An Improved Strategy for Bayesian Estimation of the Reduced Reparameterized Unified Model

Applied Psychological Measurement

2018, Vol. 42(2) 99–115

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146621617707511

journals.sagepub.com/home/apm

Steven Andrew Culpepper¹ and Aaron Hudson¹

Abstract

A Bayesian formulation for a popular conjunctive cognitive diagnosis model, the reduced reparameterized unified model (rRUM), is developed. The new Bayesian formulation of the rRUM employs a latent response data augmentation strategy that yields tractable full conditional distributions. A Gibbs sampling algorithm is described to approximate the posterior distribution of the rRUM parameters. A Monte Carlo study supports accurate parameter recovery and provides evidence that the Gibbs sampler tended to converge in fewer iterations and had a larger effective sample size than a commonly employed Metropolis–Hastings algorithm. The developed method is disseminated for applied researchers as an R package titled “rRUM.”

Keywords

Bayesian, diagnostic testing, latent class models, MCMC

Introduction

Cognitive diagnosis models (CDMs) offer a useful framework for relating a collection of discrete latent attributes to the probability of success on observed tasks. CDMs were originally developed to classify students into educationally relevant attribute profiles (Culpepper, 2015; Li, Cohen, Bottge, & Templin, 2015; K. K. Tatsuoaka, 1984; C. Tatsuoaka, 2002; Templin & Hoffman, 2013), and recent research has found applications in the diagnosis of pathological gambling (Templin & Henson, 2006) and social anxiety disorders (Chen, Liu, Xu, & Ying, 2015). CDMs specify either conjunctive, compensatory, or disjunctive relationships between latent attributes and observed responses. A task is conjunctive (de la Torre & Douglas, 2004; Junker & Sijtsma, 2001) if *every* attribute in a prespecified set must be correctly applied to ensure an affirmative or correct response. In contrast, a task is compensatory if lacking some of the required attributes can be offset or compensated for by having other attributes (see de la Torre & Douglas, 2004, for a summary of relevant models). Finally, a task is considered disjunctive (Maris, 1999; Templin & Henson, 2006) if at least one attribute is needed for “success” on the task as found in psychopathology research.

¹University of Illinois at Urbana–Champaign, IL, USA

Corresponding Author:

Steven Andrew Culpepper, Department of Statistics, University of Illinois at Urbana–Champaign, 725 South Wright Street, Champaign, IL 61820, USA.

Email: sculpepp@illinois.edu

Conjunctive CDMs are conceptually appealing for modeling educational tasks where success requires all the necessary attributes. The class of conjunctive CDMs includes the following models: (a) noisy inputs, deterministic, “and” gate (NIDA; Maris, 1999); (b) deterministic inputs, noisy “and” gate (DINA; de la Torre, 2009; Junker & Sijtsma, 2001); (c) reduced reparameterized unified model (rRUM; DiBello, Stout, & Roussos, 1995; Hartz, 2002; Roussos et al., 2007); and (d) generalized NIDA (GNIDA; de la Torre, 2011; Jiang, 1996; Maris, 1999).

The rRUM has received significant attention in large part due to the flexibility of the model over the NIDA and DINA models (Henson & Douglas, 2005; Henson, Roussos, Douglas, & He, 2008; Kim, 2011; Liu, Douglas, & Henson, 2009; Templin, Henson, Templin, & Roussos, 2008). Several studies considered alternative strategies for estimating the rRUM parameters, which include the use of Markov Chain Monte Carlo (MCMC; Chung, 2014; Hartz, 2002; Henson, Templin, & Willse, 2009; Rupp, Templin, & Henson, 2010), marginalized maximum likelihood estimation (Chiu, Köhn, & Wu, 2016; de la Torre, 2011), and expectation–maximization (EM; Chiu & Köhn, 2016; Feng, Habing, & Huebner, 2013). An important goal of CDMs is to classify individuals into attribute profiles, and Bayesian models offer a natural framework for summarizing the attribute posterior distribution and assessing classification uncertainty. However, there are disadvantages of previous applications of Bayesian methods for estimating rRUM parameters. That is, prior Bayesian estimation approaches (e.g., see Chung, 2014) used the Metropolis–Hastings (MH) algorithm, which requires the tedious specification of tuning parameters for rRUM item parameters in addition to a significant increase in computational burden to achieve convergence to the parameter posterior distribution (Chiu & Köhn, 2016).

This article improves Bayesian methods for estimating rRUM parameters. A data augmentation strategy is presented that uses a conjunctive condensation rule involving latent responses (Embretson, 1984; Maris, 1992, 1995, 1999; Maris, De Boeck, & Van Mechelen, 1996; Whitely, 1980). The introduction of augmented data for the rRUM allows sampling each parameter from its full conditional distribution. A more efficient Gibbs sampling algorithm is described that can be more easily implemented than existing MH samplers. In fact, Monte Carlo simulation evidence suggests that the developed Gibbs sampler achieved larger effective sample sizes and could converge in fewer iterations than a MH sampler. The developed methodology is disseminated as an R package titled “rRUM.” The Gibbs sampler was written in C++ and the program required 257 s to complete a replication of Chiu and Köhn’s (2016) $K = 4$ simulation condition with 10,000 iterations on a computer with 4 GB RAM, 3.10 GHz Intel Core processor, and 64-bit OS.

