This is a "preproof" accepted article for *Psychometrika*.
This version may be subject to change during the production process.
DOI: 10.1017/psy.2024.7

EFFICIENT AND EFFECTIVE VARIATIONAL BAYESIAN INFERENCE METHOD FOR LOG-LINEAR COGNITIVE DIAGNOSTIC MODEL

Xue Wang, Jiwei Zhang, Jing Lu

Abstract

In this paper, we propose a novel and highly effective variational Bayesian Expectation Maximization-Maximization (VBEM-M) inference method for log-linear cognitive diagnostic model (CDM). In the implementation of the variational Bayesian approach for the saturated log-linear CDM, the conditional variational posteriors of the parameters that need to be derived are in the same distributional family as the priors; the VBEM-M algorithm overcomes this problem. Our algorithm can directly estimate the item parameters and the latent attribute-mastery pattern simultaneously. In contrast, Yamaguchi and Okada's (2020a) variational Bayesian algorithm requires a transformation step to obtain the item parameters for the LCDM model. We conducted multiple simulation studies to assess the performance of the VBEM-M algorithm in terms of parameter recovery, execution time, and convergence rate. Furthermore, we conducted a series of comparative studies on the accuracy of parameter estimation for the DINA model and the saturated LCDM, focusing on the

VBEM-M, VB, expectation-maximization (EM), and Markov chain Monte Carlo (MCMC) algorithms. The results indicated that our method can obtain more stable and accurate estimates, especially for the small sample sizes. Finally, we demonstrated the utility of the proposed algorithm using two real datasets.

Keywords: cognitive diagnostic assessments; Expectation-Maximization algorithm; log-linear cognitive diagnostic model; Markov chain Monte Carlo; variational Bayesian algorithm

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

1. Introduction

Cognitive diagnostic assessments (CDAs) have developed rapidly over the past several decades, and they are widely used in educational and psychological research (de la Torre, 2009, 2011; de la Torre & Douglas, 2004; DiBello et al., 2007; Haberman & von Davier, 2007; Henson et al. 2009; Junker & Sijtsma, 2001; Rupp et al., 2010; von Davier, 2014a; Templin & Henson, 2006). The primary motivation for the development of CDAs is to ascertain whether or not a student has mastered some fine-grained skills or attributes that are required to solve a particular item. More specifically, not only can CDAs be used to analyze in detail the strengths and weaknesses of students in the areas they are learning, but they can also provide powerful tools to help teachers improve classroom instruction.

There is a wide variety of cognitive diagnostic models (CDMs) available in the published CDA literature (DiBello et al., 2007; Rupp & Templin, 2008b), and many of these are built on strong cognitive assumptions about the processes involved in problem-solving. These CDMs can be broadly classified into three different types: compensatory, non-compensatory, and general models. Compensatory models are based on the assumption of attribute compensation, which means that although the examinee may not have mastered all the attributes involved in an item, they are still more likely to score well on that item if they have mastered some of its attributes. This is because the attributes that the examinee has mastered can "compensate" for the other attributes that they have not mastered. The most famous compensatory model is the deterministic inputs,

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

noisy "or" gate (DINO) model (Templin & Henson, 2006) and the linear logistic model (LLM; Maris, 1999). In contrast, non-compensatory models are constructed under the assumption of attribute conjunction, which means that under the assumption of an ideal response, an examinee can score on an item only after mastering all of the attributes involved in that item; otherwise, he or she will not be able to answer the item correctly. The widely used non-compensatory (conjunctive) models are the deterministic inputs, noisy and gate (DINA) model (Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977) and the reduced reparameterized unified model (rRUM; Hartz, 2002). Some general CDM frameworks have also been established that include a variety of widely applied CDMs, such as the log-linear CDM (LCDM; Henson et al., 2009), the generalized DINA (GDINA; de la Torre, 2011) model, and the general diagnostic model (von Davier, 2008). Although DINA, DINO, rRUM, and LLM were developed from different application backgrounds, they can in fact be viewed as special cases of the LCDM by restricting certain parameters to zero in its saturated version. Henson et al. (2009) detailed how the LCDM can be transformed into our traditional models such as DINA, DINO, rRUM, and LLM through parameter restrictions. Additionally, Ma and de la Torre (2016) elucidated that the LCDM and GDINA models are equivalent in their saturated forms.

Parameter estimation is the basis of model applications, and it is a prerequisite for interpretation of complicated data in the field of educational psychology. Several strategies have been developed to estimate the parameters of CDMs. Algorithms based on maximum

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

likelihood have been widely used to estimate CDMs in the frequency framework. Examples using a marginal maximum likelihood method to estimate the parameters of several CDMs via an Expectation-Maximization (EM) algorithm (Dempster et al., 1977) can be found in the literature (de la Torre, 2009, 2011; Ma & de la Torre, 2016; Ma & Guo, 2019; Maris, 1999). Some available R packages, such as "CDM" (George et al., 2016) and "GDINA" (Ma & de la Torre, 2020), have been developed to estimate CDM parameters. However, algorithms based on maximum likelihood have some disadvantages, as elaborated by Yamaguchi and Templin (2022); for example, there is the possibility of a local maximum being reached by a maximum likelihood algorithm. Accordingly, it is challenging to discern whether parameter estimates are obtained from a global maximum, even if a multiple starting value method is used to evaluate their optimality. In addition, calculation of the variability (standard errors) of parameter estimates depends on asymptotic theory in the likelihood framework, and an asymptotic distribution with parameter restrictions may not be correct when small sample sizes are involved.

In parallel with maximum likelihood-based methods, Bayesian statistical methods have also gained widespread attention for inferring various types of CDM parameters (e.g., Chung, 2019; Culpepper, 2015, 2019; Culpepper & Hudson, 2018; DeCarlo, 2012; de la Torre & Douglas, 2004; Henson et al., 2009; Jiang & Cater, 2019; Liu, 2022; Liu et al., 2020; Zhan et al., 2019). More specifically, de la Torre and Douglas (2004) implemented a Metropolis–Hastings (MH) algorithm for estimating the higher-order DINA model

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

parameters. Henson et al. (2009) also adopted the MH algorithm to estimate LCDM parameters. Liu et al. (2020) and Liu (2022) developed the Metropolis–Hastings Robbins–Monro (MH-RM) algorithm (Cai, 2010) to estimate CDM parameters. With the help of conjugate prior distributions, Culpepper (2015) proposed a Gibbs sampling algorithm to estimate the parameters of the DINA model; the corresponding R package "dina" was developed by Culpepper in 2015. On the basis of the work of Culpepper (2015), a new No-U-Turn Gibbs sampler was proposed by da Silva et al. (2018) to estimate the parameters of the DINA model. In addition, the Gibbs sampling algorithm has also been used for updating the Q-matrix in CDMs (Chung, 2019; Culpepper, 2019; Culpepper & Hudson, 2018). DeCarlo (2012) developed the software OpenBUGS (Thomas et al., 2006) for estimating reparameterized DINA model parameters. Zhan et al. (2019) published a tutorial for estimating various types of CDM estimation using the R package "R2jags" (Su & Yajima, 2015), which is associated with the JAGS program (Plummer, 2003). Jiang and Cater (2019) estimated the parameters of the LCDM by means of the Hamiltonian Monte Carlo (HMC) algorithm (Neal, 2011) in the Stan program (Carpenter et al., 2017). However, the computationally intensive nature of Markov chain Monte Carlo (MCMC) estimation for the CDM parameters presents a major hurdle to its widespread use in the empirical application of Bayesian approaches to the study of education when faced with large samples, numerous items, numerous attributes, and complex models (Yamaguchi & Okada, 2020a; Oka et al., 2022).

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Researchers have recently become interested in the variational inference (VI) method as a more flexible and less computationally intensive alternative to traditional Bayesian statistical methods (Bishop, 2006; Blei et al., 2017; Cho et al., 2021; Grimmer, 2011; Jaakkola & Jordan, 2000; Jeon et al., 2017; Oka & Okada, 2022; Rijmen et al., 2016; Urban & Bauer, 2021; Yamaguchi, 2020; Yamaguchi & Martinez, 2021; Yamaguchi & Okada, 2020a, 2020b). Compared to the traditional MCMC methods, the VI method is a deterministic approximation approach that is based on posterior density factorization. This method accomplishes its goal of rapidly and efficiently dealing with large amounts of complex educational psychology data (e.g., large numbers of samples, items, and attributes) by transforming the statistical inference problem of the posterior density into an optimization problem. In view of their many benefits, VI algorithms have been developed to estimate a variety of psychological models such as item response theory models (Rijmen et al., 2016; Urban & Bauer, 2021), generalized linear mixed models (Jeon et al., 2017), and CDMs (Oka & Okada, 2023; Oka, Saso, & Okada, 2023; Yamaguchi, 2020; Yamaguchi & Martinez, 2023; Yamaguchi & Okada, 2020a, 2020b).

Recently, Yamaguchi and Okada (2020b) introduced a VI method specifically tailored for the DINA model, marking a significant advancement in this field. This method was derived based on the optimal variational posteriors for each model parameter. Subsequently, Yamaguchi (2020) further extended VB inference applications by developing an algorithm for the multiple-choice item of the DINA model (MC-DINA). This extension

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

to MC-DINA demonstrated the flexibility and computational efficiency of VB methods. Subsequently, Yamaguchi and Okada (2020a) developed a VB inference algorithm for saturated CDMs. They ingeniously introduced a G-matrix, reformulating existing generalized CDMs, typically parameterized by attribute parameters, into a Bernoulli mixture model. This reformulation facilitated conditionally conjugate priors for model parameters, simplifying the derivation process and enhancing algorithmic efficiency.Oka et al. (2023) sustained this trajectory of innovation by developing a VB algorithm for a polytomous-attribute saturated CDM. Their work, building on the foundational research of Yamaguchi and Okada (2020a), not only advanced the field but also incorporated parallel computing configuration. This significantly improved the computational efficiency of the VB algorithm, demonstrating its evolving capability to handle more complex CDM structures. Simultaneously, Oka and Okada (2023) tackled scalability challenges in CDMs by developing an estimation algorithm for the Q-matrix of DINA model. Their approach, integrating stochastic optimization with variational inference in an iterative algorithm, showcased the adaptability and robustness of VB methods in dealing with large-scale CDMs. This series of developments highlight the ongoing progress and effectiveness of VB methods in the estimation of diverse models within the CDMs framework.

To date, no VB algorithms have been developed to directly estimate the item parameters in the LCDM with a logit link function. This is largely due to the challenges in directly deriving the conditional posterior density of these item parameters. Although

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Yamaguchi and Okada (2020a) proposed the variational EM (VEM) algorithm to estimate the LCDM, they actually used the least-squares transformation method (de la Torre et al., 2011) to convert the estimates of the item response probability of the item-specific attribute-mastery pattern parameters, obtained through the VEM algorithm, into the corresponding item parameters of the LCDM. Furthermore, Yamaguchi and Templin (2022) employed a one-to-one mapping within the Bayesian framework to equivalently transform the item response probability parameters, obtained through the Gibbs sampling algorithm, into item parameters in the LCDM model. This paper effectively bridges this gap by proposing a novel and highly effective variational Bayesian EM-maximization (VBEM-M) algorithm for estimating the saturated LCDM. Briefly, we obtained a tight lower bound on the likelihood function of the LCDM model using Taylor expansion (Jaakkola & Jordan, 2000), where the item parameters take a quadratic form. This allows for the existence of a conjugate prior distribution, enabling the implementation of the VI method. Consequently, the VI algorithm can be executed in the LCDM by deriving a specific posterior distribution for the item parameters, originating from the Gaussian prior distribution that serves as the conjugate prior for item parameters.

We outline the benefits from the following perspectives to highlight the advantages by which the VBEM-M algorithm excels above the other algorithms. Firstly, our VBEM-M algorithm overcomes the problem of the conditional variational posteriors of the parameters that need to be derived being in the same distributional family as the priors in

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

the implementation of the VI method for the saturated LCDM formulation. Secondly, the VBEM-M algorithm can directly estimate the item parameters and latent attribute-mastery pattern (also called "attribute profile") simultaneously, unlike Yamaguchi and Okada's (2020a) VEM algorithm, which requires a two-step process to acquire the estimation of item parameters. Thirdly, the VBEM-M algorithm can obtain a more stable and accurate estimate than an EM algorithm, especially in high-dimensional and small sample size conditions. Finally, our VBEM-M algorithm offers considerable benefits in computing time compared to the time-consuming traditional MCMC algorithms. This is because we use the VI method to transform a posterior inference issue into an optimization problem.

The rest of this paper is organized as follows. Section 2 presents the LCDM and its special case, the DINA model; Section 3 introduces the specific implementation of the VBEM-M algorithm for estimating the LCDM. Section 4 presents three simulation studies that evaluate the performance of the VBEM-M algorithm in parameter recovery across different simulation conditions, and compares the performance of the VBEM-M, VB, MCMC, and EM algorithms. Section 5 uses two empirical examples to demonstrate the model estimation results of these four algorithms. Finally, some concluding remarks are presented in Section 6.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

2. Cognitive Diagnostic Models

2.1. Log-Linear Cognitive Diagnostic Model

In this study, we focused on the LCDM. This is because it is a general model that contains a large number of models that have been previously discussed, such as DINA, DINO, rRUM, and LLM (Henson et al., 2009). More importantly, the LCDM can provide a parameterization that not only enables it to characterize the differences between the various models but also offers support for more complex data structures (Henson et al., 2009). In fact, any possible set of constraints for the saturated form LCDM can be used to define a model that fits the item response in the framework of cognitive theory. Moreover, a better understanding of the relationships between compensatory models and non-compensatory models can be described in the general parametric form. After this, a brief introduction to the LCDM will be given.

First, we define several indices that will be important throughout this paper. Each examinee is denoted by i ($i = 1, \cdots, N$), each item by j ($j = 1, \cdots, J$), each attribute by k ($k = 1, \cdots, K$), and the latent class corresponding to an attribute profile is denoted by l ($l = 1, \cdots, L$). We consider the latent attribute α_{ik} to be a binary variable, where the absence or presence of the corresponding attribute is represented by the values 0 and 1, respectively. $\boldsymbol{\alpha}_i = (\alpha_{i1}, \cdots, \alpha_{ik}, \cdots, \alpha_{iK})^{\mathrm{T}}$ is a vector of K-dimensional latent attribute profiles for the ith examinee. In light of the categorical nature of the latent classes, $\boldsymbol{\alpha}_i$ belongs to one of $L = 2^K$ latent attribute profiles. Defining $\widetilde{\boldsymbol{\alpha}}_l = (\widetilde{\alpha}_{l1}, \cdots, \widetilde{\alpha}_{lk}, \cdots, \widetilde{\alpha}_{lK})^{\mathrm{T}}$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

as the attribute profile for examinees of class l, where $\widetilde{\alpha}_{lk}$ is 1 if the examinees of class l acquire skill k and 0 otherwise, will be useful in the following.

$\widetilde{\mathbf{A}} = (\widetilde{\boldsymbol{\alpha}}_1, \cdots, \widetilde{\boldsymbol{\alpha}}_l, \cdots, \widetilde{\boldsymbol{\alpha}}_L)^{\mathrm{T}}$ denotes a matrix of $L \times K$ dimensions containing all the attribute profiles. The Q-matrix(Tatsuoka, 1983) is a $J \times K$ matrix used to describe the relationship between attributes and items, where $\boldsymbol{q}_j^{\mathrm{T}} = (q_{j1}, \cdots, q_{jk}, \cdots, q_{jK})$, and $q_{ik} \in \{0, 1\}$ is a vector of the jth row of the Q-matrix; that is, $\mathbf{Q} = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_j, \cdots, \boldsymbol{q}_J)^{\mathrm{T}}$: $q_{ik} = 1$ if the attribute k is required by item j, and $q_{ik} = 0$ otherwise. Next, a binary latent indicator variable $\boldsymbol{z}_i = [z_{i1}, \cdots, z_{il}, \cdots, z_{iL}]^{\mathrm{T}}$ is introduced, which satisfies $\sum_{l=1}^{L} z_{il} = 1$, where $z_{il} = 1$ denotes the ith examinee belonging to the lth attribute profile (i.e., $\boldsymbol{\alpha}_i = \widetilde{\boldsymbol{\alpha}}_l$). Let x_{ij} be the observed item response for the ith examinee to the jth item: $x_{ij} = 1$ if the ith examinee gives the correct answer for the jth item, and it is 0 otherwise. The corresponding item response matrix for all examinees answering all items is $\mathbf{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_i, \cdots, \boldsymbol{x}_N)^{\mathrm{T}}$, where $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ij}, \cdots, x_{iJ})^{\mathrm{T}}$, $i = 1, \cdots, N$. Then, the probability of a correct response for the LCDM can be expressed as

$$P(x_{ij} = 1 | \boldsymbol{\alpha}_i = \widetilde{\boldsymbol{\alpha}}_l, \eta_j, \boldsymbol{\lambda}_j, \boldsymbol{q}_j) = \frac{\exp(\boldsymbol{\lambda}_j^{\mathrm{T}} \mathbf{h}(\widetilde{\boldsymbol{\alpha}}_l, \boldsymbol{q}_j) + \eta_j)}{1 + \exp(\boldsymbol{\lambda}_j^{\mathrm{T}} \boldsymbol{h}(\widetilde{\boldsymbol{\alpha}}_l, \boldsymbol{q}_j) + \eta_j)}, \tag{1}$$

where η_j is the intercept parameter, and $\exp(\eta_j)/(1 + \exp(\eta_j))$ indicates the probability that an examinee answers correctly on item j if he or she does not master any of the attributes examined on that item. $\boldsymbol{\lambda}_j = (\lambda_{j1}, \cdots, \lambda_{jK}, \lambda_{j12}, \cdots, \lambda_{j12\cdots(K-1)K})^{\mathrm{T}}$ is the slope parameter vector, which is composed of a $D \times 1$ vector, where $D = 2^K - 1$. $\mathbf{h}(\widetilde{\boldsymbol{\alpha}}_l, \boldsymbol{q}_j)$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

represents a set of linear combinations of $\widetilde{\boldsymbol{\alpha}}_l$ and $\boldsymbol{q}_j$:

$$\boldsymbol{\lambda}_j^{\mathrm{T}}\mathbf{h}(\widetilde{\boldsymbol{\alpha}}_l, \boldsymbol{q}_j) = \sum_{k=1}^{K}\lambda_{jk}\widetilde{\alpha}_{lk}q_{jk} + \sum_{k=1}^{K}\sum_{k'>k}\lambda_{jkk'}\widetilde{\alpha}_{lk}\widetilde{\alpha}_{ik'}q_{jk}q_{jk'} + \cdots + \lambda_{j12\cdots(K-1)K}\prod_{k=11}^{K}\widetilde{\alpha}_{lk}q_{jk}. \quad (2)$$

