	1.0013
30	0.2800	1.0731	0.4221	1.0215	-0.3751	1.0011
31	0.5042	1.0055	0.2511	1.0202	-0.9097	1.0005
32	0.6259	1.0326	0.3144	1.0243	-0.9608	1.0122
33	0.2224	1.0256	0.1425	1.0027	-0.3856	1.0012
34	0.5000	1.0014	0.0972	1.0461	-0.7270	1.0012
35	0.5429	1.0106	0.2085	1.0693	-0.2614	1.0024
36	0.4537	1.0295	0.1994	1.0359	0.4221	1.0060
37	0.4280	1.0037	0.0814	1.0230	0.3117	1.0018
38	0.4286	0.9996	0.0981	1.0251	-0.4062	0.9997
39	0.3901	1.0492	0.3072	1.0351	-0.6239	1.0044
40	0.6940	1.0144	0.0697	1.0656	-0.3961	0.9997
41	0.3930	1.0049	0.0436	1.0270	-0.3781	1.0006

well as the intertrait correlation hyperparameter are fairly close to the specified parameters, suggesting that the algorithm performs well in recovering these parameters when the latent dimensions have a low to medium correlation. Further, the two sets of posterior estimates, resulting from different prior distributions, differ only slightly from each other, signifying that the posterior estimates are not sensitive to the choice of noninformative or informative priors for the slope and intercept parameters.

In the context of real data, a subset of the College Basic Academic Subjects Examination (CBASE; [54]) English data was used to demonstrate the methodology. Specifically, these data contain independent binary responses of 1,200 college students to 41 multiple-choice items. The English test is further organized to have two subtests, namely, reading and writing, so that 25 items are in the reading subtest and 16 are in the writing subtest. It is noted that the test was designed in such a manner that it conforms to the multi-unidimensional model, as each item measures one of the two latent traits. However, one may use the more general MIRT model to explore the latent structure, and in particular, to assess individual test items (i.e., to determine if the trait mainly involved in answering each item agrees with the one that it is supposed to measure). This can be accomplished by examining the estimated slope parameters, as a larger α_{ij} corresponds to a latent dimension that is more important in determining a person's success on the item. Hence, assuming uniform priors for ξ_j , Gibbs sampling was implemented to fit the MIRT model to the CBASE data with a run length of 10,000 iterations and a burn-in period of 5,000, which was sufficient for the chains to converge. An examination of the posterior estimates of α shown in Table 4 suggests that all 16 items in the writing subtest relies on the second dimension writing more than the first dimension reading. However, some items in the reading subtest, such as items 17, 19–26, 28, and 30, require further attention and modification, as they do not seem to measure mainly reading as the rest of the items do.

In summary, the proposed MCMC algorithm provides computationally efficient and accurate estimation in the context of both simulated and real data examples. Not only does the algorithm appropriately model parameters, but also the algorithm efficiently models the intertrait correlations for the compensatory MIRT model, which provides an exploratory approach for examining the latent structure of a test and detecting items that do not measure the trait they are designed to measure.

5. Concluding Remarks

The MCMC algorithm presented in this paper offers solutions for directly modeling the underlying structure of IRT models with multiple continuous latent traits. The algorithm works well when the actual intertrait correlation is low to moderate (less than 0.8), as a high correlation tends to result in high collinearity, which makes it difficult to distinguish among multiple latent traits and estimate them. With model parameters being accurately estimated, the compensatory MIRT model can be used to explore the relative importance of a latent trait in answering each test item. This is particularly useful when the underlying structure is not known, or when it is desirable to confirm the structure by examining the performance of individual items.

References

- [1] R. D. Bock and M. Aitkin, "Marginal maximum likelihood estimation of item parameters: application of an EM algorithm," *Psychometrika*, vol. 46, no. 4, pp. 443–459, 1981.