The remainder of this article includes three sections. The first section introduces a Bayesian formulation for the rRUM and discusses a new data augmentation strategy that allows sampling from tractable full conditional distributions. The second section reports results from Monte Carlo studies that compare parameter recovery (i.e., bias and root mean square error [RMSE]), mixing, and convergence between the developed Gibbs sampler and existing methods that use MH sampling. The final section provides concluding remarks.

Model Specification

Throughout this article, let $i = 1, \dots, N$, $k = 1, \dots, K$, $j = 1, \dots, J$, and $c = 1, \dots, C$ denote individuals, attributes, items, and latent classes, respectively. Let $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)'$ be a $N \times J$ matrix of binary random variables where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$ and $Y_{ij} = 1$ indicates that individual i responded correctly to item j and $Y_{ij} = 0$ otherwise. Let $\mathbf{y}$ be an $N \times J$ matrix with elements that includes the corresponding observed values of $\mathbf{Y}$. The attribute vector equals one of $C = 2^K$ attribute profiles such that $\alpha_c \in \{0, 1\}^K$, where K denotes the number of attributes. Furthermore,

$\alpha_i = (\alpha_{i1}, \dots, \alpha_{iK})'$ is a K vector of attributes for individual i with $\alpha_{ik} = 1$ if individual i mastered attribute k and 0 otherwise. Individual i belongs to class c if $\alpha_i = \alpha_c$. CDMs assume that the elements of $\mathbf{Y}$ are independent when conditioned upon the $N \times K$ matrix of latent attributes, $\alpha = (\alpha_1, \dots, \alpha_N)'$. In addition, the model assumes that the attributes required to answer each question correctly are known (i.e., the $J \times K$ $\mathbf{Q}$ matrix is correctly specified).

The GNIDA is the most general of the aforementioned conjunctive CDMs. For the GNIDA, the probability of a correct response (or endorsement) by individual i to item j is

$$P(Y_{ij} = 1 | \alpha_i, \mathbf{s}_j, \mathbf{g}_j) = \prod_{k=1}^K \left[(1 - s_{jk})^{\alpha_{ik}} g_{jk}^{1-\alpha_{ik}} \right]^{q_{jk}}, \quad (1)$$

where $q_{jk} = (\mathbf{Q})_{jk}$ is an element of the $J \times K$ $\mathbf{Q}$ matrix, which denotes whether attribute k is needed for item j . Specifically, $q_{jk} = 1$ if attribute k is needed for item j and 0 otherwise. The parameters for item j are $\mathbf{s}_j = (s_{j1}, \dots, s_{jK})$ and $\mathbf{g}_j = (g_{j1}, \dots, g_{jK})$, which are vectors of attribute-specific slipping and guessing parameters, respectively. Note the NIDA is a restricted version of the GNIDA where the slipping and guessing parameters are fixed across items such that $s_{jk} = s_k$ and $g_{jk} = g_k$ for all j .

The GNIDA is a more general model; however, Jiang (1996) showed that the GNIDA model parameters are not identified. Let $K_j = \sum_{k=1}^K q_{jk}$ be the number of attributes needed to respond as a content master to item j . Jiang (1996) showed that at most $K_j + 1$ of the $2K_j$ slipping and guessing parameters are identified. Consequently, Jiang estimated a model that included a reduced form of the GNIDA by setting the first $K_j - 1$ of the slipping parameters equal to 0. Alternatively, Hartz (2002) enforced the identifiability conditions by creating a reparameterized version of the GNIDA known as the rRUM that estimated $K_j + 1$ parameters that are functions of the $2K_j$ GNIDA slipping and guessing parameters. The item response function (IRF) of the rRUM is

$$P(Y_{ij} = 1 | \alpha_i, \pi_j^*, \mathbf{r}_j^*) = \pi_j^* \prod_{k=1}^K r_{jk}^* (1 - \alpha_{ik})^{q_{jk}}, \quad (2)$$

where π_j^* is the probability of a correct response for individuals with all requisite attributes, $\mathbf{r}_j^* = (r_{j1}^*, \dots, r_{jK}^*)'$, and $0 \leq r_{jk}^* \leq 1$ is interpreted as a penalty parameter for missing a required attribute k . More precisely, Hartz (2002) defined identified versions of the $2K_j$ GNIDA model parameters π_j^* and r_{jk}^* as follows:

$$\pi_j^* = \prod_{k=1}^K (1 - s_{jk})^{q_{jk}}, \quad r_{jk}^* = \frac{g_{jk}}{1 - s_{jk}}, \quad k = 1, \dots, K. \quad (3)$$

The probability that $Y_{ij} = y_{ij}$ for the rRUM given α_i and the item parameters is

$$p(Y_{ij} = y_{ij} | \alpha_i, \pi_j^*, \mathbf{r}_j^*) = \left(\pi_j^* \prod_{k=1}^K r_{jk}^{*(1-\alpha_{ik})q_{jk}} \right)^{y_{ij}} \left(1 - \pi_j^* \prod_{k=1}^K r_{jk}^{*(1-\alpha_{ik})q_{jk}} \right)^{1-y_{ij}}. \quad (4)$$

The likelihood of observing the $N \times J$ matrix of responses $\mathbf{y}$ conditioned on the item parameters is

$$p(\mathbf{Y} = \mathbf{y} | \pi^*, \mathbf{r}^*) = \prod_{i=1}^N \sum_{\alpha_c \in \{0,1\}^K} \pi_c \prod_{j=1}^J p(y_{ij} | \alpha_c, \pi_j^*, \mathbf{r}_j^*), \quad (5)$$

where $\boldsymbol{\pi}^* = (\pi_1^*, \dots, \pi_J^*)$, $\mathbf{r}^* = (\mathbf{r}_1^*, \dots, \mathbf{r}_J^*)$, and the probability that individual i is a member of attribute class c is $\pi_c = P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c)$. As noted by previous investigators (Chung, 2014; Hartz, 2002), the likelihood function in Equation 5 does not imply an obvious strategy for constructing full conditional distributions to employ a Gibbs sampler. Instead, researchers used MH steps to sample $\boldsymbol{\pi}^*$ and $\mathbf{r}^*$.