Combining the latent variable $\boldsymbol{z}_i$ and Eq. (2), the LCDM can be rewritten as:

$$P(x_{ij}=1|\widetilde{\mathbf{A}}, \boldsymbol{z}_i, \eta_j, \boldsymbol{\lambda}_j, \boldsymbol{q}_j) = \prod_{l=1}^{L}P(x_{ij}=1|\boldsymbol{\alpha}_i = \widetilde{\boldsymbol{\alpha}}_l, \eta_j, \boldsymbol{\lambda}_j, \boldsymbol{q}_j)^{z_{il}}. \quad (3)$$

2.2. DINA Model

The DINA model, as a special case of the LCDM, has a relatively straightforward structure and widespread adoption in cognitive diagnostic assessments; specialized software packages are also available for a number of estimation techniques grounded in the model. Therefore, we provide a short overview of the traditional DINA model and its interconversion with the LCDM. Two item parameters have been introduced in the traditional DINA models for each item j: s_j is the slipping parameter and g_j is the guessing parameter, and the probability of a correct response can be written as

$$\begin{aligned} P(x_{ij}=1|\boldsymbol{\alpha}_i = \widetilde{\boldsymbol{\alpha}}_l, g_j, s_j) &= g_j^{1-\gamma_{lj}}(1-s_j)^{\gamma_{lj}}, \\ \gamma_{lj} &= \prod_{k=1}^{K}\alpha_{lk}^{q_{jk}}, \end{aligned} \quad (4)$$

where γ_{ij} is the ideal response pattern. $\gamma_{lj}=1$ indicates that examinee i possesses all the required attributes for item j; otherwise, $\gamma_{lj}=0$. The parameters s_j and g_j can be formally defined by

$$\begin{aligned} s_j &= P(x_{ij}=0|\gamma_{lj}=1), \\ g_j &= P(x_{ij}=1|\gamma_{lj}=0). \end{aligned} \quad (5)$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Since the estimation approach presented in this work is based on the LCDM, we must first convert the DINA model to LCDM format. Our next topic is the connection between the DINA model and the LCDM and how they may be converted back and forth.

Let $\widetilde{\boldsymbol{K}}_j = \{k : \text{attribute } k \text{ is measured by item } j\}$ denote an indicator set of attributes investigated by item j and K_j^* denote the number of investigated attributes. Then, the DINA model can be rewritten in the form:

$$P(x_{ij} = 1|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, g_j, s_j) = \frac{\exp\left(\eta_j + \lambda_{j\tilde{K}_{j1}\tilde{K}_{j2}\cdots\tilde{K}_{jK_j^*}} \prod\limits_{k^*\in\tilde{K}_j} \alpha_{lk^*} q_{jk^*}\right)}{1 + \exp\left(\eta_j + \lambda_{j\tilde{K}_{j1}\tilde{K}_{j2}\cdots\tilde{K}_{jK_j^*}} \prod\limits_{k^*\in\tilde{K}_j} \alpha_{lk^*} q_{jk^*}\right)}, \tag{6}$$

where

$$\begin{aligned} \eta_j &= -\log\left(\frac{g_j}{1-g_j}\right), \\ \lambda_{j\tilde{K}_{j1}\tilde{K}_{j2}\cdots\tilde{K}_{jK_j^*}} &= -\eta_j + \log\left(\frac{1-s_j}{s_j}\right). \end{aligned} \tag{7}$$

For simplicity, we denote $\lambda_{j\tilde{K}_{j1}\tilde{K}_{j2}\cdots\tilde{K}_{jK_j^*}}$ as λ_j and the DINA model is equivalent to the following form:

$$P(x_{ij} = 1|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, g_j, s_j) = \frac{\exp\left(\eta_j + \lambda_j \prod\limits_{k=1}^{K} \alpha_{lk}^{q_{jk}}\right)}{1 + \exp\left(\eta_j + \lambda_j \prod\limits_{k=1}^{K} \alpha_{lk}^{q_{jk}}\right)}. \tag{8}$$

While this study focuses mostly on the LCDM, various variants of the LCDM, such as the DINO model, LLM, and saturated LCDM, are also discussed. We will therefore not go into great depth here; instead, the reader should refer to the online supplemental materials for the necessary information.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

3. Variational Bayesian EM-Maximization Algorithm for the LCDM

3.1. Variational Bayesian EM algorithm

Since it is straightforward to convert an approximate conditional posterior distribution problem into an optimization problem using VI methods, these techniques see extensive application in inferring Bayesian models in the area of machine learning (Jordan et al., 1999; Bishop, 2006; Beal, 2003). Next, we briefly outline the implementation process of the variational Bayesian EM (VBEM) algorithm (Beal, 2003). Assume that the observed dataset $\boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i, \cdots, \boldsymbol{y}_N)$ is produced by model $\mathcal{M}$, where model $\mathcal{M}$ consists of the latent variables $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_i, \cdots, \boldsymbol{\zeta}_N)$ and model parameters $\boldsymbol{\theta}$. Next, we specify a variational density family $\mathcal{Q}$ over the unknown variables $\boldsymbol{\zeta}$ and $\boldsymbol{\theta}$. The purpose of this is to establish the optimal approximation $q(\boldsymbol{\zeta}, \boldsymbol{\theta}) \in \mathcal{Q}$ to their posterior distribution using this specified variational density (i.e., $q(\boldsymbol{\zeta}, \boldsymbol{\theta}) \rightsquigarrow p(\boldsymbol{\zeta}, \boldsymbol{\theta}|\boldsymbol{y})$). Next, we introduce the concept of the evidence lower bound (ELBO), which is critical for determining the optimal $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$. Let $p(\boldsymbol{y}|\mathcal{M})$ be a marginal density of the model $\mathcal{M}$; the ELBO can then be represented as

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

a lower bound of the logarithm marginal density $\log p(\boldsymbol{y}|\mathcal{M})$:

$$\begin{aligned}
\log p(\boldsymbol{y}|\mathcal{M}) &= \log \int p(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\theta})\, d\boldsymbol{\zeta} d\boldsymbol{\theta} \\
&= \log \int q(\boldsymbol{\zeta}, \boldsymbol{\theta}) \frac{p(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\theta})}{q(\boldsymbol{\zeta}, \boldsymbol{\theta})}\, d\boldsymbol{\zeta} d\boldsymbol{\theta} \\
&\geq \int q(\boldsymbol{\zeta}, \boldsymbol{\theta}) \log \frac{p(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\theta})}{q(\boldsymbol{\zeta}, \boldsymbol{\theta})}\, d\boldsymbol{\zeta} d\boldsymbol{\theta} \qquad \textbf{Jensen's inequality} \\
&= \mathrm{E}_{q(\boldsymbol{\zeta}, \boldsymbol{\theta})}[\log p(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\theta}) - \log q(\boldsymbol{\zeta}, \boldsymbol{\theta})] \\
&\triangleq \underline{\mathcal{L}}\left(q(\boldsymbol{\zeta}, \boldsymbol{\theta})\right),
\end{aligned} \tag{9}$$

where $\underline{\mathcal{L}}\left(q(\boldsymbol{\zeta}, \boldsymbol{\theta})\right)$ is denoted as the ELBO, which is a function of the free distribution $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$. We need to maximize $\underline{\mathcal{L}}\left(q(\boldsymbol{\zeta}, \boldsymbol{\theta})\right)$ with respect to $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$ so that it tends more closely to $\log p(\boldsymbol{y}|\mathcal{M})$. Blei et al.(2017) presented a formula connecting $\log p(\boldsymbol{y}|\mathcal{M})$ with the ELBO and the Kullback–Leibler (KL) divergence:

$$\log p(\boldsymbol{y}|\mathcal{M}) = \underline{\mathcal{L}}\left(q(\boldsymbol{\zeta}, \boldsymbol{\theta})\right) + \mathrm{KL}(q(\boldsymbol{\zeta}, \boldsymbol{\theta}) \| p(\boldsymbol{\zeta}, \boldsymbol{\theta}|\boldsymbol{y})). \tag{10}$$

Since $\log p(\boldsymbol{y}|\mathcal{M})$ is a constant with respect to $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$, maximizing the ELBO is actually equivalent to minimizing the KL distance. Specifically, the optimal $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$ we obtained in the variational density family $\mathcal{Q}$ is the density that minimizes the KL divergence between the posterior distribution $p(\boldsymbol{\zeta}, \boldsymbol{\theta}|\boldsymbol{y})$ and itself. To further simplify the variational density $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$, we assume that it satisfies mean-field theory. Mean-field theory has been widely used in variational Bayesian inference (Beal, 2003; Blei et al., 2017; Jordan et al., 1999; Wand et al., 2011). In the mean-field theory, latent variables are mutually independent and each is governed by a separate factor in the variational density, allowing the variational

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

density $q(\boldsymbol{\zeta}, \boldsymbol{\theta})$ to be decomposed into $q(\boldsymbol{\zeta})q(\boldsymbol{\theta})$. An iterative optimization procedure is implemented by seeking to maximize the mean-field variational density of a parameter of interest while fixing the others. The VB algorithm can be divided into the following two steps:

$$
\begin{aligned}
\textbf{VBE step:} \qquad q^{\text{new}}(\boldsymbol{\zeta}_i) &= \frac{1}{\mathcal{Z}_{\boldsymbol{\zeta}_i}} \exp\left[\int q^{\text{old}}(\boldsymbol{\theta}) \log p(\boldsymbol{\zeta}_i, \boldsymbol{y}_i|\boldsymbol{\theta})\, d\boldsymbol{\theta}\right] \\
&\propto \exp\left\{\mathrm{E}_{q^{\text{old}}(\boldsymbol{\theta})}[\log p(\boldsymbol{y}_i, \boldsymbol{\zeta}, \boldsymbol{\theta})]\right\}, \textit{for } \forall i, \\
\textbf{VBM step:} \qquad q^{\text{new}}(\boldsymbol{\theta}) &= \frac{1}{\mathcal{Z}_{\boldsymbol{\theta}}} p(\boldsymbol{\theta}) \exp\left[\int q^{\text{new}}(\boldsymbol{\zeta}) \log p(\boldsymbol{\zeta}, \boldsymbol{y}|\boldsymbol{\theta})\, d\boldsymbol{\zeta}\right] \\
&\propto \exp\left\{\mathrm{E}_{q^{\text{new}}(\boldsymbol{\zeta})}[\log p(\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\theta})]\right\},
\end{aligned}
\tag{11}
$$

where $\mathcal{Z}_{\boldsymbol{\zeta}_i}$ and $\mathcal{Z}_{\boldsymbol{\theta}}$ are the normalizing constants. To sum up, the variational density for the latent variable is updated in the VBE step, while the variational density for the model parameters is updated in the VBM step. Therefore, the prerequisite to be able to implement the VBEM algorithm is that the posterior distribution of all parameters, either latent variables or model parameters, should have a closed form. The VEM algorithm proposed by Yamaguchi and Okada (2020a, 2020b) in educational psychometric research is essentially identical to the VBEM algorithm provided by Beal (2003), with the only differences being in nomenclature.

3.2. Variational Methods in Bayesian Logistic Regression

As mentioned above, implementing the VBEM algorithm requires a closed form for the posterior distributions of each parameter. Therefore, the VBEM algorithm cannot be

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

directly applied to the LCDM based on the logit link function. To overcome this challenge, we adopt Jaakkola and Jordan's (2000) variational Bayesian method for logistic regression models to estimate the more complex LCDM in the cognitive diagnostic framework. Specifically, their method uses a Taylor expansion on the logistic function to obtain a tight lower bound, facilitating parameter representation in a Gaussian distribution form that is easily implementable for variational inference. Next, we will provide the mathematical expression that Jaakkola and Jordan (2000) used for performing the first-order Taylor expansion and the specific derivation of the tight lower bound.

Consider the logistic function $\sigma(\omega) = 1/\left(1 + \exp(-\omega)\right)$. The corresponding log logistic function can be derived as

$$\log \sigma(\omega) = -\log\left(1 + \exp(-\omega)\right) = \frac{\omega}{2} - \log\left(\exp\left(\frac{\omega}{2}\right) + \exp\left(-\frac{\omega}{2}\right)\right). \tag{12}$$

Denote that

$$f(\omega) = -\log\left(\exp\left(\frac{\omega}{2}\right) + \exp\left(-\frac{\omega}{2}\right)\right).$$

By calculating the second derivative, we can determine that $f(\omega)$ is a convex function about the variable ω^2. Therefore, any tangent line of $f(\omega)$ can serve as its lower bound, as it will always be less than or equal to $f(\omega)$. A tight lower bound function for $f(\omega)$ can be obtained by executing a first-order Taylor expansion on the function $f(\omega)$ in terms of the variable ω^2 at the point ξ^2,

$$f(\omega) \geq f(\xi) + \frac{\partial f(\xi)}{(\partial \xi^2)}(\omega^2 - \xi^2) = f(\xi) - \frac{1}{2\xi}(\sigma(\xi) - \frac{1}{2})(\omega^2 - \xi^2). \tag{13}$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

According to Eq.(12) and Eq.(13), we can derive a tight lower bound of $\sigma(\omega)$ with the specific form as

$$\sigma(\omega) \geq \sigma(\xi) \exp\left(\frac{(\omega-\xi)}{2} - \tau(\xi)(\omega^2-\xi^2)\right), \quad \tau(\xi) = \frac{1}{2\xi}\left(\sigma(\xi) - \frac{1}{2}\right), \tag{14}$$

which results in a quadratic form on ω.

Regarding to the LCDM, which also employs a logistic form, ω represents a set of linear combinations. These combinations involve unknown item parameters, an individual's latent attribute vector, and the known Q-matrix within the LCDM framework (for further details, please refer to Eqs. (1) and (2)). Based on Eq. (14), we can derive a quadratic form for the item parameter. Consequently, the VI algorithm can be implemented in the LCDM by deriving a specific posterior distribution for item parameters using the Gaussian prior distribution, which serves as the conjugate prior for these parameters. In the next subsection, we will focus on elucidating the process of deriving the tight lower bound in the LCDM using Eq. (14).

3.3. Tight Lower Bound for the LCDM

In this section, the goal is to derive the tight lower bound for LCDM as outlined above. We first conduct a transformation on the item response data in the LCDM to make it easier to acquire the tight lower bound term of the likelihood function before providing the implementation of our VBEM-M algorithm. The item response data $x_{ij} = \{0, 1\}$ is transformed into $y_{ij} = \{-1, 1\}$ with the help of the equation $y_{ij} = 2x_{ij} - 1$. Let

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

$\boldsymbol{\lambda}_j^* = (\eta_j, \boldsymbol{\lambda}_j^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{h}_{jl}^* = (1, h(\tilde{\boldsymbol{\alpha}}_l, \boldsymbol{q}_j))$; the item response probability of y_{ij} is then given by

$$\begin{aligned} P(y_{ij} = 1|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, \boldsymbol{\lambda}_j^*, \boldsymbol{q}_j) &= P(x_{ij} = 1|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, \boldsymbol{\lambda}_j^*, \boldsymbol{q}_j) = \frac{1}{1 + \exp(-\boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*)}, \\ P(y_{ij} = -1|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, \boldsymbol{\lambda}_j^*, \boldsymbol{q}_j) &= P(x_{ij} = 0|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, \boldsymbol{\lambda}_j^*, \boldsymbol{q}_j) = \frac{1}{1 + \exp(\boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*)}. \end{aligned} \tag{15}$$

Recall the logistic function form, the item response probability of y_{ij} can then be rewritten as follows,

$$p(y_{ij}|\boldsymbol{\alpha}_i = \tilde{\boldsymbol{\alpha}}_l, \boldsymbol{\lambda}_j^*, \boldsymbol{q}_j) = \sigma(y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*). \tag{16}$$

Therefore, the likelihood based on the introduced latent variable z can be represented by

$$p(\mathbf{Y}|\boldsymbol{z}, \widetilde{\mathbf{A}}, \boldsymbol{\lambda}^*, \mathbf{Q}) = \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l=1}^{L} \sigma(y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*)^{z_{il}}. \tag{17}$$

According to Eq. (14), the tight lower bound function for $\sigma(y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*)$ is determined by performing a first-order Taylor expansion with respect to the variable $(y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*)^2$ at the point ξ_{ijl}^2. Therefore, a tight lower bound of the likelihood for the LCDM can be derived by:

$$\begin{aligned} & p(\mathbf{Y}|\boldsymbol{z}, \widetilde{\mathbf{A}}, \boldsymbol{\lambda}^*, \mathbf{Q}) = \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l=1}^{L} \sigma(y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^*)^{z_{il}} \\ & \geq \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l=1}^{L} \left\{ \sigma(\xi_{ijl}) \exp\left(\frac{y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^* - \xi_{ijl}}{2} - \tau(\xi_{ijl})(\boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^* \boldsymbol{h}_{jl}^{*\mathrm{T}} \boldsymbol{\lambda}_j^* - \xi_{ijl}^2) \right) \right\}^{z_{il}} \\ & \triangleq \underline{p(\mathbf{Y}|\boldsymbol{z}, \widetilde{\mathbf{A}}, \boldsymbol{\lambda}^*, \mathbf{Q})}. \end{aligned} \tag{18}$$

Given that the tight lower bound for the likelihood function is of exponential form, using the multivariate normal distribution as a conjugate prior distribution for $\boldsymbol{\lambda}^*$ will yield a closed-form posterior distribution. Due to these considerations, in the subsequent

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

computations, we implement the VBEM-M algorithm using the tight lower bound of the likelihood function rather than the original likelihood function. Moreover, it's important to highlight that a new local parameter, ξ_{ijl}, has been introduced at this stage. Determining the optimal value for ξ_{ijl} is an essential part of our analysis. In this paper, we implement a maximization process to ascertain the most suitable value for ξ_{ijl}. The detailed methodology behind this process will be elaborated in the following subsection.

