



The Bayesian Expectation-Maximization-Maximization for the 3PLM

Shaoyang Guo¹ and Chanjin Zheng^{2,3*}

¹ Institute of Curriculum and Instruction, Faculty of Education, East China Normal University, Shanghai, China, ² Department of Educational Psychology, Faculty of Education, East China Normal University, Shanghai, China, ³ Words up your way, Beijing, China

The current study proposes an alternative feasible Bayesian algorithm for the three-parameter logistic model (3PLM) from a mixture-modeling perspective, namely, the Bayesian Expectation-Maximization-Maximization (Bayesian EMM, or BEMM). As a new maximum likelihood estimation (MLE) alternative to the marginal MLE EM (MMLE/EM) for the 3PLM, the EMM can explore the likelihood function much better, but it might still suffer from the unidentifiability problem indicated by occasional extremely large item parameter estimates. Traditionally, this problem was remedied by the Bayesian approach which led to the Bayes modal estimation (BME) in IRT estimation. The current study attempts to mimic the Bayes modal estimation method and develop the BEMM which, as a combination of the EMM and the Bayesian approach, can bring in the benefits of the two methods. The study also devised a supplemented EM method to estimate the standard errors (SEs). A simulation study and two real data examples indicate that the BEMM can be more robust against the change in the priors than the Bayes modal estimation. The mixture modeling idea and this algorithm can be naturally extended to other IRT with guessing parameters and the four-parameter logistic models (4PLM).

Keywords: 3PL, Bayesian EMM, Bayesian EM, mixture modeling, estimation

OPEN ACCESS

Edited by:

Holmes Finch,
Ball State University, United States

Reviewed by:

Xin Luo,
Uber, United States
Lietta Marie Scott,
Arizona Department of Education,
United States

*Correspondence:

Chanjin Zheng
russelzheng@gmail.com

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 22 September 2018

Accepted: 03 May 2019

Published: 31 May 2019

Citation:

Guo S and Zheng C (2019) The
Bayesian Expectation-Maximization-
Maximization for the 3PLM.
Front. Psychol. 10:1175.
doi: 10.3389/fpsyg.2019.01175

INTRODUCTION

The field of educational testing has witnessed successful development and implementation of a great variety of test item formats, including multiple-choice questions, constructed response questions, and complex performance-based questions. For the past decades, however, multiple-choice questions have been the dominant item format, especially in standardized testing. One major downside of this item format is that examinees may exploit various specific test-taking strategies to improve their performance such as guessing, especially in a low-stakes test (Lord, 1980; Baker and Kim, 2004; Cao and Stokes, 2008; Woods, 2008; Cui et al., 2018).

Consequently, researchers and practitioners have devised powerful statistical tools to model dichotomously scored examinee responses to multiple-choice items. As early as in the nascent stage of IRT, Birnbaum (1968) proposed the three-parameter logistic model (3PLM) and its equivalent model in the normal ogive form to accommodate this need. Since then, the 3PLM has become one of the major statistical tools to analyze multiple-choice data. Various more complicated three-parameter models have been developed and important examples include three-parameter multilevel models (Fox, 2010), three-parameter multidimensional normal ogive model (Samejima, 1974; McDonald, 1999; Bock and Schilling, 2003), three-parameter multidimensional logistic model (Reckase, 2009) and three-parameter partially compensatory multidimensional models (Sympson, 1978).

One major problem that hinders the widespread application of the three-parameter models is the huge challenge of item parameter estimation brought by the guessing parameter. Even for the simplest three-parameter model, the 3PLM, this caused issues for researchers. The marginal maximum likelihood estimation (MMLE) with expectation maximization (EM) algorithm (MMLE/EM) represents the major breakthrough in the estimation techniques for the full-information item factor analysis, but it often runs into convergence problem for data sparseness for the guessing parameter. Mislevy (1986) offered a practical Bayesian solution, namely, the Bayes modal estimation (BME) or the Bayesian EM (BEM), for a moderate sample size, although an MLE solution was not available. Its implementation in BILOG-MG (Zimowski et al., 2003) paved the way for the wide application of the 3PLM in practice. The priors in the Bayesian method can provide extra information and shrink the estimates back to the conditional mean of item parameter, and this shrinkage depends on how informative the priors are. In this case, an informative prior with a smaller variance may have greater influence on the estimation, while a non-informative prior with a larger variance would be relatively weak (Baker and Kim, 2004). However, it is important to emphasize that an informative prior is not the same as an appropriate prior. Mislevy (1986) has warned that, a prior with an incorrect mean and a very small variance is likely to result in a systematic bias. Moreover, when the likelihood function of the MMLE/EM is flat, the Bayesian estimates might be highly dependent on the priors, and lead to a potentially undesirable result. So, it is vital to specify appropriate priors in the Bayesian EM algorithm and one crucial element of BILOG-MG is that the default priors for item parameters are generally uninformative but functional (Mislevy, 1986).

Recently, Zheng et al. (2017) pursued a different direction and developed a more powerful MLE algorithm based on a mixture modeling reformation of 3PLM, namely the Expectation-Maximization-Maximization (EMM). The EMM essentially is modified variant of MMLE/EM and its major difference from the traditional method is to expand the complete data from (\mathbf{U}, θ) to $(\mathbf{U}, \mathbf{Z}, \theta)$ where \mathbf{U} and θ are the response and examinee matrices, and \mathbf{Z} is the latent indicator for whether some examinees use the guessing strategy for one item. EMM can be summarized conceptually as

a) introducing a new latent variable \mathbf{Z} to construct a space one more dimension than the old one, which appears to be unwise because the original 3PLM estimation problem has been made more difficult with one added dimension.

b) invoking the independent assumption of (\mathbf{Z}, θ) to approximate the joint distribution which is one the most commonly used method in statistics to address high-dimension space problems.

c) using the approximation as a surrogate of the original 3PLM likelihood function to obtain item parameter estimate.