Data Augmentation

Data augmentation is an effective strategy for posterior computation (Tanner & Wong, 1987), and this subsection outlines an approach for conjunctive CDMs. Following previous studies (Maris, 1995, 1999; Maris et al., 1996), let $\mathbf{X}_{ij\bullet} = (X_{ij1}, \dots, X_{ijK})'$ be a vector containing attribute-specific *latent responses* associated with the random variable Y_{ij} and denote a realization of the random vector of latent responses as $\mathbf{x}_{ij\bullet}$. Note that “•” is used throughout this article to denote a vector of elements over the missing subscript. Let $X_{ijk} = 1$ if attribute k was correctly applied to item j by individual i and 0 otherwise. The latent responses are deterministically related to observed responses according to a conjunctive condensation rule $Y_{ij} = \prod_{k=1}^K X_{ijk}^{q_{jk}}$ (Maris, 1995). That is, a correct response is observed if and only if all of the necessary attributes are correctly applied and $Y_{ij} = 0$ if a least one necessary attribute is incorrectly applied.

In addition to introducing X_{ijk} , the developed Bayesian model for the rRUM reintroduces the GNIDA random variables $\mathbf{s}_j$ and $\mathbf{g}_j$ to model the relationship between $\boldsymbol{\alpha}_i$ and the latent responses $\mathbf{X}_{ij\bullet}$. The IRF for Y_{ij} can be written as a function of augmented data as

$$p(Y_{ij} | \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \boldsymbol{\pi}_j^*, \mathbf{r}_j^*) = \sum_{\mathbf{x}_{ij\bullet} \in \{0,1\}^K} p(\mathbf{x}_{ij\bullet} | \boldsymbol{\alpha}_c, \mathbf{s}_j, \mathbf{g}_j) \mathcal{I}\left(\prod_{k=1}^K x_{ijk}^{q_{jk}} = y_{ij}\right), \quad (6)$$

where $\mathcal{I}(\cdot)$ denotes the indicator function. Let $\mathbf{X} \in \mathbb{R}^{N \times J \times K}$ denote a three-way array (e.g., see Hoff, 2015) of the augmented latent response data. The augmented likelihood is

$$p(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x} | \boldsymbol{\pi}^*, \mathbf{r}^*) = \prod_{i=1}^N \sum_{\boldsymbol{\alpha}_c \in \{0,1\}^K} \pi_c \prod_{j=1}^J \left[p(\mathbf{x}_{ij\bullet} | \boldsymbol{\alpha}_c, \mathbf{s}_j, \mathbf{g}_j) \mathcal{I}\left(\prod_{k=1}^K x_{ijk}^{q_{jk}} = y_{ij}\right) \right]. \quad (7)$$

The data augmentation strategy of introducing $\mathbf{X}_{ij\bullet}$, $\mathbf{s}_j$, and $\mathbf{g}_j$ is valid given that summing over the arrangements for $\mathbf{x}_{ij\bullet}$ in Equation 7 implies the rRUM likelihood in Equation 5.

Bayesian Formulation for rRUM

The previous subsection outlined a data augmentation strategy for the rRUM, and Figure 1 presents a directed acyclic graph of the rRUM. It is important to note that the aforementioned conjunctive condensation rule has been documented (Maris, 1995) and is not in itself a new concept. Instead, the novel contribution of this article is the development of a Bayesian formulation involving the latent response condensation rule to estimate parameters of the rRUM. The Bayesian formulation for the rRUM follows:

$$Y_{ij} = \prod_{k=1}^K X_{ijk}^{q_{jk}}, \quad (8)$$

Figure 1. Directed acyclic graph of the rRUM Bayesian model formulation.

Note. Circles denote random variables. Solid arrows indicate stochastic dependence, whereas dotted arrows denote deterministic dependence. The rectangular plates enclose variables with a common subscript.

$$p(\mathbf{x}_{ij\bullet} | \boldsymbol{\alpha}_i, \mathbf{s}_j, \mathbf{g}_j) = \prod_{k=1}^K \left\{ \left[(1 - s_{jk})^{\alpha_{ik}} g_{jk}^{1-\alpha_{ik}} \right]^{x_{ijk}} \left[s_{jk}^{\alpha_{ik}} (1 - g_{jk})^{1-\alpha_{ik}} \right]^{1-x_{ijk}} \right\}^{q_{jk}}, \quad (9)$$

$$p(\mathbf{s}_j, \mathbf{g}_j) \propto \prod_{k=1}^K s_{jk}^{a_s-1} (1 - s_{jk})^{b_s-1} g_{jk}^{a_g-1} (1 - g_{jk})^{b_g-1} \mathcal{I}(0 \leq g_{jk} < 1 - s_{jk} \leq 1), \quad (10)$$

$$P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c | \boldsymbol{\pi}) = \sum_{c=1}^C \pi_c \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c), \quad (11)$$

$$\boldsymbol{\pi} = (\pi_1, \dots, \pi_C)' \sim \text{Dirichlet}(\boldsymbol{\delta}_0), \quad \boldsymbol{\delta}_0 = (\delta_{01}, \dots, \delta_{0C}). \quad (12)$$