3.4. Fully Bayesian Representation of the Joint Posterior Distribution

In the fully Bayesian framework, statistical inference relies on the selection of the prior distribution. The posterior distribution can be derived by combining the prior distribution (prior information) with the likelihood function (sample information). Prior distributions from the following Bayesian hierarchical structures will be considered in this study:

$$
\begin{aligned}
& y_{ij} \sim p(y_{ij}|\boldsymbol{z}_i, \widetilde{\mathbf{A}}, \boldsymbol{\lambda}^*, \mathbf{Q}),\ p\left(\boldsymbol{z}_i|\boldsymbol{\pi}\right) = \prod_{l=1}^{L} \pi_l^{z_{il}}, 0 \leq \pi_l \leq 1, \sum_{l=1}^{L} \pi_l = 1, \\
& p(\boldsymbol{\pi}) = p\left(\pi_1, \ldots, \pi_L\right) \ = \ \text{Dirichlet}\ \left(\boldsymbol{\delta}_0\right), \boldsymbol{\delta}_0 = \left(\delta_{01}, \ldots, \delta_{0L}\right), \\
& p(\boldsymbol{\lambda}_j^*) \ = \ \text{MVN}\left(\boldsymbol{\lambda}_0^*, \mathbf{I}_{D+1}\right), \boldsymbol{\lambda}_0^* = (\eta_0, \boldsymbol{\lambda}_0), \\
& \boldsymbol{\lambda}_0 \ = \ (\underbrace{\lambda_{0,1}, \cdots, \lambda_{0,K}}_{\textbf{main effect terms}}, \underbrace{\lambda_{0,K+1}, \cdots, \lambda_{0,D}}_{\textbf{interaction terms}}) \ = \ (\underbrace{\lambda_{0,main}, \cdots, \lambda_{0,main}}_{\textbf{main effect terms}}, \underbrace{\lambda_{0,inter}, \cdots, \lambda_{0,inter}}_{\textbf{interaction terms}}), \\
& p(\eta_0) = \text{N}\left(\mu_{\eta_0}, \sigma_{\eta_0}^2\right), \\
& p(\lambda_{0,main}) \ = \ \text{N}\left(\mu_{\lambda_{0,main}}, \sigma_{\lambda_{0,main}}^2\right) \mathcal{I}\left(c, \infty\right), \\
& p(\lambda_{0,inter}) \ = \ \text{N}\left(\mu_{\lambda_{0,inter}}, \sigma_{\lambda_{0,inter}}^2\right),
\end{aligned}
\tag{19}
$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

where $\mathbf{I}_{D+1}$ is a $(D+1)$-dimensional identity matrix. Parameter c is a truncation parameter. Some literature restricts the main effect terms of $\boldsymbol{\lambda}$ to non-negative values (Zhan et al., 2019). To address this, a truncation parameter c is introduced to adjust the range of values for the prior parameter $\lambda_{0,main}$. For example, when c is set to $-\infty$, there is no restriction on $\lambda_{0,main}$, while setting $c = 0$ restricts $\lambda_{0,main}$ to non-negative values. In practice, users can adjust the value of c to restrict the range of $\lambda_{0,main}$ according to their specific requirements. Let $\boldsymbol{\Omega} = (\boldsymbol{\delta}_0, \mu_\eta, \sigma_\eta^2, \mu_\lambda, \sigma_\lambda^2)$, the joint posterior distribution of $(\mathbf{Y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^* | \mathbf{Q}, \boldsymbol{\Omega}, \widetilde{\mathbf{A}})$ based on the tight lower bound can be represented by

$$
\begin{aligned}
\underline{p(\mathbf{Y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^* | \mathbf{Q}, \boldsymbol{\Omega}, \widetilde{\mathbf{A}})} \;&=\; \underline{p(\mathbf{Y} | \boldsymbol{z}, \widetilde{\mathbf{A}}, \boldsymbol{\lambda}^*, \mathbf{Q})} p(\boldsymbol{z} | \boldsymbol{\pi}) p(\boldsymbol{\pi}) p(\boldsymbol{\lambda}^* | \boldsymbol{\lambda}_0^*) p(\boldsymbol{\lambda}_0^*) \\
&\propto \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{l=1}^{L} \left\{ \sigma(\xi_{ijl}) \exp\left(\frac{y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^* - \xi_{ijl}}{2} - \tau(\xi_{ijl}) (\boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^* \boldsymbol{h}_{jl}^{*\mathrm{T}} \boldsymbol{\lambda}_j^* - \xi_{ijl}^2) \right) \right\}^{z_{il}} \\
&\times \prod_{i=1}^{N} \prod_{l=1}^{L} \pi_l^{z_{il}} \prod_{l=1}^{L} \pi_l^{\delta_{0l}} \prod_{j=1}^{J} \exp\left\{ -\frac{(\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_0^*)^{\mathrm{T}} (\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_0^*)}{2} \right\} \exp\left\{ -\frac{(\eta_0 - \mu_{\eta_0})^2}{2\sigma_{\eta_0}^2} \right\} \\
&\times \exp\left\{ -\frac{(\lambda_{0,main} - \mu_{\lambda_{0,main}})^2}{2\sigma_{\lambda_{0,main}}^2} \right\} \mathcal{I}(\lambda_{0,main} > c) \exp\left\{ -\frac{(\lambda_{0,inter} - \mu_{\lambda_{0,inter}})^2}{2\sigma_{\lambda_{0,inter}}^2} \right\} \times const,
\end{aligned}
\tag{20}
$$

where *const* denotes a constant. The logarithm of $\underline{p(\mathbf{Y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^* | \mathbf{Q}, \boldsymbol{\Omega}, \widetilde{\mathbf{A}})}$ can be further expressed as

$$
\begin{aligned}
&\log \underline{p(\mathbf{Y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^* | \mathbf{Q}, \boldsymbol{\Omega}, \widetilde{\mathbf{A}})} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{l=1}^{L} z_{il} \left\{ \log(\sigma(\xi_{ijl})) + \frac{y_{ij} \boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^* - \xi_{ijl}}{2} - \tau(\xi_{ijl}) (\boldsymbol{\lambda}_j^{*\mathrm{T}} \boldsymbol{h}_{jl}^* \boldsymbol{h}_{jl}^{*\mathrm{T}} \boldsymbol{\lambda}_j^* - \xi_{ijl}^2) \right\} \\
&+ \sum_{i=1}^{N} \sum_{l=1}^{L} z_{il} \log \pi_l + \sum_{l=1}^{L} \delta_{0l} \log \pi_l - \sum_{j=1}^{J} \frac{(\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_0^*)^{\mathrm{T}} (\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_0^*)}{2} \\
&- \frac{(\eta_0 - \mu_{\eta_0})^2}{2\sigma_{\eta_0}^2} - \frac{(\lambda_{0,main} - \mu_{\lambda_{0,main}})^2}{2\sigma_{\lambda_{0,main}}^2} \mathcal{I}(\lambda_{0,main} > c) - \frac{(\lambda_{0,inter} - \mu_{\lambda_{0,inter}})^2}{2\sigma_{\lambda_{0,inter}}^2} + const.
\end{aligned}
\tag{21}
$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

3.5. Implementation of VBEM-M Algorithm for LCDM

Assuming that the joint variational density of $(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^*)$ for the LCDM satisfies mean-field theory, the following equation holds:

$$q(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^*) = \left(\prod_{i=1}^{N} q(\boldsymbol{z})\right)\left(q(\boldsymbol{\pi})\prod_{j=1}^{J} q(\boldsymbol{\lambda}_j^*)q(\eta_0)\prod_{d=1}^{D} q(\lambda_{0d})\right). \tag{22}$$

Let $\Theta = (\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^*)$; in terms of Eq. (9) and Eq. (21), the ELBO $\underline{\mathcal{L}}(q(\Theta))$ can then be derived as

$$\begin{aligned}
&\underline{\mathcal{L}}(q(\Theta)) = \mathrm{E}_{q(\boldsymbol{z},\boldsymbol{\pi},\boldsymbol{\lambda}^*,\boldsymbol{\lambda}_0^*)}[\log p(\mathbf{Y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^* | \mathbf{Q}, \boldsymbol{\Omega}, \widetilde{\mathbf{A}}) - \log q(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^*)] \\
&\geq \mathrm{E}_{q(\boldsymbol{z},\boldsymbol{\pi},\boldsymbol{\lambda}^*,\boldsymbol{\lambda}_0^*)}[\log \underline{p(\mathbf{Y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^* | \mathbf{Q}, \boldsymbol{\Omega}, \widetilde{\mathbf{A}})} - \log q(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_0^*)] \\
&= \mathrm{E}_{q(\boldsymbol{z},\boldsymbol{\pi},\boldsymbol{\lambda}^*,\boldsymbol{\lambda}_0^*)}\Bigg[\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L} z_{il}\left\{\log(\sigma(\xi_{ijl})) + \frac{y_{ij}\boldsymbol{\lambda}_j^{*\mathrm{T}}\boldsymbol{h}_{jl}^* - \xi_{ijl}}{2} - \tau(\xi_{ijl})(\boldsymbol{\lambda}_j^{*\mathrm{T}}\boldsymbol{h}_{jl}^*\boldsymbol{h}_{jl}^{*\mathrm{T}}\boldsymbol{\lambda}_j^* - \xi_{ijl}^2)\right\} \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\log\pi_l + \sum_{l=1}^{L}(\delta_{0l}-1)\log\pi_l - \sum_{j=1}^{J}\frac{(\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_0^*)^{\mathrm{T}}(\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_0^*)}{2} \\
&- \frac{(\eta_0 - \mu_{\eta_0})^2}{2\sigma_{\eta_0}^2} - \frac{(\lambda_{0,main} - \mu_{\lambda_{0,main}})^2}{2\sigma_{\lambda_{0,main}}^2}\mathcal{I}(\lambda_{0,main} > c) - \frac{(\lambda_{0,inter} - \mu_{\lambda_{0,inter}})^2}{2\sigma_{\lambda_{0,inter}}^2} \\
&- \sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\log\pi_{il}^* - \sum_{l=1}^{L}(\delta_l^* - 1)\log\pi_l - \sum_{j=1}^{J}\frac{(\boldsymbol{\lambda}_j^* - \boldsymbol{m}_j^*)^{\mathrm{T}}\boldsymbol{V}_j^{*-1}(\boldsymbol{\lambda}_j^* - \boldsymbol{m}_j^*)}{2} \\
&+ \frac{(\eta_0 - \mu_{\eta_0}^*)^2}{2\sigma_{\eta_0}^{*2}} + \frac{(\lambda_{0,main} - \mu_{\lambda_{0,main}}^*)^2}{2\sigma_{\lambda_{0,main}}^{*2}}\mathcal{I}(\lambda_{0,main} > c) - \frac{(\lambda_{0,inter} - \mu_{\lambda_{0,inter}}^*)^2}{2\sigma_{\lambda_{0,inter}}^{*2}}\Bigg] + const \\
&\triangleq \underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi}),
\end{aligned} \tag{23}$$

where $\underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi})$ is a tight lower bound of $\underline{\mathcal{L}}(q(\Theta))$. Next, we maximize $\underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi})$ to obtain estimates of latent variables $\boldsymbol{z}$, model parameters $(\boldsymbol{\pi}, \boldsymbol{\lambda}^*)$, hyperparameters $\boldsymbol{\lambda}_0^*$ and local point parameter $\boldsymbol{\xi}$. Specifically, there are three steps to the implementation process:

(a) VBE step: update variational density for latent variable;

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

(b) VBM step: update variational densities for model parameters and hyperparameters;

(c) M step: update local point parameter $\boldsymbol{\xi}$ by maximizing $\underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi})$.

In the following text, $q^*(\cdot)$ denotes the optimal variational posterior in each iteration. To keep things simple, we only present the core formulation for updating. The specifics can be found in the online supplemental materials. The estimation procedure of the VBEM-M algorithm is shown in Table 1. In Table 1 and subsequent tables, all parameters are estimated using their posterior means. In addition, the specific implementation process of each step for the VBEM-M algorithm is shown in Figure 1.

==========

Insert Table 1 about here

==========

==========

Insert Figure 1 about here

==========

(a) VBE step. In this step, we update the variational density of $\boldsymbol{z}_i$ for each i, where $i = 1, \cdots, N$. $q^*(\boldsymbol{z}_i)$ is derived to be a categorical distribution with parameter $\boldsymbol{\pi}_i^*$. That is,

$$q^*(\boldsymbol{z}_i|\boldsymbol{\pi}_i^*) = \prod_{l=1}^{L} \pi_{il}^{*z_{il}}, \tag{24}$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

where

$$\pi_{il}^* = \frac{\rho_{il}}{\sum_{l=1}^{L} \rho_{il}},$$
$$\rho_{il} = \exp\left\{ \sum_{j=1}^{J} \left\{ \log(\sigma(\xi_{ijl})) + \frac{y_{ij}\mathrm{E}_{q(\boldsymbol{\lambda}_j^*)}[\boldsymbol{\lambda}_j^*]^{\mathrm{T}}\boldsymbol{h}_{jl}^* - \xi_{ijl}}{2} \right.\right. \tag{25}$$
$$\left.\left. - \tau(\xi_{ijl})(\boldsymbol{h}_{jl}^{*\mathrm{T}}\mathrm{E}_{q(\boldsymbol{\lambda}_j^*)}[\boldsymbol{\lambda}_j^*\boldsymbol{\lambda}_j^{*\mathrm{T}}]\boldsymbol{h}_{jl}^* - \xi_{ijl}^2) \right\} + \mathrm{E}_{q(\boldsymbol{\pi})}\log(\pi_l) \right\}.$$

(b) VBM step. In this step, we update the variational density for $\boldsymbol{\pi}$, $\boldsymbol{\lambda}_j^*$ $(j = 1, \cdots, J)$, η_0, $\lambda_{0,main}$ and $\lambda_{0,inter}$.

(b1) Update the variational density for $\boldsymbol{\pi}$

$q^*(\boldsymbol{\pi})$ is derived to be a Dirichlet distribution with parameter $\boldsymbol{\delta}^*$. That is,

$$q^*(\boldsymbol{\pi}|\boldsymbol{\delta}^*) \propto \prod_{l=1}^{L} \pi_l^{\delta_l^* - 1}, \tag{26}$$

where

$$\delta_l^* = \sum_{i=1}^{N} \mathrm{E}_{q(\boldsymbol{z}_i)}[z_{il}] + \delta_{0l}. \tag{27}$$

(b2) Update the variational density for $\boldsymbol{\lambda}_j^*$

$q(\boldsymbol{\lambda}_j^*)$ is proportional to a multivariate normal distribution with mean vector $\boldsymbol{m}_j^*$ and covariance $\boldsymbol{V}_j^*$. That is,

$$q^*(\boldsymbol{\lambda}_j^*|\boldsymbol{m}_j^*, \boldsymbol{V}_j^*) \propto \exp\left\{ -\frac{(\boldsymbol{\lambda}_j^* - \boldsymbol{m}_j^*)^{\mathrm{T}}\boldsymbol{V}_j^{*-1}(\boldsymbol{\lambda}_j^* - \boldsymbol{m}_j^*)}{2} \right\}, \tag{28}$$

where

$$\begin{aligned} \boldsymbol{V}_j^{*-1} &= \mathbf{I}_{D+1}^{-1} + 2\sum_{i=1}^{N}\sum_{l=1}^{L} \mathrm{E}_{q(\boldsymbol{z}_i)}[z_{il}]\tau(\xi_{ijl})\boldsymbol{h}_{jl}^*\boldsymbol{h}_{jl}^{*\mathrm{T}}, \\ \boldsymbol{m}_j^* &= \boldsymbol{V}_j^* \left(\boldsymbol{\lambda}_0^* + \frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{L} \mathrm{E}_{q(\boldsymbol{z}_i)}[z_{il}]y_{ij}\boldsymbol{h}_{jl}^* \right). \end{aligned} \tag{29}$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

(b3) Update the variational density for η_0

$q^*(\eta_0)$ is proportional to a normal distribution with mean $\mu^*_{\eta_0}$ and variance $\sigma^{*2}_{\eta_0}$. That is,

$$q^*(\eta_0|\mu^*_{\eta_0}, \sigma^{*2}_{\eta_0}) \propto \exp\left\{-\frac{(\eta_0 - \mu^*_{\eta_0})^2}{2\sigma^{*2}_{\eta_0}}\right\}, \tag{30}$$

where

$$\begin{aligned}(\sigma^{*2}_{\eta_0})^{-1} &= J + \frac{1}{\sigma^2_{\eta_0}},\\ \mu^*_{\eta_0} &= \sigma^{*2}_{\eta_0}\left(\frac{\mu_{\eta_0}}{\sigma^2_{\eta_0}} + \sum_{j=1}^{J}\left(\mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j]\right)_{\eta}\right),\end{aligned} \tag{31}$$

where $\mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j]_\eta$ is the corresponding expected value of the element η_j in the vector $\boldsymbol{\lambda}^*$.