- [2] R. J. Mislevy, "Estimation of latent group effects," *Journal of the American Statistical Association*, vol. 80, no. 392, pp. 993–997, 1985.
- [3] R. J. Patz and B. W. Junker, "A straightforward approach to Markov chain Monte Carlo methods for item response models," *Journal of Educational and Behavioral Statistics*, vol. 24, no. 2, pp. 146–178, 1999.
- [4] R. K. Tsutakawa and H. Y. Lin, "Bayesian estimation of item response curves," *Psychometrika*, vol. 51, no. 2, pp. 251–267, 1986.
- [5] J. Bafumi, A. Gelman, D. K. Park, and N. Kaplan, "Practical issues in implementing and understanding Bayesian ideal point estimation," *Political Analysis*, vol. 13, no. 2, pp. 171–187, 2005.
- [6] C. S. Martin, T. Chung, L. Kirisci, and J. W. Langenbucher, "Item response theory analysis of diagnostic criteria for alcohol and cannabis use disorders in adolescents: implications for DSM-V," *Journal of Abnormal Psychology*, vol. 115, no. 4, pp. 807–814, 2006.
- [7] U. Feske, L. Kirisci, R. E. Tarter, and P. A. Pilkonis, "An application of item response theory to the DSM-III-R criteria for borderline personality disorder," *Journal of Personality Disorders*, vol. 21, no. 4, pp. 418–433, 2007.
- [8] C. L. Beseler, L. A. Taylor, and R. F. Leeman, "An item-response theory analysis of DSM-IV Alcohol-Use disorder criteria and "binge" drinking in undergraduates," *Journal of Studies on Alcohol and Drugs*, vol. 71, no. 3, pp. 418–423, 2010.
- [9] D. A. Gilder, I. R. Gizer, and C. L. Ehlers, "Item response theory analysis of binge drinking and its relationship to lifetime alcohol use disorder symptom severity in an american indian community sample," *Alcoholism*, vol. 35, no. 5, pp. 984–995, 2011.
- [10] A. T. Panter and B. B. Reeve, "Assessing tobacco beliefs among youth using item response theory models," *Drug and Alcohol Dependence*, vol. 68, pp. S21–S39, 2002.
- [11] D. Courvoisier and J. F. Etter, "Using item response theory to study the convergent and discriminant validity of three questionnaires measuring cigarette dependence," *Psychology of Addictive Behaviors*, vol. 22, no. 3, pp. 391–401, 2008.
- [12] J. S. Rose and L. C. Dierker, "An item response theory analysis of nicotine dependence symptoms in recent onset adolescent smokers," *Drug and Alcohol Dependence*, vol. 110, no. 1-2, pp. 70–79, 2010.
- [13] S. E. Fienberg, M. S. Johnson, and B. W. Junker, "Classical multilevel and Bayesian approaches to population size estimation using multiple lists," *Journal of the Royal Statistical Society. Series A*, vol. 162, no. 3, pp. 383–405, 1999.
- [14] M. Reiser, "An application of the item-response model to psychiatric epidemiology," *Sociological Methods and Research*, vol. 18, pp. 66–103, 1989.
- [15] M. Orlando, C. D. Sherbourne, and D. Thissen, "Summed-score linking using item response theory: application to depression measurement," *Psychological Assessment*, vol. 12, no. 3, pp. 354–359, 2000.
- [16] A. Tsutsumi, N. Iwata, N. Watanabe et al., "Application of item response theory to achieve cross-cultural comparability of occupational stress measurement," *International Journal of Methods in Psychiatric Research*, vol. 18, no. 1, pp. 58–67, 2009.
- [17] D. F. Andrade and H. R. Tavares, "Item response theory for longitudinal data: population parameter estimation," *Journal of Multivariate Analysis*, vol. 95, no. 1, pp. 1–22, 2005.
- [18] J. M. te Marvelde, C. A. W. Glas, G. Van Landeghem, and J. Van Damme, "Application of multidimensional item response theory models to longitudinal data," *Educational and Psychological Measurement*, vol. 66, no. 1, pp. 5–34, 2006.
- [19] M. von Davier and S. Sinharay, "An importance sampling algorithm for latent regression models," *Journal of Educational and Behavioral Statistics*, vol. 32, no. 3, pp. 233–251, 2007.
- [20] M. von Davier and S. Sinharay, "Stochastic approximation methods for latent regression item response models," *Journal of Educational and Behavioral Statistics*, vol. 35, no. 2, pp. 174–193, 2010.
- [21] R. Holman and C. A. W. Glas, "Modelling non-ignorable missing-data mechanisms with item response theory models," *The British Journal of Mathematical and Statistical Psychology*, vol. 58, no. 1, pp. 1–17, 2005.
- [22] A. Birnbaum, "Statistical theory for logistic mental test models with a prior distribution of ability," *Journal of Mathematical Psychology*, vol. 6, no. 2, pp. 258–276, 1969.
- [23] F. B. Baker and S.-H. Kim, *Item response theory: Parameter Estimation Techniques*, vol. 176, Marcel Dekker, New York, NY, USA, 2nd edition, 2004.
- [24] I. W. Molenaar, "Estimation of item parameters," in *Rasch Models: Foundations, Recent Developments, and Applications*, G. H. Fischer and I. W. Molenaar, Eds., pp. 39–51, Springer, New York, NY, USA, 1995.