Equation 8 is the previously stated deterministic function relating the response data to latent responses. Note the technical convention that $0^0 = 1$ is used, so Y_{ij} is only deterministically related to required latent responses where $q_{jk} = 1$. Equation 9 indicates that the latent responses are probabilistically related to $\boldsymbol{\alpha}_i$, which is a difference with the DINA model where Y_{ij} is probabilistically related to $\boldsymbol{\alpha}_i$ and the ideal responses are deterministically related to latent attributes. Equation 9 assumes $\mathbf{X}_{ij\bullet}$ are conditionally independent given $\boldsymbol{\alpha}_i$, $\mathbf{s}_j$, and $\mathbf{g}_j$. The slipping parameters are defined as $s_{jk} = P(X_{ijk} = 0 | \alpha_{ik} = 1)$, which is interpreted as the probability of slipping on latent response k and $g_{jk} = P(X_{ijk} = 1 | \alpha_{ik} = 0)$ is the probability of guessing.

Equations 8 and 9 together imply the rRUM IRF. Namely, the IRF in Equation 2 for the probability that $Y_{ij} = 1$ is obtained when $\prod_{k=1}^K X_{ijk}^{q_{jk}} = 1$. The probability that $Y_{ij} = 0$ is one minus Equation 2, which is identical to summing $p(\mathbf{x}_{ij\bullet} | \boldsymbol{\alpha}_i, \mathbf{s}_j, \mathbf{g}_j)$ over the $2^K - 1$ possible arrangements where at least one $X_{ijk} = 0$ (i.e., $\prod_{k=1}^K X_{ijk}^{q_{jk}} = 0$). Recall from Equation 3 that Hartz (2002) imposed the constraint that $0 \leq g_{jk} < 1 - s_{jk} \leq 1$ by restricting $0 \leq r_{jk}^* \leq 1$. Accordingly, the prior for the slipping and guessing parameters in Equation 10 incorporates the restriction by specifying $\mathbf{s}_j$ given $\mathbf{g}_j$ as independent truncated Beta distributions and vice versa. Note that uninformative uniform priors can be employed by setting $a_s = b_s = a_g = b_g = 1$. Equations 11 and 12

were described by Culpepper (2015) in the context of Bayesian estimation of the DINA model. That is, Equation 11 specifies a categorical prior for α_i conditioned upon π and Equation 12 indicates π has a conjugate Dirichlet distribution with a parameter vector δ_0 . Note that employing Equations 11 and 12 represents an unstructured model for the 2^K classes rather than an approximation to the latent class distribution (e.g., Hartz, 2002).

Posterior Approximation With the Gibbs Sampler

The introduction of augmented data offers a straightforward strategy for approximating the posterior distribution with MCMC. The algorithm iteratively samples from the full conditional distributions using the following steps:

1. Randomly initialize parameters $\mathbf{X}^{(0)}$, $\alpha^{(0)}$, $\mathbf{s}^{(0)}$, $\mathbf{g}^{(0)}$, and $\pi^{(0)}$.
2. For iteration $t = 1, \dots, T$, sequentially sample $\mathbf{X}^{(t)}$, $\alpha^{(t)}$, $\pi^{(t)}$, $\mathbf{s}^{(t)}$, and $\mathbf{g}^{(t)}$ as follows:
 - a. Sample the latent responses as $\mathbf{X}^{(t)} \sim p(\mathbf{X} | \mathbf{Y}, \alpha^{(t-1)}, \mathbf{s}^{(t-1)}, \mathbf{g}^{(t-1)})$.
 - b. Sample latent attributes as $\alpha^{(t)} \sim p(\alpha | \mathbf{X}^{(t)}, \mathbf{s}^{(t-1)}, \mathbf{g}^{(t-1)}, \pi^{(t-1)})$.
 - c. Sample the structural parameters as $\pi^{(t)} \sim p(\pi | \alpha^{(t)})$.
 - d. Sample $\mathbf{s}^{(t)} \sim p(\mathbf{s} | \mathbf{X}^{(t)}, \alpha^{(t)}, \mathbf{g}^{(t-1)})$.
 - e. Sample $\mathbf{g}^{(t)} \sim p(\mathbf{g} | \mathbf{X}^{(t)}, \alpha^{(t)}, \mathbf{s}^{(t)})$.
 - f. Compute $\pi^{*(t)}$ and $\mathbf{r}^{*(t)}$ in the posterior using $\mathbf{s}^{(t)}$ and $\mathbf{g}^{(t)}$ with Equation 3.

Sampling from the full conditional distributions with a large enough T yields a Markov chain that converges to the posterior distribution.