(b4) Update the variational density for $\lambda_{0,main}$

$q^*(\lambda_{0,main})$ is proportional to a truncated normal distribution with mean $\mu^*_{\lambda_{0,main}}$ and variance $\sigma^{*2}_{\lambda_{0,main}}$. Specifically,,

$$q^*(\lambda_{0,main}|\mu^*_{\lambda_{0,main}}, \sigma^{*2}_{\lambda_{0,main}}) \propto \exp\left\{-\frac{(\lambda_{0,main} - \mu^*_{\lambda_{0,main}})^2}{2\sigma^{*2}_{\lambda_{0,main}}}\right\}\mathcal{I}\left(c, \infty\right), \tag{32}$$

where

$$\begin{aligned}(\sigma^{*2}_{\lambda_{0,main}})^{-1} &= J^*_{main} + \frac{1}{\sigma^2_{\lambda_{0,main}}},\\ \mu^*_{\lambda_{0,main}} &= \sigma^{*2}_{\lambda_{0,main}}\left(\frac{\mu_{\lambda_{0,main}}}{\sigma^2_{\lambda_{0,main}}} + \sum_{d=1}^{K}\sum_{j\in J_d}\left(\mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j]\right)_{\lambda_d}\right),\end{aligned} \tag{33}$$

where J^*_{main} denotes the number of all main effect terms, $J_d = \{j : \lambda_{jd} \neq 0\}$, and $\mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j]_{\lambda_d}$ is the corresponding expected value of the element λ_{jd} in the vector $\boldsymbol{\lambda}^*$.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

(b5) Update the variational density for $\lambda_{0,inter}$

$q^*(\lambda_{0,inter})$ is proportional to a truncated normal distribution with mean $\mu^*_{\lambda_{0,inter}}$ and variance $\sigma^{*2}_{\lambda_{0,inter}}$. Specifically,

$$q^*(\lambda_{0,inter}|\mu^*_{\lambda_{0,inter}}, \sigma^{*2}_{\lambda_{0,inter}}) \propto \exp\left\{-\frac{(\lambda_{0,inter} - \mu^*_{\lambda_{0,inter}})^2}{2\sigma^{*2}_{\lambda_{0,inter}}}\right\}, \tag{34}$$

where

$$\begin{aligned}
(\sigma^{*2}_{\lambda_{0,inter}})^{-1} &= J^*_{inter} + \frac{1}{\sigma^2_{\lambda_{0,inter}}}, \\
\mu^*_{\lambda_{0,inter}} &= \sigma^{*2}_{\lambda_{0,inter}} \left(\frac{\mu_{\lambda_{0,inter}}}{\sigma^2_{\lambda_{0,inter}}} + \sum_{d=K+1}^{D} \sum_{j \in J_d} \left(\mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j]\right)_{\lambda_d}\right),
\end{aligned} \tag{35}$$

where J^*_{inter} denotes the number of all interaction terms.

(c) M step. In this step, we update the local point parameter ξ_{ijl} $(i = 1, \cdots, N; j = 1, \cdots, J; l = 1, \cdots, L)$. To obtain the optimal ξ_{ijl}, we need to maximize $\underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi})$ by computing the derivative of ξ_{ijl} to zero:

$$\frac{\partial \underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi})}{\partial \xi_{ijl}} = 0. \tag{36}$$

Therefore, we have

$$\xi^2_{ijl} = \xi^2_{jl} = \boldsymbol{h}^{*\mathrm{T}}_{jl} \mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j \boldsymbol{\lambda}^{*\mathrm{T}}_j] \boldsymbol{h}^*_{jl}. \tag{37}$$

Considering the aforementioned presentation of the VBEM-M algorithm, it is clear that we need to compute a large number of expectations using categorical, Dirichlet, normal, multivariate normal, and truncated normal distributions. Some formulae for calculating

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

these expectations are as follows:

$$
\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{z}_i)}[z_{il}] = \pi^*_{il}, \quad \mathrm{E}_{q(\boldsymbol{\pi})} \log(\pi_l) = \psi(\delta^*_l) - \psi\left(\sum_{l=1}^{L} \delta^*_l\right), \\
&\mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j] = \boldsymbol{\mu}^*_j, \quad \mathrm{E}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j \boldsymbol{\lambda}^{*\mathrm{T}}_j] = \boldsymbol{\Sigma}^*_j + \boldsymbol{\mu}^*_j \boldsymbol{\mu}^{*\mathrm{T}}_j, \\
&\mathrm{E}_{q(\eta_0)}[\eta_0] = \mu^*_{\eta_0}, \quad \mathrm{E}_{q(\eta_0)}[(\eta_0 - \mu^*_{\eta_0})^2] = \sigma^{*2}_{\eta_0}, \\
&\mathrm{E}_{q(\lambda_{0,main})}[\lambda_{0,main}] = \mu^*_{\lambda_{0,main}} + \sigma^{*2}_{\lambda_{0,main}} \frac{\phi(u)}{\Phi(u)}, u = \frac{c - \mu^*_{\lambda_{0,main}}}{\sigma^*_{\lambda_{0,main}}}, \\
&\mathrm{E}_{q(\lambda_{0,main})}[(\lambda_{0,main} - \mu^*_{\lambda_{0,main}})^2] = \sigma^{*2}_{\lambda_{0,main}} \left(1 - \mu^*_{\lambda_{0,main}} \frac{\phi(u)}{\Phi(u)}\right), \\
&\mathrm{E}_{q(\lambda_{0,inter})}[\lambda_{0,inter}] = \mu^*_{\lambda_{0,inter}}, \quad \mathrm{E}_{q(\lambda_{0,inter})}[(\lambda_{0,inter} - \mu^*_{\lambda_{0,inter}})^2] = \sigma^{*2}_{\lambda_{0,inter}},
\end{aligned}
\tag{38}
$$

where $\psi(\cdot)$ is $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x)$, $\Gamma(x) = \int_0^\infty t^{(x-1)} \exp(-t)\, dt$, $\phi(\cdot)$ is the density function of a standard normal distribution, and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution.

4. Simulation Study

In the following simulation studies, we address three primary concerns: First, the performance of the VBEM-M algorithm under various conditions for the DINA model; second, the performance of the VBEM-M algorithm, based on the DINA model, compares to Yamaguchi and Okada's (2020b) VB method, the MCMC algorithms within the full Bayesian framework, and the EM algorithm in the frequency framework under different simulation settings; third, the performance of the VBEM-M algorithm is compared with the VB, MCMC, and EM algorithms under the saturated LCDM with different simulation conditions. Online supplement showcases the performance of the VBEM-M algorithm for

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

the DINA model under different initial values and in other widely used CDMs, including the DINO model and LLM.

Data generation. Item response data x_{ij} is generated from a Bernoulli distribution with probability of correct response $P(x_{ij} = 1|\boldsymbol{\alpha}_i, \boldsymbol{\lambda}^*, \boldsymbol{q}_j)$. The true values of the item parameters based on DINA model are constrained by considering four different levels of noise to investigate the correlation between noise and recovery. For each item, the following scenarios are considered. (a1) Low noise level (LNL): $s_j = g_j = 0.1$, with corresponding true values $\eta_j = -2.1972$, $\lambda_j = 4.3944$. (a2) High noise level (HNL): $s_j = g_j = 0.2$, with corresponding true values $\eta_j = -1.3863$, $\lambda_j = 2.7726$. (a3) Slipping higher than guessing (SHG): $s_j = 0.2$, $g_j = 0.1$, with corresponding true values $\eta_j = -2.1972$, $\lambda_j = 3.5835$. (a4) Guessing higher than slipping (GHS): $s_j = 0.1$, $g_j = 0.2$, with corresponding true values $\eta_j = -1.3863$, $\lambda_j = 3.5835$.

To generate the attribute-mastery patterns, we used the same procedure as Chiu and Douglas (2013), which takes into account the correlations among the attributes. Specifically, $\boldsymbol{\alpha}_i^* = (\alpha_{i1}^*, \cdots, \alpha_{ik}^*, \cdots, \alpha_{iK}^*)^{\mathrm{T}}$ are generated from a multivariate normal distribution; that is, $\boldsymbol{\alpha}_i^* \sim \mathrm{N}(\mathbf{0}_K, \boldsymbol{\Sigma}_{K\times K})$, where $\mathbf{0_K} = (0, \ldots, 0)_{K\times 1}^{\mathrm{T}}$ and

$$\boldsymbol{\Sigma_{K\times K}} = \begin{bmatrix} 1 & \ldots & \sigma \\ \vdots & \ddots & \vdots \\ \sigma & \ldots & 1 \end{bmatrix}_{K\times K},$$

where the off-diagonal elements of $\Sigma_{K\times K}$ are σ. As σ increases from 0 to 1, the correlation

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

between attributes also increases from 0 to maximum. The relationships between the attribute profiles $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_i^*$ can be expressed as $\alpha_{ik} = 1$ if $\alpha_{ik}^* > 0$ and $\alpha_{ik} = 0$ otherwise. Although the Q-matrices are created randomly, they still conform to the identifiability constraints outlined by Chen et al. (2015, 2017), Liu and Andersson (2020), and Xu and Shang (2018). We present the Q-matrices used in these simulations in online supplement.

Prior distributions. The prior parameter $\boldsymbol{\delta}_0$ is set as $\boldsymbol{\delta}_0 = \mathbf{1}_L$ (Culpepper, 2015; Zhan et al., 2019), where $\mathbf{1}_L$ denotes a L-dimensional vector with all elements equal to 1. The hyperparameters are chosen as follows: $\mu_{\eta_0} = -2$, $\mu_{\lambda_{main}} = \mu_{\lambda_{inter}} = 0$, and $\sigma^2_{\eta_0} = \sigma^2_{\lambda_{main}} = \sigma^2_{\lambda_{inter}} = 10$.

Estimation software. We implemented four different approaches, namely, the VBEM-M algorithm, VB algorithm, MCMC sampling algorithm, and EM algorithm, using the R programming language (R Core Team, 2017) on a desktop computer equipped with Intel (R) Core (TM) i5-10400 CPU @ 2.90GHz, 16GB RAM. To enhance the computational efficiency of the VBEM-M method, we utilized two R packages, "Rcpp" (Eddelbuettel & Francois, 2011) and "RcppArmadillo" (Eddelbuettel & Sanderson, 2014), to call the C++ programming language. The R code of our VBEM-M algorithm can be found in the online supplementary materials. We used the R package "variationalDCM" (Hijikata et al., 2023) to implement the VB method. The MCMC sampling algorithms were implemented separately using the R packages "dina" (Culpepper & Balamuta, 2019) which is integrated with the C++ program, and "R2jags" (Su & Yajima, 2015) which is

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

associated with the JAGS program (Plummer, 2003). The EM algorithm was implemented using the R packages "GDINA" (Ma & de la Torre, 2020) and "CDM" (George et al., 2016), respectively.

Convergence diagnosis. The VBEM-M algorithm was considered converged if the absolute difference between two consecutive iterations was less than $e_0 = 10^{-4}$, or if the number of iterations T had reached 2000. When using the R packages "dina" and "R2jags" to implement the MCMC sampling algorithms, for the DINA model, Culpepper (2015) demonstrated that it would have converged after 750 iterations, thus the chain length was set to 2000 and the first 1000 iterations were set as a 'burn-in' period. For the saturated LCDM, we chose a chain length of 10000, with a burn-in of 5000. For the EM algorithm, when employing the R package "GDINA" , the convergence criteria is when the maximum absolute change in item success probabilities between consecutive iterations was smaller than $e_0 = 10^{-4}$ or when T exceeded 2000. In addition, when using the R package "CDM", iteration will end if the maximal change in parameter estimates is below $e_0 = 0.001$.

Evaluation Criteria. For item parameters and class membership probability parameters, we assess the accuracy of parameter estimation using bias and RMSE (Root Mean Square Error). For attribute parameters, we adopt the following two evaluation indices: the pattern-wise agreement rate (PAR), which indicates the rates of correct classification for attribute patterns, and is formulated as

$$\text{PAR} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{I}(\hat{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i), \tag{39}$$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

and the attribute-wise agreement rate (AAR), which signifies the rates of correct classification for individual attributes, and is defined as

$$\text{AAR}(k) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{I}(\hat{\alpha}_{ik} = \alpha_{ik}), \tag{40}$$

where $\boldsymbol{\alpha}_i$ is the true value of the ith student's attribute profile and $\hat{\boldsymbol{\alpha}}_i$ is the estimated value of $\boldsymbol{\alpha}_i$. $\hat{\alpha}_{ik}$ is the estimated value of α_{ik} for the specific attribute k.

4.1. Simulation Study 1

In this simulation study, we explored the performance of the VBEM-M algorithm under various simulation conditions. We set the test length to $J = 30$, the number of attributes was set to $K = 5$, and the corresponding Q-matrix is shown in the online supplemental materials. The following manipulated conditions were considered: (A) number of examinees $N = 1000$ and 2000; (B) correlation among attributes $\sigma = 0$, 0.3 and 0.7; and (C) noise levels LNL, HNL, SHG, and GHS. Fully crossing different levels of these three factors yields 24 conditions (2 sample sizes $\times$ 3 correlations $\times$ 4 noise levels). There were 100 replications for each simulation condition. The recovery results of parameters are displayed in Tables 2 and 3 and Figure 2.

==========

Insert Table 2 about here

==========

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

==========

Insert Table 3 about here

==========

==========

Insert Figure 2 about here

==========

The following conclusions can be drawn from Tables 2 and 3. (1) Given the correlation and noise levels, when the number of examinees is increased from 1000 to 2000, the average RMSE, the average bias, and standard deviation (SD) for $\boldsymbol{\eta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\pi}$ show decreasing trends. For example, when the correlation among attributes is 0.3 and the LNL is applied, increasing the number of examinees from 1000 to 2000 results in the average bias of $\boldsymbol{\eta}$ decreasing from -0.0140 to -0.0077, and the average bias of $\boldsymbol{\lambda}$ decreasing from 0.0307 to 0.0133. The average RMSE of $\boldsymbol{\eta}$ decreases from 0.1369 to 0.0981, the average RMSE of $\boldsymbol{\lambda}$ from 0.2337 to 0.1669, and the average RMSE of $\boldsymbol{\pi}$ from 0.0022 to 0.0016. The SD of $\boldsymbol{\eta}$ decreases from 0.0937 to 0.0664, the SD of $\boldsymbol{\lambda}$ decreases from 0.1617 to 0.1152, and the SD of $\boldsymbol{\pi}$ decreases from 0.0051 to 0.0037. (2) When the number of examinees and the noise level are given, with increasing σ, the average RMSE for $\boldsymbol{\eta}$ increase somewhat. This indicates that $\boldsymbol{\eta}$ is less impacted by the correlation between attributes. $\boldsymbol{\lambda}$ is substantially more impacted by σ; specifically, the average bias and RMSE for $\boldsymbol{\lambda}$ tend to decrease markedly as $\boldsymbol{\sigma}$ increases. In the meanwhile, RMSE for $\boldsymbol{\pi}$ also tend to decrease as $\boldsymbol{\sigma}$

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

increases. For example, when the number of examinees is fixed at 1000 and the LNL noise level is applied, the average bias are –0.0118, 0.140, 0.104 respectively, and the average RMSE rises from 0.1351 to 0.1388 when σ increases. The change in bias and RMSE of $\boldsymbol{\eta}$ are found to be slight. However, the decreases in bias and RMSE are markedly greater for $\boldsymbol{\lambda}$, with the average bias of $\boldsymbol{\lambda}$ decreasing from 0.0365 to 0.0279 and the corresponding average RMSE decreasing from 0.2560 to 0.2216. For $\boldsymbol{\pi}$, the average bias remains at 0.0000 in all conditions, while the average RMSE exhibits the largest change in the HNL condition, decreasing from 0.0058 to 0.0048. (3) The accuracy of attribute profile recovery is highest under the LNL condition because the noise is the lowest. For example, with a fixed number of examinees at 1000 and a correlation of $\sigma = 0$, the PAR is 0.9025 under the LNL condition and only 0.6736 under the HNL condition. Under the LNL condition, the AAR values for five attributes exceed 0.9667 across various sample sizes and levels of attribute correlation. Moreover, the accuracy of attribute profile recovery tends to improve as σ increases.

In Figure 2, as an explanation, we only show the recovery results for the LNL and HNL based on the sample size $N = 1000$. On each item, the bias of $\boldsymbol{\eta}$ are almost the same for the LNL and the HNL. Furthermore, when the correlation between attributes is strengthened (σ from 0 to 0.7), there is no difference between the bias and RMSE of $\boldsymbol{\eta}$ in the LNL (HNL). It was also discovered that, for both low and high levels of noise, the RMSE of $\boldsymbol{\eta}$ is lower when the items evaluate more attributes. At low noise levels, for

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

instance, the RMSE of $\boldsymbol{\eta}$ for the first item evaluating one attribute is greater than that for the eleventh item evaluating the first three attributes together. For $\boldsymbol{\lambda}$, although the bias of $\boldsymbol{\lambda}$ differs on each item at low and high noise levels, the values of bias are basically around 0. Similarly, for both low and high levels of noise, the RMSE of $\boldsymbol{\lambda}$ is lower when items have higher correlation amongst themselves. This is because as the attribute correlation increases, more accurate estimates of α are obtained, which in turn enhances the accuracy of λ estimates. This also provides an empirical guarantee for our later practical research. That is, when designing the items, we should aim to achieve higher correlations between attributes to increase the accuracy of parameter estimation.

Additionally, we assess the performance of the VBEM-M algorithm under different initial values (please see online supplement for details), and the results showed that our VBEM-M algorithm is not affected by the different initial values.