- [25] A. F. M. Smith and G. O. Roberts, "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society. Series B*, vol. 55, no. 1, pp. 3–23, 1993.
- [26] L. Tierney, "Markov chains for exploring posterior distributions," *The Annals of Statistics*, vol. 22, no. 4, pp. 1701–1762, 1994.
- [27] J. H. Albert, "Bayesian estimation of normal ogive item response curves using gibbs sampling," *Journal of Educational Statistics*, vol. 17, pp. 251–269, 1992.
- [28] R. J. Patz and B. W. Junker, "Applications and extensions of MCMC in IRT: multiple item types, missing data, and rated responses," *Journal of Educational and Behavioral Statistics*, vol. 24, no. 4, pp. 342–366, 1999.
- [29] S. K. Sahu, "Bayesian estimation and model choice in item response models," *Journal of Statistical Computation and Simulation*, vol. 72, no. 3, pp. 217–232, 2002.
- [30] A. A. Béguin and C. A. W. Glas, "MCMC estimation and some model-fit analysis of multidimensional IRT models," *Psychometrika*, vol. 66, no. 4, pp. 541–561, 2001.
- [31] H. Lee, *Markov Chain Monte Carlo Methods for Estimating Multidimensional Ability in Item Response Analysis*, ProQuest LLC, Ann Arbor, Mich, USA, 1995.
- [32] Y. Sheng and C. K. Wikle, "Comparing multiunidimensional and unidimensional item response theory models," *Educational and Psychological Measurement*, vol. 67, no. 6, pp. 899–919, 2007.
- [33] Y. Sheng and C. K. Wikle, "Bayesian multidimensional IRT models with a hierarchical structure," *Educational and Psychological Measurement*, vol. 68, no. 3, pp. 413–430, 2008.
- [34] Y. Sheng and C. K. Wikle, "Bayesian IRT models in incorporating general and specific abilities," *Behaviormetrika*, vol. 36, no. 1, pp. 27–48, 2009.
- [35] L. Yao, *BMIRT: Bayesian Multivariate Item Response Theory [Computer Software]*, CTB/McGraw-Hill, Monterey, Calif, USA, 2003.
- [36] L. Yao and K. A. Boughton, "A multidimensional item response modeling approach for improving subscale proficiency estimation and classification," *Applied Psychological Measurement*, vol. 31, no. 2, pp. 83–105, 2007.
- [37] R. Levy, "The rise of markov chain monte carlo estimation for psychometric modeling," *Journal of Probability and Statistics*, vol. 2009, Article ID 537139, 18 pages, 2009.
- [38] M. D. Reckase, Ed., *Multidimensional Item Response Theory*, Springer, New York, NY, USA, 2009.
- [39] D. M. Bolt and V. F. Lall, "Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo," *Applied Psychological Measurement*, vol. 27, no. 6, pp. 395–414, 2003.
- [40] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [41] S. E. Embretson and S. P. Reise, *Item Response Theory for Psychologists*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 2000.
- [42] R. P. McDonald, *Factor Analysis and Related Methods*, Lawrence Erlbaum, Hillsdale, NJ, USA, 1985.
- [43] R. K. Tsutakawa, "Estimation of two-parameter logistic item response curves," *Journal of Educational Statistics*, vol. 9, pp. 263–276, 1985.
- [44] Y. Sheng, "A MATLAB package for Markov chain Monte Carlo with a multi-unidimensional IRT model," *Journal of Statistical Software*, vol. 28, no. 10, pp. 1–20, 2008.
- [45] Y. Sheng, "Bayesian estimation of MIRT models with general and specific latent traits in MATLAB," *Journal of Statistical Software*, vol. 34, no. 3, pp. 1–27, 2010.
- [46] K. J. Holzinger and F. Swineford, "The Bi-factor method," *Psychometrika*, vol. 2, no. 1, pp. 41–54, 1937.
- [47] J. Barnard, R. McCulloch, and X.-L. Meng, "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage," *Statistica Sinica*, vol. 10, no. 4, pp. 1281–1311, 2000.
- [48] M. Pourahmadi, "Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters," *Biometrika*, vol. 94, no. 4, pp. 1006–1013, 2007.
- [49] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–550, 1987.
- [50] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2004.
- [51] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, UK, 1996.
- [52] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, pp. 457–511, 1992.

- [53] J. P. Fox, "Multilevel IRT modeling in practice with the package mlirt," *Journal of Statistical Software*, vol. 20, no. 5, pp. 1–16, 2007.
- [54] S. Osterlind, "A national review of scholastic achievement in general education: how are we doing and why should we care?" vol. 25 of *ASHE-ERIC Higher Education Report*, George Washington University Graduate School of Education and Human Development, Washington, DC, USA, 1997.