Full Conditional Distributions

This subsection presents the full conditional distributions implied by the rRUM Bayesian model formulation. Let $\mathbf{X}_{ij(k)}$ be a $K - 1$ vector that excludes the k th element X_{ijk} . The first step in the Gibbs sampler is to sample X_{ijk} from the full conditional distribution given the observed data Y_{ij} , latent response item parameters s_{jk} and g_{jk} , latent attribute α_{ik} , and $\mathbf{X}_{ij(k)}$ for all i, j , and k . The full conditional distribution for X_{ijk} is

$$X_{ijk} | Y_{ij} = y_{ij}, \mathbf{X}_{ij(k)}, \alpha_{ik}, s_{jk}, g_{jk} \sim \text{Bernoulli}(\tilde{\pi}_{ijk}), \quad (13)$$

$$\begin{aligned} \tilde{\pi}_{ijk} &= \frac{p(y_{ij} | \mathbf{X}_{ij(k)}, X_{ijk} = 1) P(X_{ijk} = 1 | \alpha_{ik}, s_{jk}, g_{jk})}{\sum_{x_{ijk}=0}^1 p(y_{ij} | \mathbf{X}_{ij(k)}, X_{ijk} = x_{ijk}) P(X_{ijk} = x_{ijk} | \alpha_{ik}, s_{jk}, g_{jk})} \\ &= \left\{ \left(1 - \prod_{k' \neq k} x_{ijk'}^{q_{jk'}} \right) \left[(1 - s_{jk})^{\alpha_{ik}} g_{jk}^{(1-\alpha_{ik})} \right]^{q_{jk}} \right\}^{1-y_{ij}}, \end{aligned} \quad (14)$$

where Equation 14 for $\tilde{\pi}_{ijk}$ is obtained using Bayes's theorem and the deterministic relationship in Equation 8 for $p(y_{ij} | \mathbf{X}_{ij(k)}, X_{ijk} = x_{ijk})$ which is

$$p(y_{ij} | \mathbf{X}_{ij(k)}, X_{ijk} = x_{ijk}) = \left(x_{ijk}^{q_{jk}} \prod_{k' \neq k} x_{ijk'}^{q_{jk'}} \right)^{y_{ij}} \left(1 - x_{ijk}^{q_{jk}} \prod_{k' \neq k} x_{ijk'}^{q_{jk'}} \right)^{1-y_{ij}}, \quad (15)$$

where $\prod_{k' \neq k} x_{ijk'}^{q_{jk'}}$ denotes the product of latent responses that excludes x_{ijk} (i.e., the product of elements in $\mathbf{X}_{ij(k)}$). Equation 15 implies restrictions must be placed on X_{ijk} given the observed

Y_{ij} and the other latent responses $\mathbf{X}_{ij(k)}$. Clearly, $Y_{ij} = 1$ implies $\tilde{\pi}_{ijk} = 1$ and $X_{ijk} = 1$ for all k . Also, $X_{ijk} = 1$ with probability $[(1 - s_{jk})^{\alpha_{ik}} g_{jk}^{(1-\alpha_{ik})}]^{q_{jk}}$ when $Y_{ij} = 0$ and at least one other latent response is 0 (i.e., $\prod_{k' \neq k} x_{ijk'} = 0$). Otherwise, $Y_{ij} = 0$ and $\prod_{k' \neq k} x_{ijk'} = 1$ implies $\tilde{\pi}_{ijk} = 0$ and $X_{ijk} = 0$.

The second step in the Gibbs sampler is to sequentially sample α_{ik} for $k = 1, \dots, K$ for all i . Let $\boldsymbol{\alpha}_{i(k)}$ denote a $K - 1$ attribute vector for individual i that excludes α_{ik} . The full conditional distribution for α_{ik} is

$$\begin{aligned} p(\alpha_{ik} | \mathbf{X}_{i\bullet k}, \mathbf{s}_{\bullet k}, \mathbf{g}_{\bullet k}, \boldsymbol{\alpha}_{i(k)}, \boldsymbol{\pi}) &\propto p(\mathbf{X}_{i\bullet k} | \alpha_{ik}, \mathbf{s}_{\bullet k}, \mathbf{g}_{\bullet k}) p(\alpha_{ik} | \boldsymbol{\pi}, \boldsymbol{\alpha}_{i(k)}), \\ &\propto \left[\prod_{j=1}^J p(x_{ijk} | \alpha_{ik}, s_{jk}, g_{jk}) \right] p(\alpha_{ik} | \boldsymbol{\pi}, \boldsymbol{\alpha}_{i(k)}), \end{aligned} \quad (16)$$

where $\mathbf{X}_{i\bullet k} = (X_{i1k}, \dots, X_{iJk})'$ is a J vector of latent responses for individual i on attribute k , $\mathbf{s}_{\bullet k} = (s_{1k}, \dots, s_{Jk})'$, and $\mathbf{g}_{\bullet k} = (g_{1k}, \dots, g_{Jk})'$. Notice that α_{ik} is influenced by $\boldsymbol{\alpha}_{i(k)}$ in the prior distribution $p(\alpha_{ik} | \boldsymbol{\pi}, \boldsymbol{\alpha}_{i(k)})$. The full conditional distribution for α_{ik} is

$$\alpha_{ik} | X_{i\bullet k}, \mathbf{s}_{\bullet k}, \mathbf{g}_{\bullet k}, \boldsymbol{\alpha}_{i(k)}, \boldsymbol{\pi} \sim \text{Bernoulli}(\tilde{\pi}_{ik}), \quad (17)$$

$$\tilde{\pi}_{ik} = \frac{\pi_{k1} \prod_{j=1}^J \left[(1 - s_{jk})^{x_{ijk}} s_{jk}^{1-x_{ijk}} \right]^{q_{jk}}}{\pi_{k1} \prod_{j=1}^J \left[(1 - s_{jk})^{x_{ijk}} s_{jk}^{1-x_{ijk}} \right]^{q_{jk}} + \pi_{k0} \prod_{j=1}^J \left[g_{jk}^{x_{ijk}} (1 - g_{jk})^{1-x_{ijk}} \right]^{q_{jk}}},$$

where $\pi_{k1} = p(\alpha_{ik} = 1 | \boldsymbol{\pi}, \boldsymbol{\alpha}_{i(k)})$ and $\pi_{k0} = p(\alpha_{ik} = 0 | \boldsymbol{\pi}, \boldsymbol{\alpha}_{i(k)})$ are elements of $\boldsymbol{\pi}$.