4.2. Simulation Study 2

The purpose of this simulation study is to compare the proposed method with Yamaguchi and Okada's (2020b) VB method, the MCMC sampling algorithms, and the EM algorithm in terms of parameter accuracy for the DINA model. Specifically, the R package "variationalDCM" was used to implement Yamaguchi and Okada's (2020b) VB method, while the R packages "dina" and "R2jags" were used to implement the MCMC sampling algorithms. The EM algorithm was implemented using the R packages "GDINA"

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

and "CDM".

The simulation design is as follows: the test length was fixed at $J = 30$, and the number of attributes was set to $K = 5$. The varying conditions of the simulation are as follows: (D) The number of examinees $N = 200$, 500, 1000, and 2000; (E) correlation among attributes $\sigma = 0$, 0.3, and 0.7; and (F) LNL and HNL conditions. Fully crossing different levels of these two factors yields 24 conditions (4 sample sizes $\times$ 3 correlations $\times$ 2 noise levels). Each simulation condition was replicated 100 times. The recovery results of item parameters and attribute profile recovery for all six methods are shown in Tables 4 and 5. Due to the space limit, we only present the results with the correlation $\sigma = 0.3$ in Tables 4 and 5; the other two correlation cases ($\sigma = 0$ and $\sigma = 0.7$) are given in online supplement. Figure 3 depicts the boxplots of the bias and RMSE for $\boldsymbol{\eta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\pi}$ estimated by the six methods with $\sigma = 0.3$ under the LNL condition. Table 6 shows the computation time for these six methods under the same conditions. Here, the displayed computation time is the average time across 100 replications.

==========

Insert Table 4 about here

==========

==========

Insert Table 5 about here

==========

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

==========

Insert Table 6 about here

==========

In Tables 4 and 5, as well as in the subsequent simulation studies, the RMSE and bias mentioned are the average RMSE and average bias. From Tables 4 and 5, we can draw the following conclusions: (1) The VBEM-M algorithm consistently outperforms the other five methods in terms of achieving lower RMSE values for item parameters $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ under all four sample sizes, regardless of LNL or HNL condition. (2) For the EM algorithm, both EM-GDINA and EM-CDM methods have higher bias and RMSE for item parameters $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ than four other methods, especially for a small sample size of N=200, under both LNL and HNL conditions. (3) With the same sample size and noise level, both MCMC methods (MCMC-dina and MCMC-R2jags) show similar estimation accuracy, as do the two EM methods (EM-GDINA and EM-CDM). (4) For parameter $\boldsymbol{\pi}$, the estimated bias and RMSE of the six methods are basically the same under various identical simulation conditions, with no significant differences. (5) In terms of the accuracy of attribute profile recovery, the results of the six methods are essentially the same under each simulation condition.

From Table 6, we can see that the VBEM-M algorithm is highly efficient in terms of computation time. It performs faster than the VB method across most simulation conditions, and this speed advantage is more noticeable as sample sizes increase. Overall, the computation speed of the VBEM-M algorithm is second only to the two EM

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

algorithms, i.e., EM-GDINA and EM-CDM. The two Bayesian methods, MCMC-dina and MCMC-R2jags, have longer computation time than the other four methods. Additionally, MCMC-dina is faster than MCMC-R2jags due to its use of the "Rcpp" and "RcppArmadillo" packages, which are built on C++ programming language.

4.3. Simulation Study 3

This simulation study aims to evaluate the effectiveness of the VBEM-M algorithm on the saturated LCDM by comparing it with Yamaguchi and Okada's (2020a) VB method, the MCMC sampling algorithms, and the EM algorithm. Specifically, the R package "variationalDCM" was used to implement Yamaguchi and Okada's (2020a) VB method, the R package "R2jags" was used to implement the MCMC sampling algorithms, and the EM algorithm was implemented using the R package "GDINA".

This simulation was designed with an attribute number of $K = 3$ and a test length of $J = 18$. In the saturated LCDM, each item's $\boldsymbol{\lambda}_j^*$ is an 8-dimensional vector ($2^3 = 8$). The true values of $\boldsymbol{\lambda}^*$ are shown in Table 7. We conducted simulations across different sample sizes (N=1000, 2000) and attribute correlations ($\sigma = 0, 0.3, 0.7$), resulting in six different conditions. Each condition was replicated 100 times. Notably, an additional calculation procedure was needed for Yamaguchi and Okada's (2020a) VB method, as the R package "variationalDCM" only reports the correct response probabilities for different attribute mastery patterns. We transformed these probabilities into LCDM parameters by solving a

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

linear system of equations (Liu & Johnson, 2019; Yamaguchi & Templin, 2022). The parameter recovery results for the $\sigma = 0.3$ condition are displayed in Tables 8 and 9. The estimation results for $\sigma = 0$ and 0.7 are available in the online supplementary materials.

==========

Insert Table 7 about here

==========

==========

Insert Table 8 about here

==========

==========

Insert Table 9 about here

==========

==========

Insert Table 10 about here

==========

For a more detailed analysis, we split the parameter $\boldsymbol{\lambda}$ into two parts: $\boldsymbol{\lambda}_{main}$ (i.e. $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3$) and $\boldsymbol{\lambda}_{inter}$ (i.e. $\boldsymbol{\lambda}_{12}, \boldsymbol{\lambda}_{13}, \boldsymbol{\lambda}_{23}, \boldsymbol{\lambda}_{123}$), which represent the main effects and interactions, respectively. From the results, we can draw the following conclusions: (1) As the number of examinees increases, the RMSE for item parameters of all algorithms

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

decreases. (2) The proposed VBEM-M algorithm performs better than other algorithms on all item parameters across all conditions, especially on the interactions. Specifically, in terms of the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\pi}$, VBEM-M has a slight advantage over the other algorithms, whereas it shows a significant advantage in estimating $\boldsymbol{\lambda}_{main}$ and $\boldsymbol{\lambda}_{inter}$, particularly for the parameter $\boldsymbol{\lambda}_{inter}$. On the other hand, the EM algorithm performs poorly with small sample sizes. For the $\boldsymbol{\lambda}_{inter}$ parameter, its RMSE exceeds 2 when $N = 200$. (3) Compared to other algorithms, VBEM-M performs significantly better with small sample sizes ($N = 200, 500$), with noticeably lower RMSE. (4) It is worth noting that the results from all algorithms indicate that, although the interaction terms have smaller true values compared to the main effects, their estimation accuracy is worse. This suggests that estimating interaction effects is the most challenging aspect of the saturated LCDM model. (5) As for the accuracy of attribute profiles, there is no obvious difference among these algorithms, but VBEM-M still shows slightly higher accuracy than the others.

Table 10 shows the average computation time across 100 replications for the four algorithms under the $\sigma = 0.3$ condition. The results indicate that our algorithm performs better than the other algorithms in terms of computational efficiency. Additionally, an interesting observation is that the EM algorithm takes the longest time when the sample size is small ($N = 200$). This suggests that the EM algorithm converges more slowly with smaller sample sizes.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

5. Empirical Example

5.1. Empirical Example 1

In this example, a fraction subtraction test dataset (Tatsuoka, 1990, Tatsuoka, 2002; de la Torre & Douglas, 2004) was investigated using the DINA model. The VBEM-M algorithm, VB algorithm (implemented in the "variationalDCM" package), MCMC sampling technique (implemented in the "dina" package), and EM algorithm (implemented in the "GDINA" package) were used for the parameter estimation of the DINA model. This test involves 2144 middle school students responding to 15 fraction subtraction items, including five measured attributes: subtract basic fractions, reduce and simplify, separate whole from fraction, borrow from whole, and convert whole to fraction; 536 of 2144 students were chosen for this study (Zhang et al., 2020). The corresponding Q-matrix, parameter estimates, and SDs are shown in Table 11.

To facilitate the following item analysis, we transformed the estimates of the intercept and interaction parameters into the traditional estimates of slipping and guessing parameters, as shown in Table 11. Additionally, the comparison of the parameter estimates among the four algorithms can be found in the supplementary materials. Based on Table 11, we found that the estimates of the five items with the lowest slipping are items 3, 8, 9, 10, and 7, in that order. The estimated values of the slipping parameters for the five items are 0.0395, 0.0480, 0.0652, 0.0664, and 0.0773, respectively. This demonstrates that the five items are less likely to slip than the other ten items. Furthermore, the five items

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

with the highest guessing are items 2, 10, 8, 5, and 13, in that order. For these five items, the estimated guessing parameters are 0.2035, 0.1658, 0.1417, 0.1307, and 0.1293, respectively. Moreover, items 3, 8, and 10 have low slipping parameters and high guessing parameters, indicating that these items are more likely to be correctly guessed. It is worth noting that there's an interesting observation regarding the results for item 1: since g_1 is very small and s_1 is very large, it is difficult for students who do not master the first attribute to get a correct response by guessing (the probability of a correct response is lower than 0.0200), and even if they do master the first attribute, the probability of a correct response is still only about 0.7000 due to the possibility of slipping.

Based on the results in Table S11 in the supplementary materials, we investigated the relationship between the VBEM-M algorithm and the other three algorithms in parameter estimation by analyzing the correlations of parameters $\boldsymbol{s}$ and $\boldsymbol{g}$ across these algorithms. The correlations between $\boldsymbol{s}$ estimates from the VBEM-M and VB algorithms is 0.9984, between VBEM-M and MCMC algorithms is 0.9979, and between VBEM-M and EM algorithms is 0.9989. The correlations between the estimators of $\boldsymbol{g}$ calculated using the VBEM-M algorithm and those obtained from the VB, MCMC, and EM algorithms are 0.9488, 0.9552, and 0.8632, respectively. These findings suggest that the VBEM-M algorithm's parameter estimates align more closely with those from VB and MCMC algorithms, as indicated by the high correlations. In addition, the estimators of the mixing proportions of attribute-mastery patterns, $\widehat{\pi}_l$ for $l = 1, \cdots, 2^5 = 32$, are presented in

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Figure S2 of the supplementary matrials. Notably, these estimates are highly consistent across the VBEM-M algorithm, VB algorithm, MCMC sampling technique, and EM algorithm. A total of 67% of the examinees were classified into the following four attribute profiles: (1,1,1,0,0), (1,1,1,1,0), (1,1,1,0,1), and (1,1,1,1,1). This suggests that a majority of students have mastered the first three attributes. The computation time for the VBEM-M, VB, MCMC, and EM algorithms were 0.1651 s, 0.1661 s, 11.3820 s, and 0.2870 s, respectively.

==========

Insert Table 11 about here

==========

5.2. Empirical Example 2

In this section, we analyze the Examination for the Certificate of Proficiency in English (ECPE) dataset based on the LCDM. The ECPE has been widely used in previous research based on the LCDM (e.g., Liu & Johnson, 2019; Templin & Bradshaw, 2014; Templin & Hoffman, 2013; von Davier, 2014b), and it includes 0-1 response data from 2,922 examinees on 28 items. Three attributes are measured: morphosyntactic rules, cohesive rules, and lexical rules. Nine of the 28 items measure two attributes, and the others measure one. The VBEM-M algorithm, VB algorithm (implemented in the "variationalDCM" package), MCMC algotithm (implemented in the "R2jags" package),

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

and EM algorithm (implemented in the "GDINA" package) were used for the parameter estimation of the LCDM model. However, due to space limitations, we only present the estimation results of the VBEM-M method in Tables 12 and 13. The results of the other algorithms can be found in the supplementary materials.

==========

Insert Table 12 about here

==========

==========

Insert Table 13 about here

==========

The outcomes of the VBEM-M algorithm were more similar to those of the VB algorithm and the MCMC algorithm. Please refer to the supplementary materials for more details. From Table 12, we found that the estimates of the interaction terms are relatively smaller compared to the main effects, indicating that the main effects have a greater influence on the probability of a correct response. Additionally, most of the interaction effects are positive, suggesting that the interactions between skills are more likely to positively affect the probability of a correct response. Furthermore, from the estimates of $\boldsymbol{\pi}$ in Table 13, we can observe that the most prevalent attribute mastery patterns are (0, 0, 0), (0, 0, 1), (0, 1, 1), and (1, 1, 1). This suggests a possible linear hierarchy structure among the skills. Specifically, mastering lexical rules requires mastering cohesive rules first,

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

and mastering morphosyntactic rules is a prerequisite for mastering cohesive rules. This finding is consistent with previous research conclusions (Gierl et al., 2007a, 2007b).

6. Discussion

In this paper, we propose the novel VBEM-M algorithm for estimating the parameters of the LCDM, which offers fast execution and excellent estimation accuracy. While Yamaguchi and Okada (2020a) introduced a VB method for estimating LCDM parameters, their approach primarily focuses on estimating the probability of correct item responses for specific attribute-mastery patterns, without directly estimating the item parameters. In contrast, our VBEM-M algorithm can simultaneously and directly estimate both attribute-mastery patterns and item parameters.

Since the posterior distributions of the item parameters in the LCDM do not have closed forms, it is difficult to execute parameter estimation using the classic VBEM algorithm. To get around this problem, in our approach, the likelihood function for the LCDM is replaced with a tight lower bound obtained by Taylor expansion, and inference is then performed. The item parameters based on the tight lower bound take on an exponential form, allowing us to use a Gaussian distribution as its conjugate prior. Additionally, a new location parameter ξ is introduced in implementing the Taylor expansion, and an extra maximizing step is added to the typical VBEM algorithm to seek the optimal local point ξ. Three simulation studies were carried out in this study: the first two focused on DINA model as the special case of the LCDM, while the third simulation

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

study considered the saturated LCDM. The parameter recovery results from the VBEM-M algorithm were analyzed under simulated conditions. The VBEM-M algorithm was shown to be effective in terms of parameter recovery, execution time, and convergence rate. In addition, the estimation accuracy and computation time of the VB, MCMC and EM algorithms were investigated in depth.

To begin with, it was found that the VBEM-M algorithm produces favorable results in terms of parameter recovery, providing three main benefits. First, the VBEM-M algorithm can be implemented under various sample sizes, and its accuracy improves as the sample size increases. Based on the DINA model, we found that higher attribute correlation does not affect $\boldsymbol{\eta}$ estimates but improves $\boldsymbol{\lambda}$ estimation accuracy. In addition, the convergence rate of the VBEM-M algorithm is fast, and it is not sensitive to the choice of initial values. It brings considerable efficiency gains, converging to the true values in only approximately ten iterations for different simulation conditions.

The second benefit is that the VBEM-M algorithm has a considerable accuracy advantage over other algorithms, especially when the sample size is small. For instance, in the DINA model with $N = 200$, $K = 5$, and $\sigma = 0.3$, under the LNL condition, the RMSEs of $\boldsymbol{\lambda}$ using VBEM-M, VB, MCMC-dina, MCMC-R2jags, EM-GDINA, and EM-CDM are 0.4500, 0.5507, 0.5470, 0.5554, 1.0239, and 1.0238, respectively. It is evident that our method shows significant advantages, particularly outperforming EM algorithms. However, this benefit diminishes as the sample size increases. This makes the VBEM-M algorithm

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

more reliable in situations with smaller sample sizes, which are often occurs in real-world applications.

Finally, the VBEM-M algorithm stands out for its computational efficiency. While not as fast as the EM algorithms, it still holds an advantage over other algorithms. For example, based on the DINA model with $N = 2000$, $J = 30$, $K = 5$, and $\sigma = 0.3$, it takes an average of 0.3686s, 0.5136s, 93.5225s, 2061.8450s, 0.1949s, and 0.2097s for VBEM-M, VB, MCMC-dina, MCMC-R2jags, EM-GDINA, and EM-CDM, respectively, across 100 replications. Compared to the two EM algorithms, our algorithm showed the time differences of only 0.1737s and 0.1589s, respectively, and it outperformed the other algorithms. This suggests that the VBEM-M algorithm performs well in terms of computational efficiency.

While the VBEM-M algorithm has its advantages, it also has some limitations. For instance, as mentioned above, the VBEM-M algorithm could not perform as fast as EM algorithm. In addition, the VBEM-M algorithm is essentially an approximation of the posterior distribution of parameters, which works well for the DINA model and some LCDM submodels, as showed in online supplement. However, its performance in complex LCDMs with high attribute dimensions (like a 32-dimensional $\boldsymbol{\lambda}^*$ for $K = 5$) still needs to be investigated.

In future studies, first, we will consider to explore whether the VBEM-M algorithm can be generalized to other types of CDMs, such as polytomous CDMs and longitudinal

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

CDMs. Second, in this study, the Q-matrix was calibrated in advance; however, in practice, there is a potential for mis-specification (Rupp & Templin, 2008a). Therefore, we will modify the VBEM-M algorithm to simultaneously estimate the Q-matrix and model parameters. Third, while the VBEM-M algorithm converges quickly, it still operates slower than the EM algorithm in terms of computation time. We plan to further optimize the code associated with C++ or Fortran to increase its speed.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The two datasets can be found as follows: https://cran.r-project.org/web/packages/CDM/index.html.