The third step in the Gibbs algorithm is to sample $\boldsymbol{\pi}$ from the full conditional distribution. Similar to Culpepper (2015), the full conditional distribution for the vector of latent class probabilities is

$$\boldsymbol{\pi} | \boldsymbol{\alpha} \sim \text{Dirichlet}(\boldsymbol{\delta}_0 + \tilde{\boldsymbol{\delta}}), \quad \tilde{\boldsymbol{\delta}} = \left(\sum_{i=1}^N \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_1), \dots, \sum_{i=1}^N \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c) \right)'. \quad (18)$$

Note that $\tilde{\boldsymbol{\delta}}$ is a vector with elements equal to the number of individuals in each of the 2^K latent classes. The number of profiles in attribute class c for $\tilde{\boldsymbol{\delta}}$ is easily computed. First, use the bijection $\boldsymbol{\alpha} \mathbf{v}$ (Chung, 2014; von Davier, 2014), where $\mathbf{v} = (2^{K-1}, 2^{K-2}, \dots, 2, 1)$ to map each attribute profile to an integer between 0 and $2^K - 1$. The elements of $\tilde{\boldsymbol{\delta}}$ can be computed by tabulating the elements of $\boldsymbol{\alpha} \mathbf{v}$.

The fourth step is to sample the rRUM item parameters, $\boldsymbol{\pi}^*$ and $\mathbf{r}^*$. Note that sampling $\boldsymbol{\pi}^*$ and $\mathbf{r}^*$ from the posterior distribution requires two steps. First, it is necessary to sample the augmented slipping and guessing parameters $\mathbf{s}$ and $\mathbf{g}$. Specifically, the full conditional distribution for s_{jk} and g_{jk} is

$$\begin{aligned} p(s_{jk}, g_{jk} | \mathbf{X}_{\bullet jk}, \boldsymbol{\alpha}) &\propto \left[\prod_{i=1}^N p(x_{ijk} | \alpha_{ik}, s_{jk}, g_{jk}) \right] p(s_{jk}, g_{jk}), \\ &\propto s_{jk}^{a_{sk}-1} (1 - s_{jk})^{b_{sk}-1} g_{jk}^{a_{gk}-1} (1 - g_{jk})^{b_{gk}-1} \mathcal{I}(0 \leq g_{jk} < 1 - s_{jk} \leq 1), \end{aligned} \quad (19)$$

$$\begin{aligned}
a_{sk} &= q_{jk} \sum_{i=1}^N (1 - x_{ijk}) \alpha_{ik} + a_s, \\
b_{sk} &= q_{jk} \sum_{i=1}^N x_{ijk} \alpha_{ik} + b_s, \\
a_{gk} &= q_{jk} \sum_{i=1}^N x_{ijk} (1 - \alpha_{ik}) + a_g, \\
b_{gk} &= q_{jk} \sum_{i=1}^N (1 - x_{ijk}) (1 - \alpha_{ik}) + b_g,
\end{aligned}$$

where $X_{\bullet jk} = (X_{1jk}, \dots, X_{Njk})'$ is an N vector of latent responses for item j and attribute k . Note that s_{jk} and g_{jk} can be sampled from Equation 19 by sampling s_{jk} and g_{jk} from the corresponding truncated univariate Beta distributions. The second step is to transform the latent response slipping and guessing parameters into the identified rRUM item parameters π_j^* and r_j^* . Specifically, the rRUM item parameters are defined as deterministic functions of $\mathbf{s}_j$ and $\mathbf{g}_j$, so the values for π_j^* and r_j^* are computed each iteration using Equation 3.

It may be important to note that the aforementioned full conditional distributions for X_{ijk} , s_{jk} , and g_{jk} have focused on the cases where $q_{jk} = 1$. When $q_{jk} = 0$, X_{ijk} , s_{jk} , and g_{jk} are not updated.

Monte Carlo Study

Simulation Study Design

We conducted a simulation to compare the developed Gibbs sampling algorithm to an MCMC approach that employed the MH algorithm (i.e., C++ code for the MH sampler was written based on by Chung, 2014). The Gibbs and MH samplers were compared in terms of parameter recovery, mixing, and convergence using the proportional scale reduction factor, $\hat{R}$ (Brooks & Gelman, 1998), and effective sample sizes for each parameter (Kass, Carlin, Gelman, & Neal, 1998). $\hat{R}$ and effective sample sizes were computed with the R coda package (Plummer, Best, Cowles, & Vines, 2006) by running five chains each of length 20,000. Note that $\hat{R}$ is the ratio of between-chain variance to within chain variance in the model parameters and $\hat{R} < 1.2$ is generally used as a threshold for concluding the chains have converged to the posterior distribution.