References

Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College London, London.
https://www.cse.buffalo.edu/faculty/mbeal/thesis/

Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York: Springer.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association, 112*(518), 859–877.
https://doi.org/10.1080/01621459.2017.1285773

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika 75*(1), 33–57. https://doi.org/10.1007/s11336-009-9136-x

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software, 76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. (2017). Bayesian estimation of the DINA Q matrix. *Psychometrika, 83*(1), 89–108. https://doi.org/10.1007/s11336-017-9579-4

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. https://doi.org/10.1080/01621459.2014.934827

Cho, A. E., Wang, C., Zhang, X., & Xu, G. (2021). Gaussian variational estimation for multidimensional item response theory. *British Journal of Mathematical and Statistical Psychology, 74*(S1), 52–85. https://doi.org/10.1111/bmsp.12219

Chung, M. (2019). A Gibbs sampling algorithm that estimates the Q-matrix for the DINA model. *Journal of Mathematical Psychology, 93*, 102275. https://doi.org/10.1016/j.jmp.2019.07.002

Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics, 40*(5), 454–476. https://doi.org/10.3102/1076998615595403

Culpepper, S. A. (2019). Estimating the cognitive diagnosis Q matrix with expert knowledge: Application to the fraction-subtraction dataset. *Psychometrika, 84*(2), 333–357. https://doi.org/10.1007/s11336-018-9643-8

Culpepper, S. A. & Balamuta, J. J. (2019). *dina: Bayesian Estimation of DINA Model* (R package version 2.0.0). Retrieved from https://cran.r-project.org/package=dina

Culpepper, S. A., & Hudson, A. (2018). An improved strategy for Bayesian estimation of the reduced reparameterized unified model. *Applied Psychological Measurement, 42*(2), 99–115. https://doi.org/10.1177/0146621617707511

da Silva, M. A., de Oliveira, E. S., von Davier, A. A., & Bazán, J. L. (2018). Estimating the DINA model parameters using the No-U-Turn Sampler. *Biometrical Journal, 60*(2), 352–368. https//doi.org/10.1002/bimj.201600225

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(1) 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of*

Educational and Behavioral Statistics, 34(1), 115–130. https://doi.org/10.3102/1076998607309474

de la Torre, J. (2011). The generalized DINA framework. *Psychometrika, 76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*(3), 333-353. https://doi.org/10.1007/BF02295640

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement, 36*(6), 447–468. https://doi.org/10.1177/0146621612449069

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, vol. 26, psychometrics* (pp. 979–1030). Amsterdam: Elsevier.

Eddelbuettel, D., & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis, 71*, 1054–1063. https://doi.org/10.1016/j.csda.2013.02.005

Eddelbuettel, D., & Francois, R. (2011). Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software, 40*, 1–18. https://doi.org/10.18637/jss.v040.i08

Fager, M., Pace, J., Templin, J.L. (2019). Using Mplus to Estimate the Log-Linear Cognitive Diagnosis Model. In: von Davier, M., Lee, YS. (eds) *Handbook of Diagnostic*

Classification Models(pp. 581–591). Springer, Cham.

George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. *Journal of Statistical Software, 74*(2), 1–24. https://doi.org/10.18637/jss.v074.i02

Gierl, M. J., Cui, Y., & Hunka, S. (2007a). *Using connectionist models to evaluate examinees' response patterns on tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Gierl, M.J., Leighton, J.P., & Hunka, S.M. (2007b). Using the attribute hierarchy method to make diagnostic inferences about respondents' cognitive skills. In J.P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education: theory and applications* (pp. 242–274). Cambridge: Cambridge University Press.

Grimmer, J. (2011). An introduction to Bayesian inference via variational approximations. *Political Analysis, 19*(1), 32–47. https://doi.org/10.1093/pan/mpq027

Haberman, S. J., & von Davier, M. (2007). Some notes on models for cognitively based skill diagnosis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, vol. 26, psychometrics* (pp. 1031–1038). Amsterdam: Elsevier.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*(4), 301–321. https://doi.org/10.1111/j.1745-3984.1989.tb00336.x

Hartz, S. (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., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5

Hijikata, K., Oka, M., Yamaguchi, K., & Okada, K. (2023). variationalDCM: An R package for variational Bayesian inference in diagnostic classification models. *PsyArXiv.* https://doi.org/10.31234/osf.io/f2sqd

Jaakkola, T. S., Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing, 10*(1), 25–37. https://doi.org/10.1023/A:1008932416310

Jeon, M., Rijmen, F., & Rabe-Hesketh, S. (2017). A variational maximization-maximization algorithm for generalized linear mixed models with crossed random effects. *Psychometrika, 82*(3), 693–716. https://doi.org/10.1007/s11336-017-9555-z

Jiang, Z., & Carter, R. (2019). Using Hamiltonian Monte Carlo to estimate the log-linear cognitive diagnosis model via Stan. *Behavior Research Methods, 51*(2), 651–662. https://doi.org/10.3758/s13428-018-1069-9

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to

variational methods for graphical models. *Machine learning*, *37*(2), 183–233. https://doi.org/10.1023/A:1007665907178

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258–272. https://doi.org/10.1177/01466210122032064

Liu, C. W. (2022). Efficient Metropolis-Hastings Robbins-Monro algorithm for high-dimensional diagnostic classification models. *Applied Psychological Measurement, 46*(8), 662–674. https://doi.org/10.1177/01466216221123981

Liu, C. W., Andersson, B., & Skrondal, A. (2020). A constrained Metropolis-Hastings Robbins-Monro algorithm for Q matrix estimation in DINA models. *Psychometrika 85*(2), 322–357 . https://doi.org/10.1007/s11336-020-09707-4

Liu, X., Johnson, M.S. (2019). Estimating CDMs Using MCMC. In: von Davier, M., Lee, YS. (eds) *Handbook of Diagnostic Classification Models*(pp. 629–649). Springer, Cham. https://doi.org/10.1007/978-3-030-05584-4_31

Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology, 69*(3), 253–275. https://doi.org/10.1111/bmsp.12070

Ma, W., & de la Torre, J. (2020). GDINA: An R Package for Cognitive Diagnosis Modeling. *Journal of Statistical Software, 93*(14), 1–26.

https://doi.org/10.18637/jss.v093.i14

Ma, W., & Guo, W. (2019). Cognitive diagnosis models for multiple strategies. *British Journal of Mathematical and Statistical Psychology, 72*(2), 370–392. https://doi.org/10.1111/bmsp.12155

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics, 2*(2), 99–120. https://doi.org/10.3102/10769986002002099

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*(2), 187–212. https://doi.org/10.1007/BF02294535

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks (Ed.), *Handbook of Markov Chain Monte Carlo* (pp. 113–162). Boca Raton, FL: CRC Press/Taylor & Francis.

Oka, M., Okada, K.(2023) Scalable Bayesian Approach for the Dina Q-Matrix Estimation Combining Stochastic Optimization and Variational Inference.*Psychometrika,88*, 302–331. https://doi.org/10.1007/s11336-022-09884-4

Oka, M., Saso, S., & Okada, K. (2023). Variational inference for a polytomous-attribute saturated diagnostic classification model with parallel computing.*Behaviormetrika,50*(1), 63–92. https://doi.org/10.1007/s41237-022-00164-0

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *The 3rd international workshop on distributed statistical computing, 124*, 1–8. Retrieved from http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.https://www.R-project.org/

Rijmen, F., Jeon, M., & Rabe-Hesketh, S. (2016). Variational approximation methods. In W. J. van der Linden (Ed.), *Handbook of item response theory: Statistical tools* (Vol. 2, pp. 259–270). CRC Press.

Rupp, A. A., & Templin, J. L. (2008a). Effects of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement, 68*(1), 78–96. https://doi.org/10.1177/0013164407301545

Rupp, A. A., & Templin, J. L. (2008b). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspective, 6*(4), 219–262. https://doi.org/10.1080/15366360802490866

Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods and applications*. New York: Guilford.

Su, Y. S., & Yajima, M. (2015). R2jags: Using R to run "JAGS". *R package version 0.7-1.* Retrieved from http://CRAN.R-project.org/package=R2jags

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society. Series C: Applied Statistics, 51*(3), 337–350. https://doi.org/10.1111/1467-9876.00272

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*(4), 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Erlbaum.

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika, 79*(2), 317–339.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice, 32*(2), 37–50.

Thomas, A., O'Hara, B., Ligges, U., & Sturtz, S. (2006). Making BUGS open. *R News, 6*, 12–17. Retrieved from http://mathstat.helsinki.fi/openbugs/FAQFrames.html

Urban, C. J., & Bauer, D. J. (2021). A deep learning algorithm for high-dimensional exploratory item factor analysis. *Psychometrika, 86*(1), 1–29. https://doi.org/10.1007/s11336-021-09748-3

von Davier, M. (2008). A general diagnostic model applied to language testing data. *Britsh Journal of Mathematical and Statistical Psychology, 61*(2), 287–307. https://doi.org/10.1348/000711007X193957

von Davier, M. (2014a). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology, 67*(1), 49–71. https://doi.org/10.1111/bmsp.12003

von Davier, M. (2014b). The log-linear cognitive diagnostic model (LCDM) as a special case of the general diagnostic model (GDM). *ETS Research Report Series, 2014*(2), 1–13.

Wand, M. P., Ormerod, J. T., Padoan, S. A., & Frühwirth, R. (2011). Mean field variational Bayes for elaborate distributions. *Bayesian Analysis, 6*(4), 847–900. https://doi.org/10.1214/11-BA631

Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, *113*(523), 1284–1295. https://doi.org/10.1080/01621459.2017.1340889

Yamaguchi, K. (2020). Variational Bayesian inference for the multiple-choice DINA model. *Behaviormetrika 47*(1), 159–187 . https://doi.org/10.1007/s41237-020-00104-w

Yamaguchi, K., & Okada, K. (2020a). Variational Bayes inference algorithm for the saturated diagnostic classification model. *Psychometrika 85*(4), 973–995. https://doi.org/10.1007/s11336-020-09739-w

Yamaguchi, K., & Okada, K. (2020b). Variational Bayes inference for the DINA model. *Journal of Educational and Behavioral Statistics, 45*(5), 569–597. https://doi.org/10.3102/1076998620911934

Yamaguchi, K., & Martinez, A. J. (2023). Variational Bayes inference for hidden Markov diagnostic classification models. *British Journal of Mathematical and Statistical Psychology, 00*, 1–25. https://doi.org/10.1111/bmsp.12308

Yamaguchi, K., & Templin, J. L. (2022). A Gibbs sampling algorithm with monotonicity constraints for diagnostic classification models. *Journal of Classification, 39*(1), 24–54. https://doi.org/10.1007/s00357-021-09392-7

Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics, 44*(4), 473–503. https://doi.org/10.3102/1076998619826040

Zhang, Z., Zhang, J., Lu, J., & Tao, J. (2020). Bayesian estimation of the dina model with Pólya-gamma Gibbs sampling. *Frontiers in Psychology, 11*, 384. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00384

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 1: Estimation procedure of the VBEM-M algorithm.

VBEM-M Algorithm

Input: $\boldsymbol{\delta}_0$, μ_{η_0}, $\sigma^2_{\eta_0}$, $\mu_{\lambda_{0,main}}$, $\sigma^2_{\lambda_{0,main}}$,$\mu_{\lambda_{0,inter}}$, $\sigma^2_{\lambda_{0,inter}}$, e_0, T

Initialization: $\mathrm{E}^{(0)}_{q(\eta_0)}[\eta_0]$, $\mathrm{E}^{(0)}_{q(\lambda_{0d})}[\lambda_{0d}]$, $\mathrm{E}^{(0)}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j]$, $\mathrm{E}^{(0)}_{q(\boldsymbol{\lambda}^*_j)}[\boldsymbol{\lambda}^*_j\boldsymbol{\lambda}^{*\mathrm{T}}_j]$.

Repeat

(a) VBE-step: update $q^*(\boldsymbol{z}_i)$ according to Eq.(25).

(b) VBM-step:

(b1) update $q^*(\boldsymbol{\pi})$ according to Eq.(27).

(b2) update $q^*(\boldsymbol{\lambda}^*_j)$ according to Eq.(29).

(b3) update $q^*(\eta_0)$ according to Eq.(31).

(b4) update $q^*(\lambda_{0,main})$ according to Eq.(33).

(b5) update $q^*(\lambda_{0,inter})$ according to Eq.(35).

(c) M-step: update ξ_{ijl} using Eq.(37).

Until the absolute difference of $\underline{\mathcal{L}}^*(q(\Theta), \boldsymbol{\xi})$ between two adjacent iterations is less than e_0 or $t >$T, where e_0 is the convergence threshold and T is the maximum iterations.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 2: The accuracy of item parameters and class membership probability parameters using the VBEM-M algorithm in simulation study 1.

	N=1000						N=2000					
LNL	$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$		$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$	
	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD
$\sigma = 0$	0.1351(–0.0118)	0.0920	0.2560(0.0365)	0.1765	0.0022(0.0000)	0.0053	0.0976(–0.0038)	0.0652	0.1877(0.0150)	0.1260	0.0015(0.0000)	0.0038
$\sigma = 0.3$	0.1369(–0.0140)	0.0937	0.2337(0.0307)	0.1617	0.0022(0.0000)	0.0051	0.0981(–0.0077)	0.0664	0.1669(0.0133)	0.1152	0.0016(0.0000)	0.0037
$\sigma = 0.7$	0.1388(–0.0104)	0.0967	0.2216(0.0279)	0.1516	0.0021(0.0000)	0.0045	0.1004(–0.0073)	0.0687	0.1600(0.0135)	0.1078	0.0015(0.0000)	0.0032
HNL	$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$		$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$	
	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD
$\sigma = 0$	0.1131(–0.0038)	0.0844	0.2240(0.0261)	0.1621	0.0058(0.0000)	0.0053	0.0797(–0.0038)	0.0559	0.1605(0.0132)	0.1155	0.0041(0.0000)	0.0038
$\sigma = 0.3$	0.1126(–0.0088)	0.0859	0.1979(0.0138)	0.1489	0.0056(0.0000)	0.0051	0.0813(–0.0016)	0.0609	0.1417(0.0069)	0.1059	0.0039(0.0000)	0.0037
$\sigma = 0.7$	0.1130(–0.0080)	0.0889	0.1738(0.0144)	0.1393	0.0048(0.0000)	0.0045	0.0822(–0.0038)	0.0629	0.1297(0.0101)	0.0990	0.0036(0.0000)	0.0032
SHG	$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$		$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$	
	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD
$\sigma = 0$	0.1406(–0.0093)	0.0919	0.2218(0.0217)	0.1667	0.0035(0.0000)	0.0053	0.1029(–0.0061)	0.0652	0.1649(0.0156)	0.1185	0.0025(0.0000)	0.0038
$\sigma = 0.3$	0.1440(–0.0100)	0.0935	0.2135(0.0171)	0.1534	0.0033(0.0000)	0.0051	0.1026(–0.0070)	0.0665	0.1475(0.0130)	0.1090	0.0025(0.0000)	0.0037
$\sigma = 0.7$	0.1478(–0.0156)	0.0968	0.1932(0.0240)	0.1444	0.0032(0.0000)	0.0045	0.1027(–0.0106)	0.0688	0.1362(0.0144)	0.1026	0.0021(0.0000)	0.0032
GHS	$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$		$\boldsymbol{\eta}$		$\boldsymbol{\lambda}$		$\boldsymbol{\pi}$	
	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD	RMSE(Bias)	SD
$\sigma = 0$	0.1038(–0.0064)	0.0844	0.2537(0.0279)	0.1729	0.0035(0.0000)	0.0053	0.0729(–0.0011)	0.0599	0.1852(0.0141)	0.1227	0.0026(0.0000)	0.0038
$\sigma = 0.3$	0.1073(–0.0047)	0.0860	0.2262(0.0193)	0.1573	0.0035(0.0000)	0.0051	0.0752(0.0005)	0.0610	0.1650(0.0131)	0.1119	0.0025(0.0000)	0.0037
$\sigma = 0.7$	0.1071(–0.0074)	0.0889	0.2031(0.0186)	0.1466	0.0032(0.0000)	0.0045	0.0763(–0.0023)	0.0630	0.1470(0.0080)	0.1042	0.0023(0.0000)	0.0032

Note: The values outside parentheses represent the RMSE, while the values inside the parentheses indicate bias. These reflect the average RMSE, bias and SD for all intercept parameters $\boldsymbol{\eta}$, slope parameters $\boldsymbol{\lambda}$, and class membership probability parameters $\boldsymbol{\pi}$.

Table 3: The accuracy of attribute profile parameters using the VBEM-M algorithm in simulation study 1.

LNL	$N = 1000$						$N = 2000$					
	AAR1	AAR2	AAR3	AAR4	AAR5	PAR	AAR1	AAR2	AAR3	AAR4	AAR5	PAR
$\sigma = 0$	0.9792	0.9667	0.9880	0.9672	0.9903	0.9025	0.9787	0.9667	0.9876	0.9669	0.9900	0.9007
$\sigma = 0.3$	0.9807	0.9718	0.9877	0.9712	0.9918	0.9107	0.9803	0.9727	0.9876	0.9719	0.9914	0.9120
$\sigma = 0.7$	0.9821	0.9799	0.9874	0.9797	0.9941	0.9290	0.9827	0.9807	0.9872	0.9803	0.9940	0.9307
HNL	$N = 1000$						$N = 2000$					
	AAR1	AAR2	AAR3	AAR4	AAR5	PAR	AAR1	AAR2	AAR3	AAR4	AAR15	PAR
$\sigma = 0$	0.9102	0.8959	0.9334	0.8919	0.9413	0.6736	0.9098	0.8939	0.9332	0.8935	0.9413	0.6731
$\sigma = 0.3$	0.9137	0.9085	0.9372	0.9039	0.9497	0.6980	0.9150	0.9111	0.9375	0.9061	0.9491	0.7031
$\sigma = 0.7$	0.9345	0.9335	0.9512	0.9286	0.9608	0.7724	0.9364	0.9350	0.9520	0.9301	0.9628	0.7781
SHG	$N = 1000$						$N = 2000$					
	AAR1	AAR2	AAR3	AAR4	AAR5	PAR	AAR1	AAR2	AAR3	AAR4	AAR15	PAR
$\sigma = 0$	0.9525	0.9443	0.9707	0.9406	0.9766	0.8172	0.9516	0.9458	0.9724	0.9432	0.9761	0.8213
$\sigma = 0.3$	0.9596	0.9452	0.9740	0.9435	0.9762	0.8266	0.9592	0.9462	0.9739	0.9432	0.9762	0.8269
$\sigma = 0.7$	0.9689	0.9618	0.9777	0.9604	0.9826	0.9871	0.9693	0.9630	0.9785	0.9614	0.9831	0.8734
GHS	$N = 1000$						$N = 2000$					
	AAR1	AAR2	AAR3	AAR4	AAR5	PAR	AAR1	AAR2	AAR3	AAR4	AAR15	PAR
$\sigma = 0$	0.9476	0.9442	0.9689	0.9416	0.9762	0.8153	0.9489	0.9462	0.9685	0.9424	0.9762	0.8186
$\sigma = 0.3$	0.9476	0.9521	0.9667	0.9467	0.9783	0.8238	0.9485	0.9519	0.9665	0.9478	0.9780	0.8248
$\sigma = 0.7$	0.9624	0.9620	0.9755	0.9584	0.9818	0.8630	0.9623	0.9623	0.9753	0.9590	0.9813	0.8635

Note: AAR1 represents the correct classification rate for the first attribute, AAR2 for the second attribute, AAR3 for the third attribute, AAR4 for the fourth attribute, and AAR5 for the fifth attribute. PAR stands for the pattern-wise agreement rate.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 4: The accuracy of item parameters and class membership probability parameters using the VBEM-M, VB, MCMC-dina, MCMC-R2jags, EM-GDINA, and EM-CDM algorithms for the DINA model under the $\sigma = 0.3$ condition in simulation study 2.

				$\boldsymbol{\eta}$			
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
LNL	$N = 200$	0.2759(–0.0439)	0.3068(0.0255)	0.3043(0.0291)	0.3045(0.0267)	0.3660(–0.0533)	0.3659(–0.0531)
	$N = 500$	0.1881(–0.0191)	0.1973(0.0112)	0.1968(0.0127)	0.1970(0.0120)	0.2050(–0.0172)	0.2050(–0.0172)
	$N = 1000$	0.1370(–0.0103)	0.1400(0.0052)	0.1400(0.0059)	0.1400(0.0055)	0.1426(–0.0088)	0.1426(–0.0088)
	$N = 2000$	0.0974(-0.0033)	0.0987(0.0046)	0.0988(0.0050)	0.0987(0.0048)	0.0994(–0.0024)	0.0994(–0.0024)
				$\boldsymbol{\lambda}$			
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.4500(0.1151)	0.5507(–0.1284)	0.5470(–0.1441)	0.5554(–0.1670)	1.0239(0.2478)	1.0238(0.2478)
	$N = 500$	0.3284(0.0541)	0.3552(–0.0502)	0.3551(–0.0564)	0.3567(–0.0656)	0.3961(0.0555)	0.3961(0.0555)
	$N = 1000$	0.2437(0.0295)	0.2532(–0.0243)	0.2527(–0.0272)	0.2539(–0.0317)	0.2638(0.0263)	0.2638(0.0263)
	$N = 2000$	0.1780(0.0129)	0.1814(–0.0146)	0.1814(–0.0159)	0.1817(–0.0183)	0.1845(0.0104)	0.1845(0.0104)
				$\boldsymbol{\pi}$			
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.0055(0.0000)	0.0055(0.0000)	0.0055(0.0000)	0.0055(0.0000)	0.0056(0.0000)	0.0056(0.0000)
	$N = 500$	0.0034(0.0000)	0.0034(0.0000)	0.0034(0.0000)	0.0034(0.0000)	0.0034(0.0000)	0.0034(0.0000)
	$N = 1000$	0.0023(0.0000)	0.0023(0.0000)	0.0023(0.0000)	0.0023(0.0000)	0.0023(0.0000)	0.0023(0.0000)
	$N = 2000$	0.0016(0.0000)	0.0016(0.0000)	0.0016(0.0000)	0.0016(0.0000)	0.0016(0.0000)	0.0016(0.0000)
				$\boldsymbol{\eta}$			
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
HNL	$N = 200$	0.2250(–0.0207)	0.2533(0.0010)	0.2456(0.0095)	0.2465(0.0036)	0.3180(0.0351)	0.3181(–0.0351)
	$N = 500$	0.1564(–0.0122)	0.1633(0.0002)	0.1622(0.0047)	0.1625(0.0016)	0.1682(–0.0103)	0.1681(–0.0102)
	$N = 1000$	0.1113(–0.0042)	0.1136(0.0024)	0.1133(0.0051)	0.1134(0.0036)	0.1150(–0.0027)	0.1150(–0.0027)
	$N = 2000$	0.0811(–0.0045)	0.0819(–0.0011)	0.0817(0.0000)	0.0818(–0.0005)	0.0824(–0.0039)	0.0824(–0.0039)
				$\boldsymbol{\lambda}$			
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.4108(0.0928)	0.4754(–0.0237)	0.4617(–0.0591)	0.4686(–0.0897)	0.6678(0.1202)	0.6679(0.1202)
	$N = 500$	0.2874(0.0322)	0.3064(–0.0179)	0.3042(–0.0342)	0.3068(–0.0472)	0.3192(0.0249)	0.3192(0.0249)
	$N = 1000$	0.2061(0.0211)	0.2118(–0.0047)	0.2112(–0.0138)	0.2116(–0.0199)	0.2165(0.0168)	0.2165(0.0169)
	$N = 2000$	0.1489(0.0118)	0.1509(–0.0011)	0.1505(–0.0055)	0.1507(–0.0087)	0.1524(0.0094)	0.1524(0.0094)
				$\boldsymbol{\pi}$			
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.0111(0.0000)	0.0113(0.0000)	0.0107(0.0000)	0.0107(0.0000)	0.0146(0.0000)	0.0146(0.0000)
	$N = 500$	0.0077(0.0000)	0.0077(0.0000)	0.0076(0.0000)	0.0077(0.0000)	0.0088(0.0000)	0.0088(0.0000)
	$N = 1000$	0.0056(0.0000)	0.0057(0.0000)	0.0057(0.0000)	0.0057(0.0000)	0.0060(0.0000)	0.0060(0.0000)
	$N = 2000$	0.0041(0.0000)	0.0041(0.0000)	0.0042(0.0000)	0.0042(0.0000)	0.0043(0.0000)	0.0043(0.0000)

Note: The values outside the parentheses represent the RMSE, while the values inside the parentheses indicate bias. Here, RMSE and Bias denote the average RMSE and Bias, respectively, for all intercept parameters $\boldsymbol{\eta}$, all slope parameters $\boldsymbol{\lambda}$ and all class membership probability parameters $\boldsymbol{\pi}$.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 5: The accuracy of attribute profile parameters using the VBEM-M, VB, MCMC-dina, MCMC-R2jags, EM-GDINA, and EM-CDM algorithms for the DINA model under the $\sigma = 0.3$ condition in simulation study 2.

		AAR1						AAR2					
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM	VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
LNL	$N = 200$	0.9877	0.9876	0.9876	0.9878	0.9874	0.9874	0.9580	0.9572	0.9579	0.9577	0.9564	0.9564
	$N = 500$	0.9875	0.9875	0.9876	0.9873	0.9875	0.9875	0.9578	0.9576	0.9577	0.9573	0.9576	0.9576
	$N = 1000$	0.9886	0.9886	0.9887	0.9886	0.9886	0.9886	0.9622	0.9622	0.9621	0.9620	0.9621	0.9621
	$N = 2000$	0.9884	0.9884	0.9883	0.9883	0.9884	0.9884	0.9614	0.9615	0.9613	0.9614	0.9615	0.9615
		AAR3						AAR4					
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM	VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.9832	0.9828	0.9828	0.9829	0.9825	0.9825	0.9820	0.9815	0.9818	0.9815	0.9808	0.9808
	$N = 500$	0.9855	0.9854	0.9853	0.9853	0.9855	0.9855	0.9821	0.9821	0.9820	0.9820	0.9820	0.9820
	$N = 1000$	0.9856	0.9856	0.9856	0.9856	0.9856	0.9856	0.9824	0.9824	0.9825	0.9825	0.9824	0.9824
	$N = 2000$	0.9852	0.9852	0.9852	0.9852	0.9852	0.9852	0.9817	0.9817	0.9817	0.9816	0.9817	0.9817
		AAR5						PAR					
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM	VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.9699	0.9690	0.9690	0.9690	0.9686	0.9686	0.8916	0.8892	0.8900	0.8902	0.8868	0.8869
	$N = 500$	0.9729	0.9729	0.9727	0.9727	0.9728	0.9728	0.8964	0.8962	0.8960	0.8954	0.8961	0.8961
	$N = 1000$	0.9726	0.9726	0.9724	0.9723	0.9726	0.9725	0.9002	0.9001	0.9001	0.8999	0.9000	0.9000
	$N = 2000$	0.9724	0.9724	0.9724	0.9725	0.9724	0.9724	0.8990	0.8990	0.8987	0.8989	0.8990	0.8990
		AAR1						AAR2					
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM	VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.9375	0.9366	0.9370	0.9366	0.9354	0.9354	0.8781	0.8746	0.8762	0.8748	0.8702	0.8701
	$N = 500$	0.9389	0.9386	0.9383	0.9385	0.9382	0.9383	0.8830	0.8818	0.8806	0.8809	0.8802	0.8803
	$N = 1000$	0.9408	0.9408	0.9407	0.9407	0.9406	0.9406	0.8878	0.8877	0.8876	0.8877	0.8873	0.8873
	$N = 2000$	0.9403	0.9404	0.9402	0.9401	0.9403	0.9404	0.8899	0.8899	0.8900	0.8898	0.8899	0.8899
		AAR3						AAR4					
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM	VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
HNL	$N = 200$	0.9244	0.9232	0.9238	0.9235	0.9218	0.9218	0.9174	0.9158	0.9169	0.9170	0.9130	0.9130
	$N = 500$	0.9278	0.9276	0.9272	0.9279	0.9264	0.9265	0.9192	0.9188	0.9188	0.9188	0.9176	0.9176
	$N = 1000$	0.9290	0.9289	0.9287	0.9290	0.9286	0.9286	0.9212	0.9211	0.9209	0.9212	0.9208	0.9209
	$N = 2000$	0.9304	0.9304	0.9302	0.9303	0.9303	0.9303	0.9209	0.9209	0.9205	0.9208	0.9209	0.9209
		AAR5						PAR					
		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM	VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
	$N = 200$	0.8939	0.8926	0.8938	0.8922	0.8875	0.8876	0.6570	0.6520	0.6550	0.6526	0.6397	0.6397
	$N = 500$	0.9015	0.9006	0.9004	0.9000	0.8996	0.8996	0.6707	0.6686	0.6675	0.6677	0.6646	0.6647
	$N = 1000$	0.9062	0.9060	0.9057	0.9059	0.9058	0.9058	0.6812	0.6809	0.6800	0.6807	0.6798	0.6798
	$N = 2000$	0.9064	0.9064	0.9063	0.9063	0.9064	0.9064	0.6838	0.6838	0.6834	0.6833	0.6837	0.6838

Note: AAR1 represents the correct classification rate for the first attribute, AAR2 for the second attribute, AAR3 for the third attribute, AAR4 for the fourth attribute, and AAR5 for the fifth attribute. PAR stands for the pattern-wise agreement rate.

Table 6: The computational time (in seconds) for the VBEM-M, VB, MCMC-dina, MCMC-R2jags, EM-GDINA, and EM-CDM algorithms with the $\sigma = 0.3$ condition based on DINA in simulation study 2.

		VBEM-M	VB	MCMC-dina	MCMC-R2jags	EM-GDINA	EM-CDM
LNL	$N = 200$	0.0721s	0.0483s	9.7123s	163.2525s	0.0847s	0.0428s
	$N = 500$	0.1220s	0.1092s	23.6093s	467.2393s	0.0887s	0.0668s
	$N = 1000$	0.2061s	0.2415s	46.5265s	998.5244s	0.1208s	0.1126s
	$N = 2000$	0.3686s	0.5136s	93.5255s	2061.8450s	0.1949s	0.2097s
HNL	$N = 200$	0.0783s	0.1012s	9.7794s	170.3141s	0.1618s	0.0606s
	$N = 500$	0.1524s	0.2613s	23.8735s	463.9300s	0.1624s	0.0886s
	$N = 1000$	0.2617s	0.5403s	47.2185s	994.7302s	0.2078s	0.1376s
	$N = 2000$	0.4890s	1.0802s	94.3035s	2190.041s	0.2986s	0.2667s

Table 7: True values of $\boldsymbol{\lambda}^*$ for the saturated LCDM in simulation study 3.

	η	$\boldsymbol{\lambda}_{main}$			$\boldsymbol{\lambda}_{inter}$			
Item	η	λ_1	λ_2	λ_3	λ_{12}	λ_{13}	λ_{23}	λ_{123}
1	–1.5	3.5	0	0	0	0	0	0
2	–1.5	0	3.5	0	0	0	0	0
3	–1.5	0	0	3.5	0	0	0	0
4	–1.5	3.5	0	0	0	0	0	0
5	–1.5	0	3.5	0	0	0	0	0
6	–1.5	0	0	3.5	0	0	0	0
7	–1.5	2	2	0	–0.5	0	0	0
8	–1.5	2	0	2	0	–0.5	0	0
9	–1.5	0	2	2	0	0	–0.5	0
10	–1.5	1.5	1.5	1.5	–0.5	–0.5	–0.5	1
11	–1.5	2	2	0	–0.5	0	0	0
12	–1.5	2	0	2	0	–0.5	0	0
13	–1.5	0	2	2	0	0	–0.5	0
14	–1.5	1.5	1.5	1.5	–0.5	–0.5	–0.5	1
15	–1.5	2	2	0	–0.5	0	0	0
16	–1.5	2	0	2	0	–0.5	0	0
17	–1.5	0	2	2	0	0	–0.5	0
18	–1.5	1.5	1.5	1.5	–0.5	–0.5	–0.5	1

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 8: The accuracy of item parameters and class membership probability parameters using the VBEM-M, VB, MCMC-R2jags, and EM-GDINA algorithms for the LCDM model under the $\sigma = 0.3$ condition in simulation study 3.

		$\boldsymbol{\eta}$				$\boldsymbol{\lambda}_{main}$			
		VBEM-M	VB	MCMC-R2jags	EM-GDINA	VBEM-M	VB	MCMC-R2jags	EM-GDINA
$\sigma = 0.3$	$N = 200$	0.3044(–0.0454)	0.3929(–0.0562)	0.3312(0.1008)	0.4784(–0.0602)	0.4075(0.0720)	0.5998(–0.0909)	0.4922(–0.1235)	0.8472(0.0909)
	$N = 500$	0.2128(–0.0238	0.2431(–0.0297)	0.2263(0.0419)	0.2492(–0.0248)	0.2944(0.0293)	0.3701(–0.0272)	0.3469(–0.0708)	0.3883(0.0362)
	$N = 1000$	0.1489(0.0009)	0.1642(–0.0036)	0.1600(0.0348)	0.1661(–0.0002)	0.2244(0.0051)	0.2627(–0.0161)	0.2543(–0.0461)	0.2689(0.0146)
	$N = 2000$	0.1115(0.0030)	0.1177(0.0001)	0.1162(0.0208)	0.1183(0.0016)	0.1695(–0.0030)	0.1866(–0.0106)	0.1834(–0.0298)	0.1884(0.0044)
		$\boldsymbol{\lambda}_{inter}$				$\boldsymbol{\pi}$			
		VBEM-M	VB	MCMC-R2jags	EM-GDINA	VBEM-M	VB	MCMC-R2jags	EM-GDINA
	$N = 200$	0.5798(–0.0653)	1.0939(0.0215)	0.9812(0.3198)	2.3665(0.1440)	0.0177(0.0000)	0.0302(0.0000)	0.0187(0.0000)	0.0290(0.0000)
	$N = 500$	0.4880(–0.0253)	0.7015(0.0007)	0.6916(0.1476)	0.7824(0.0034)	0.0109(0.0000)	0.0179(0.0000)	0.0110(0.0000)	0.0173(0.0000)
	$N = 1000$	0.3960(–0.0014)	0.5032(0.0042)	0.5030(0.0855)	0.5311(0.0060)	0.0077(0.0000)	0.0122(0.0000)	0.0078(0.0000)	0.0119(0.0000)
	$N = 2000$	0.3140(0.0023)	0.3683(0.0019)	0.3665(0.0455)	0.3770(0.0020)	0.0054(0.0000)	0.0086(0.0000)	0.0054(0.0000)	0.0085(0.0000)

Note: The values outside the parentheses represent the RMSE, while the values inside the parentheses indicate bias. Here, RMSE and Bias denote the average RMSE and Bias, respectively, for all intercept parameters $\boldsymbol{\eta}$, all main effect slope parameters $\boldsymbol{\lambda}_{main}$, all interaction slope parameters $\boldsymbol{\lambda}_{inter}$ and all class membership probability parameters $\boldsymbol{\pi}$.

Table 9: Evaluation of the accuracy of attribute profile parameters using the VBEM-M, VB, MCMC-R2jags and EM-GDINA Algorithms for the saturated LCDM under the $\sigma = 0.3$ condition in simulation study 3.

		AAR1				AAR2			
		VBEM-M	VB	MCMC-R2jags	EM-GDINA	VBEM-M	VB	MCMC-R2jags	EM-GDINA
$\sigma = 0.3$	$N = 200$	0.9361	0.9306	0.9315	0.9284	0.9400	0.9334	0.9370	0.9334
	$N = 500$	0.9428	0.9425	0.9418	0.9421	0.9410	0.9404	0.9403	0.9400
	$N = 1000$	0.9427	0.9423	0.9420	0.9418	0.9441	0.9442	0.9440	0.9438
	$N = 2000$	0.9434	0.9434	0.9433	0.9431	0.9438	0.9438	0.9437	0.9436
		AAR3				PAR			
		VBEM-M	VB	MCMC-R2jags	EM-GDINA	VBEM-M	VB	MCMC-R2jags	EM-GDINA
	$N = 200$	0.9355	0.9310	0.9324	0.9304	0.8296	0.8126	0.8204	0.8120
	$N = 500$	0.9392	0.9391	0.9392	0.9384	0.8401	0.8387	0.8389	0.8380
	$N = 1000$	0.9434	0.9432	0.9431	0.9429	0.8468	0.8462	0.8459	0.8456
	$N = 2000$	0.9429	0.9430	0.9429	0.9428	0.8465	0.8464	0.8464	0.8461

Note: AAR1 represents the correct classification rate for the first attribute, AAR2 for the second attribute, AAR3 for the third attribute. PAR stands for the pattern-wise agreement rate.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 10: The computational time (in seconds) for the VBEM-M, VB, MCMC-R2jags and EM-GDINA algorithms based on LCDM with the $\sigma = 0.3$ condition in simulation study 3.