The Monte Carlo simulation study used the designs of Chiu, Douglas, and Li (2009) and Chiu and Köhn (2016). Specifically, 50 datasets for $K=3$ and $K=4$ were generated with $N=3,000$ and $J=30$. Note that $\pi_j^* = 0.9$ and $r_j^* = 0.60$ for all 50 datasets, $j = 1, \dots, J$, and $k = 1, \dots, K$. For $K=3$, the $\mathbf{Q}$ matrix was constructed by repeating the three patterns with one element equal to one four times, the three rows with two elements equal to one four times, and the vector of all ones six times. For $K=4$, the 15 possible arrangements for α_i that excludes the (0,0,0,0) row were repeated twice in $\mathbf{Q}$. Subject attribute profiles were generated in two steps. First, multivariate standard normal data were generated as $(\theta_{i1}, \dots, \theta_{iK})' \sim \mathcal{N}_K(\mathbf{0}_K, \Sigma)$, where Σ is a correlation matrix with 0.5 in the off-diagonal elements. Second, latent attributes were generated as,

$$\alpha_{ik} = \begin{cases} 1 & \theta_{ik} \geq \Phi^{-1}\left(\frac{k}{K+1}\right), \\ 0 & \text{otherwise} \end{cases}, \quad (20)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.

Figure 2. Estimated bias and RMSE for $\hat{\pi}^*$ and $\hat{r}^*$ of the Gibbs and MH samplers for $K=3$.

Note. Estimates are based on 50 replications. The Q matrix in the x-axis denotes ones with “•” and zeros with “o”. RMSE = root mean square error; MH = Metropolis–Hastings.

Figure 3. Estimated bias and RMSE for $\hat{\pi}^*$ and $\hat{r}^*$ of the Gibbs and MH samplers for $K=4$.
Note. Estimates are based on 50 replications. The Q matrix in the x-axis denotes ones with “●” and zeros with “○”.
RMSE = root mean square error; MH = Metropolis–Hastings.

Figure 4. Estimated bias and RMSE for $\hat{\pi}$ of the Gibbs and MH samplers for $K = 3$ and 4 .

Note. Estimates are based on 50 replications. The attribute profiles in the x-axis denote ones with “•” and zeros with “o”. RMSE = root mean square error; MH = Metropolis–Hastings.

The Markov chains were executed with 20,000 iterations. The developed Gibbs sampler was employed with uniform priors for $\mathbf{s}$, $\mathbf{g}$, and $\boldsymbol{\pi}$ by setting $a_s = b_s = a_g = b_g = 1$ and $\boldsymbol{\delta}_0 = \mathbf{1}$. Furthermore, five independent chains were run with randomly generated starting values for both the Gibbs and MH samplers. The starting values were generated as $\alpha_{ik} \sim \text{Bernoulli}(0.5)$, $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\delta}_0)$, and $p(s_{jk}, g_{jk}) \propto \mathcal{I}(0 \leq g_{jk} < 1 - s_{jk} \leq 1)$. Unless otherwise noted, the simulation results used a burn-in of 10,000 iterations for both the Gibbs and MH samplers.

Simulation Study Results

Figures 2 through 4 present estimated bias and RMSE for item and structural parameters for the Gibbs and MH samplers for $K = 3$ and 4 . The results in Figures 2 and 3 provide evidence that the Gibbs and MH samplers both yielded accurate item parameters. Furthermore, RMSE for the

Figure 5. Median PSRF across burn-in values for model parameters from Monte Carlo simulation studies for Gibbs and MH samplers and $K = 3$ and 4.

Note. Estimates are based on 50 replications. The upper bolded, dashed line plots the maximum value of PSRF for the model parameters. PSRF = proportional scale reduction factor; MH = Metropolis–Hastings.

skill-specific penalty parameters (i.e., $\mathbf{r}^*$) tended to be larger for both methods for items that required more skills. For $K = 3$, the maximum, minimum, and average deviation from the true model parameters across all 50 replications were 0.166, 2.20×10^{-5} , and 0.0185, respectively, for the Gibbs sampler and 0.161, 1.83×10^{-5} , and 0.0184, respectively, for the MH sampler. Similarly, the maximum, minimum, and average deviation for $K = 4$ were 0.229, 3.00×10^{-6} , and 0.0237, respectively, for Gibbs and 0.220, 4.67×10^{-6} , and 0.0231, respectively, for MH. Figure 4 provides additional evidence the Gibbs and MH samplers accurately recovered the structural parameter $\boldsymbol{\pi}$.

The Gibbs and MH samplers differed in terms of mixing and convergence. The simulation included 50 replications per condition, and Figure 5 plots the median $\hat{R}$ from 50 replications for all model parameters to assess the number of iterations needed for convergence to the posterior distribution for the Gibbs and MH samplers. Figure 5 provides evidence the maximum $\hat{R}$ across all parameters, and replications for the Gibbs sampler fell below 1.1 by 10,000 iterations for $K = 3$ and 4. In contrast, the maximum $\hat{R} > 2$ for the MH sampler, which provides evidence of relatively poor mixing.

Figures 6 and 7 plot the median, minimum, and maximum effective sample sizes to provide additional insight regarding the degree of autocorrelation in the Markov chains (i.e., larger effective sample sizes are associated with less autocorrelation). The simulation study computed effective sample size from five chains of length 20,000 with a burn-in of 10,000 for an apparent

Figure 6. Median, minimum, and maximum effective sample sizes (ESS) for π^* and r^* of the Gibbs and Metropolis–Hastings samplers for $K=3$.