		VBEM-M	VB	MCMC-R2jags	EM-GDINA
$\sigma = 0.3$	$N = 200$	0.0564s	0.0816s	170.1357s	1.4199s
	$N = 500$	0.0705s	0.1061s	478.8929s	0.6961s
	$N = 1000$	0.0978s	0.1827s	1072.0754s	0.6939s
	$N = 2000$	0.1680s	0.3605s	3162.5147s	0.7485s

Table 11: The Q-matrix and the estimation results of the parameters η and λ using the VBEM-M algorithm in the empirical example 1

	Q-matrix					Estimate			
Item	1	2	3	4	5	$\hat{\eta}$	$\hat{\lambda}$	$\hat{g}$	$\hat{s}$
1	1	0	0	0	0	–3.9286(0.2585)	4.8606(0.2746)	0.0193	0.2825
2	1	1	1	1	0	–1.3649(0.1225)	3.4358(0.1893)	0.2035	0.1120
3	1	0	0	0	0	–1.9954(0.2123)	5.1863(0.2440)	0.1197	0.0395
4	1	1	1	1	1	–1.9774(0.1223)	3.9219(0.1998)	0.1216	0.1252
5	0	0	1	0	0	–1.8950(0.2203)	3.0390(0.2398)	0.1307	0.2416
6	1	1	1	1	0	–3.3033(0.1546)	4.5416(0.2015)	0.0355	0.2247
7	1	1	1	1	0	–2.5221(0.1417)	5.0019(0.2047)	0.0743	0.0773
8	1	1	0	0	0	–1.8014(0.1785)	4.7880(0.2181)	0.1417	0.0480
9	1	0	1	0	0	–2.3739(0.1983)	5.0362(0.2306)	0.0835	0.0652
10	1	0	1	1	1	–1.6155(0.1180)	4.2588(0.2073)	0.1658	0.0664
11	1	0	1	0	0	–2.2268(0.1952)	4.3746(0.2246)	0.0974	0.1045
12	1	0	1	1	0	–3.2651(0.1552)	5.1252(0.2064)	0.0368	0.1347
13	1	1	1	1	0	–1.9080(0.1324)	3.6173(0.1889)	0.1293	0.1533
14	1	1	1	1	1	–3.5572(0.1459)	4.9477(0.2081)	0.0277	0.1993
15	1	1	1	1	0	–3.9134(0.1649)	5.3988(0.2111)	0.0196	0.1846

Note: The values outside the parentheses represent the posterior means of the parameters, while the values inside the parentheses indicate the standard deviation.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

Table 12: The estimation results of the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ using the VBEM-M algorithm in the empirical example 2.

Item	$\hat{\eta}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\lambda_3}$	$\hat{\lambda_{12}}$	$\hat{\lambda_{13}}$	$\hat{\lambda_{23}}$	
1	0.8043(0.0576)	0.6103(0.2493)	0.7109(0.1066)	–	0.4428(0.2724)	–	–	
2	1.0281(0.0572)	–	1.2528(0.0821)	–	–	–	–	
3	–0.3492(0.0659)	0.9689(0.2787)	–	0.3714(0.0929)	–	0.3094(0.2915)	–	
4	–0.1438(0.0642)	–	–	1.6936(0.0808)	–	–	–	
5	1.0740(0.0671)	–	–	2.0166(0.0890)	–	–	–	
6	0.8621(0.0661)	–	–	1.6847(0.0859)	–	–	–	
7	–0.0809(0.0656)	1.7865(0.2990)	–	0.9441(0.0941)	–	0.1457(0.3131)	–	
8	1.4738(0.0594)	–	1.9063(0.0895)	–	–	–	–	
9	0.1172(0.0642)	–	–	1.1930(0.0801)	–	–	–	
10	0.0708(0.0467)	2.0545(0.0841)	–	–	–	–	–	
11	–0.0525(0.0655)	1.3287(0.2892)	–	0.9845(0.0943)	–	0.2637(0.3035)	–	
12	–1.7782(0.0731)	0.5863(0.2888)	–	1.3152(0.0985)	–	0.9094(0.3008)	–	
13	0.6723(0.0476)	1.6258(0.0857)	–	–	–	–	–	
14	0.1837(0.0468)	1.3824(0.0807)	–	–	–	–	–	
15	0.9875(0.0666)	–	–	2.1183(0.0887)	–	–	–	
16	–0.0791(0.0656)	1.4896(0.2920)	–	0.8778(0.0939)	–	0.0136(0.3057)	–	
17	1.3267(0.0708)	–	1.0508(0.2745)	0.6181(0.1291)	–	–	–0.1952(0.2980)	
18	0.9132(0.0663)	–	–	1.4051(0.0851)	–	–	–	
19	–0.1952(0.0642)	–	–	1.8412(0.0812)	–	–	–	
20	–1.4189(0.0706)	1.0231(0.2775)	–	0.9529(0.0966)	–	0.6143(0.2903)	–	
21	0.1639(0.0656)	1.0841(0.2886)	–	1.1344(0.0958)	–	0.0312(0.3032)	–	
22	–0.8644(0.0661)	–	–	2.2256(0.0818)	–	–	–	
23	0.6594(0.0558)	–	2.0529(0.0834)	–	–	–	–	
24	–0.6815(0.0559)	–	1.5284(0.0758)	–	–	–	–	
25	0.0953(0.0467)	1.1596(0.0792)	–	–		–	–	–
26	0.1574(0.0642)	–	–	1.1265(0.0801)	–	–	–	
27	–0.8658(0.0481)	1.7058(0.0784)	–	–	–	–	–	
28	0.5622(0.0650)	–	–	1.7455(0.0841)	–	–	–	

Note: The values outside the parentheses represent the posterior means of the parameters, while the values inside the parentheses indicate the standard deviation.

Table 13: The estimation results of the class membership parameters $\boldsymbol{\pi}$ using the VBEM-M algorithm in the empirical example 2.

	(0,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(1,1,0)	(1,0,1)	(0,1,1)	(1,1,1)
π	0.2966	0.0098	0.0170	0.1318	0.0071	0.0145	0.1793	0.3439

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

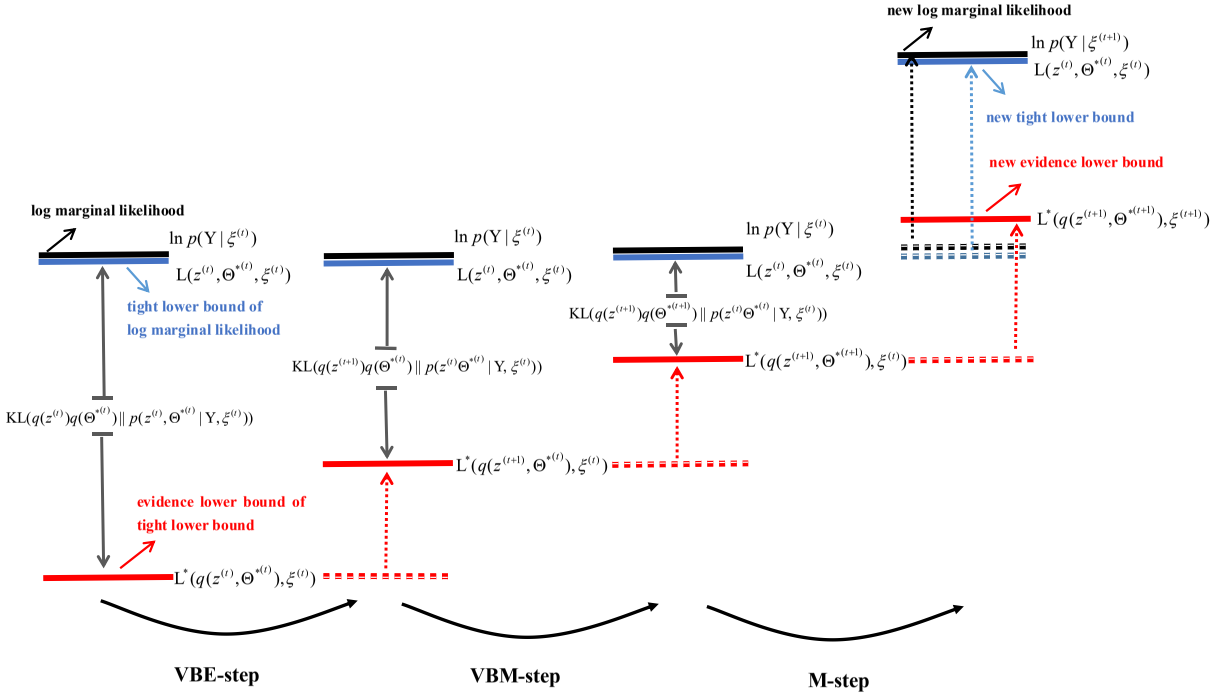

Figure 1: Graphical illustration of the VBEM-M algorithm implementation process. Let $\Theta^{*(t)} = (\boldsymbol{\pi}^{(t)}, \boldsymbol{\lambda}^{*(t)}, \boldsymbol{\lambda}_0^{*(t)})$. the variational density of the latent variable $\boldsymbol{z}^{(t+1)}$ is updated in the VBE-step. In VBM-step, the variational densities for model parameters and hyperparameters $\Theta^{*(t+1)}$ are updated. In M-step, we update $\boldsymbol{\xi}^{*(t+1)}$ by maximizing $\mathrm{L}(q(\boldsymbol{z}^{(t+1)}, \Theta^{*(t+1)}), \boldsymbol{\xi}^{*(t)})$.

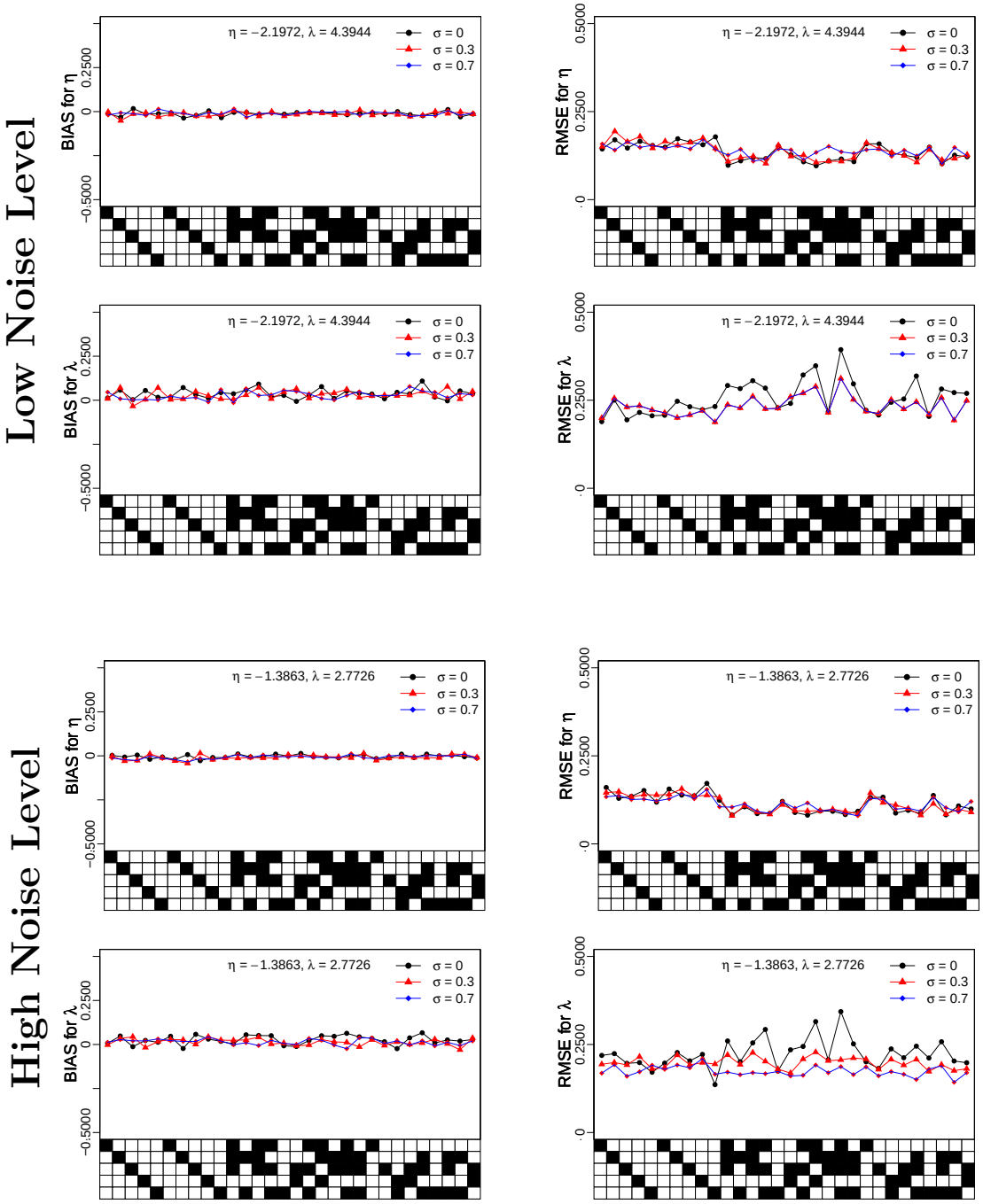

Figure 2: The bias and RMSE of η and λ for each item in the simulation study 1. The Q-Matrix denotes the skills required for each item along the x axis, where the black square ="1" and white square ="0".

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press

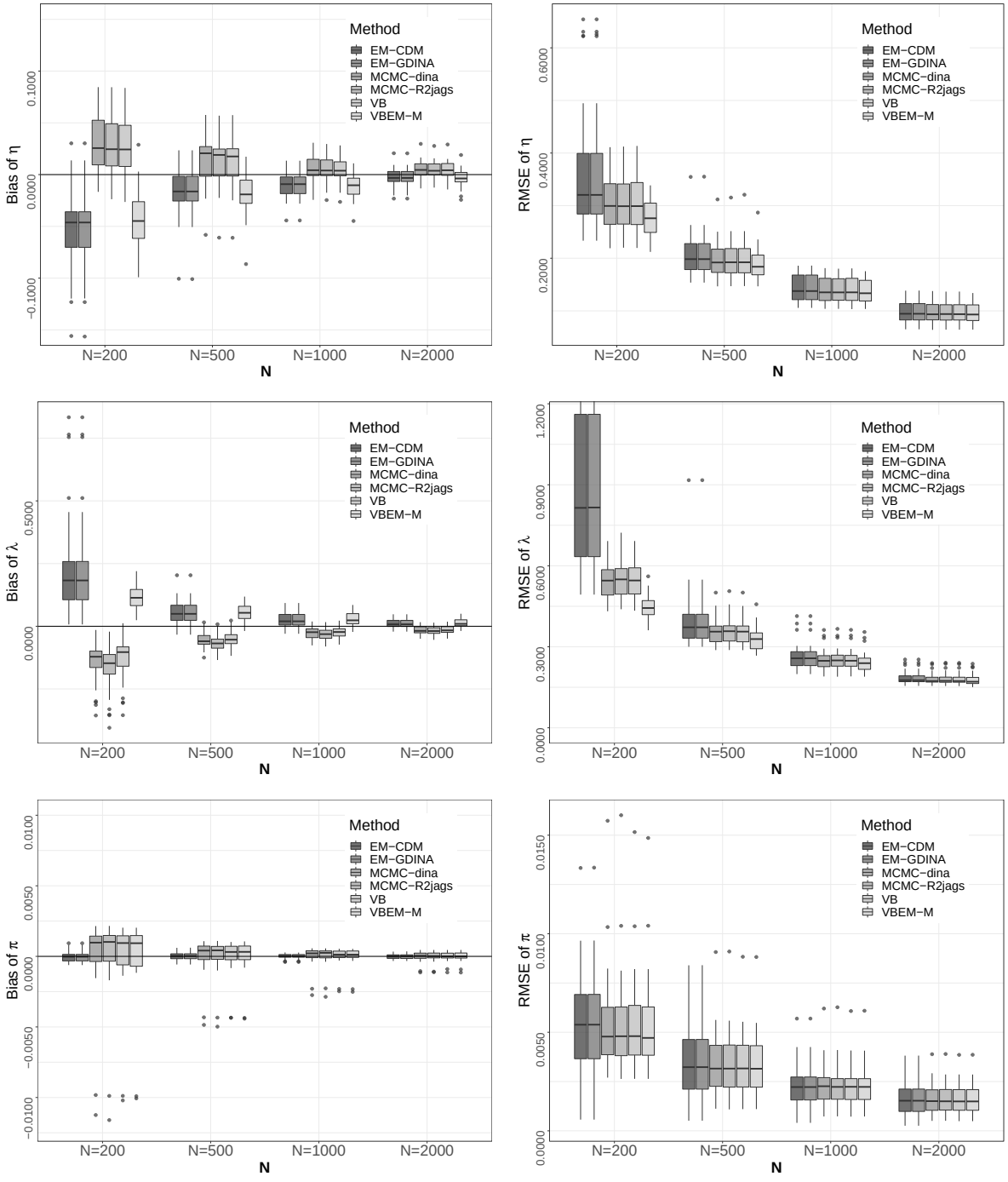

Figure 3: The boxplots of bias and RMSE for η, λ and π estimated by the VBEM-M, VB, MCMC-dina, MCMC-R2jags, EM-GDINA and EM-CDM with $\sigma = 0.3$ under the LNL condition in simulaion study 2.

https://doi.org/10.1017/psy.2024.7 Published online by Cambridge University Press