Note. Estimates are based on 50 replications. The Q matrix in the x-axis denotes ones with “•” and zeros with “o”.

sample size (i.e., the number of iterations used to estimate the parameters after burn-in) of 50,000. The results in Figures 6 and 7 provide evidence that the Gibbs sampler produced samples with less autocorrelation. For instance, for $K=3$, the effective sample for the MH sampler never exceeded 626 and was as low as 11 based upon an apparent sample size of 50,000. The minimum and maximum effective sample sizes for $K=4$ were 9 and 678, respectively. In contrast, the effective sample sizes of item parameters for the Gibbs sampler were larger than the MH sampler with values ranging between 973 and 13,991 for $K=3$ and 491 and 12,396 for $K=4$. The Gibbs effective sample sizes were generally larger for π_j^* and smaller for r_j^* for items with complex structure. One explanation could be that there is dependence among the r_j^* for items with complex structure. In addition, the effective sample sizes for r_{j1}^* tended to be smaller for the $K=4$ case across item complexity, which could be attributed to the fact that a smaller threshold was used to generate α_{i1} and there were fewer nonmasters on the first attribute.

Figure 7. Median, minimum, and maximum effective sample sizes (ESS) for π^* and r^* of the Gibbs and Metropolis–Hastings samplers for $K=4$.

Note. Estimates are based on 50 replications. The $\mathbf{Q}$ matrix in the x-axis denotes ones with “●” and zeros with “○”.

Discussion

This article presented a novel data augmentation strategy for the conjunctive rRUM. The primary contribution of the article is the introduction of a data augmentation strategy that implies tractable full conditional distributions for rRUM parameters. One consequence is that Gibbs sampling can be implemented as opposed to the traditionally employed MH sampling. The Monte Carlo simulation provided evidence that the Gibbs sampler achieved better mixing than MH.

Another potential advantage of the developed Bayesian formulation for the rRUM relates to the modeling of α_i and π . Specifically, the proposed formulation improves upon existing Bayesian strategies (e.g., Culpepper, 2015) for estimating attribute vectors by sampling α_{ik} conditioned upon the remaining attributes as opposed to sampling from a 2^K multinomial distribution. The developed approach requires only $2K$ computations per individual to sample α_i rather than 2^K . Furthermore, the developed model employs an unstructured model for π and may be more flexible than approximations that assume an underlying multivariate normal distribution (e.g., see Chiu & Köhn, 2016; Hartz, 2002). Future research should consider the relative merits of the unstructured model in this article to commonly employed approximations for the rRUM latent class distribution (e.g., see Hartz, 2002; Templin et al., 2008).

Prior applications of MH samplers for estimating rRUM parameters have been computationally intensive. Another contribution of this study is that the new Gibbs sampler is disseminated as an R package ("rRUM"). As noted above, the R code showed significant improvement in computational speed. Consequently, the availability of the new R package provides applied researchers with reasonably efficient software for estimating rRUM parameters.

In conclusion, this article reports a new data augmentation scheme for efficient Bayesian estimation of the rRUM. The rRUM is a more flexible conjunctive model than the NIDA and DINA and the data augmentation strategy developed in this article supports future applications and theoretical work.

Acknowledgment

This article benefited from the comments and suggestions of Jeffrey Douglas. Any remaining errors belong to the authors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.
- Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110(510), 850-866.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74, 633-665.

- Chiu, C.-Y., & Köhn, H.-F. (2016). The reduced RUM as a logit model: Parameterization and constraints. *Psychometrika*, 81, 350-370.
- Chiu, C.-Y., Köhn, H.-F., & Wu, H.-M. (2016). Fitting the reduced RUM with Mplus: A tutorial. *International Journal of Testing*, 16, 331-351.
- Chung, M. (2014). *Estimating the Q-matrix for cognitive diagnosis models in a Bayesian framework* (Unpublished doctoral dissertation). Columbia University, New York, NY.
- Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics*, 40, 454-476.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Routledge.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*, 49, 175-186.
- Feng, Y., Habing, B. T., & Huebner, A. (2013). Parameter estimation of the reduced RUM using the EM algorithm. *Applied Psychological Measurement*, 38, 137-150.
- Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Champaign.
- Henson, R. A., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29, 262-277.
- Henson, R. A., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. *Applied Psychological Measurement*, 32, 275-288.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.
- Hoff, P. D. (2015). Equivariant and scale-free Tucker decomposition models. *Bayesian Analysis*.
- Jiang, H. (1996). *Applications of computational statistics in cognitive diagnosis and IRT modeling* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Champaign.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, 52, 93-100.
- Kim, Y.-H. (2011). Diagnosing EAP writing ability using the reduced reparameterized unified model. *Language Testing*, 28, 509-541.
- Li, F., Cohen, A., Bottge, B., & Templin, J. L. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, 76, 181-204.
- Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, 33, 579-598.
- Maris, E. (1992). *Psychometric models for psychological processes and structures* (Unpublished doctoral dissertation). University of Leuven, Belgium.
- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, 60, 523-547.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika*, 61, 7-29.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC (R package Version 0.16-1) [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=coda>
- Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnosis system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275-318). Cambridge University Press.

- Rupp, A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford Press.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398), 528-540.
- Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 51, 337-350.
- Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Computer-Based Education Research Laboratory, University of Illinois at Urbana-Champaign, Champaign.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.
- Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement*, 32, 559-574.
- Templin, J. L., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, 32(2), 37-50.
- von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67, 49-71.
- Whitley, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479-494.