Simulation studies indicated that in the expanded space, one can better explore the likelihood function and thus is able to obtain the MLE solution with a moderate sample size. But the EMM is only a partial solution to the possible unidentifiability issue of the 3PLM, evidenced by occasional

improbable parameter estimates in the simulation study, and thus further improvement is necessary.

This paper attempts to propose a Bayesian version of the EMM (BEMM) which will bring in the advantages of both the EMM and the Bayesian method. On one hand, the BEMM can solve the unidentifiability issue with the 3PLM by adding additional prior information as in the BEM; on the other hand, it is expected to be more robust against the change in the item parameter priors than the Bayesian EM due to its power in exploitation of the likelihood function.

The rest of the paper is organized as follows: First, the mixture modeling approach to the 3PLM in IRT literature are carefully summarized and an alternative mixture modeling reformulation of the 3PLM with less stringent assumptions is presented. Then, two BEMM algorithms (BEMM-P and BEMM-C) and the estimating method for standard errors (SEs) are derived. One simulation study is carried out to demonstrate the feasibility of the BEMMs, compared to the Bayesian EM and the EMM; and two real data calibrations are presented to show the advantage of the BEMM algorithms. Lastly, future directions are discussed.

MIXTURE-MODELING APPROACH TO THE 3PLM

Mixture modeling is a powerful statistical tool for accommodating heterogeneity among an overall population. The similar idea for the 3PLM is not even entirely new. A two-process theory on the 3PLM, a p-process which represents the answering behavior based on examinee's ability and a g-process for the guessing strategy, was mentioned by Hutchinson (1991). Later this idea was extensively discussed by San Martín et al. (2006) to justify the development of IRT models of ability-based guessing, and by Maris and Bechger (2009) to demonstrate the identifiability and interpretability issue of the 3PLM. Although von Davier (2009) provided a clear summary of the status quo, differences on the nature of the two processes are far from being settled.

The current paper focus on taking advantage of the two-process reformulation to address the item parameter estimation difficulty of the 3PLM, but the discussion on the conceptual differences is beyond its scope. Two arrangements of these processes can be identified: the g-process comes first or the way around, and thus two different versions of reformulation of the 3PLM can be developed. Interesting enough, the one proposed by Zheng et al. (2017) corresponds to the one with the g-process coming first and Béguin and Glas (2001) proposed an ability-based reformulation for the three-parameter normal ogive model (3PNO) which coincides with the one with the p-process coming first. The two reformulations are briefly reviewed here as the starting point of the two BEMM algorithms

The basic formulation of the 3PLM is defined as:

$$P(u_{ij} = 1 | \theta_j, a_i, b_i, c_i) \equiv P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp(-Da_i(\theta_j - b_i))} \quad (1)$$

where u_{ij} represents the response of examinee j ($j = 1, 2, \dots, N$) on item i ($i = 1, 2, \dots, n$); a_i , b_i , and c_i are the discrimination, difficulty, and guessing parameters for the i^{th} item, respectively; θ_j is the ability parameter of the examinee j ; and D is the scaling constant, 1.702. Let $\xi_i = (a_i, b_i, c_i)$ represents the item parameter vector for i^{th} item, the function $P(u_{ij} = 1|\theta_j, \xi_i)$ can be abbreviated as $P_i(\theta_j)$, which is the probability of the correct response $u_{ij} = 1$ to item i given θ_j . The 3PLM can be conceived as an extension of the 2PLM with an item-specific guessing parameter:

$$P_i(\theta_j) = c_i + (1 - c_i) \times P_i^*(\theta_j) \tag{2}$$

with

$$P_i^*(\theta_j) = \frac{1}{1 + \exp(-Da_i(\theta_j - b_i))} \tag{3}$$

as the 2PLM. Following Béguin and Glas (2001)'s ability-based representation for the three-parameter normal ogive model (3PNO), the 3PLM can also be written as

$$P_i(\theta_j) = P_i^*(\theta_j) + [1 - P_i^*(\theta_j)] \times c_i. \tag{4}$$

A reformulation of the ability-based 3PLM, similar to Zheng et al. (2017), can be derived readily. Following Culpepper (2015), we may introduce a latent variable $v_{ij} \in \mathbf{V}$, and $v_{ij} \sim \text{Bernoulli}(P_i^*(\theta_j))$:

$$v_{ij} = \begin{cases} 1, & \text{if examinee } j \text{ has ability to answer item } i \text{ correctly;} \\ 0, & \text{if examinee } j \text{ does not have ability to answer item } i \text{ correctly.} \end{cases}$$

Reasonably, v_{ij} follows a Bernoulli distribution with parameter $P_i^*(\theta_j)$, or $P(v_{ij} = 1, u_{ij} = 1|\xi_i, \theta_j) = P(u_{ij} = 1|\xi_i, \theta_j) = P_i^*(\theta_j)$. From the mixture-modeling perspective, depending on the value of v_{ij} , the possibilities of 3PLM can be decomposed into two irrelevant parts: 1 and c_i .

Furthermore, the conditional possibilities $P(u_{ij}|v_{ij}, \theta_j, \xi_i)$ can be easily obtained as:

$$\begin{aligned} P(u_{ij} = 1|v_{ij} = 1, \theta_j, \xi_i) &= 1, P(u_{ij} = 1|v_{ij} = 0, \theta_j, \xi_i) = c_i, \\ P(u_{ij} = 0|v_{ij} = 1, \theta_j, \xi_i) &= 0, P(u_{ij} = 0|v_{ij} = 0, \theta_j, \xi_i) = 1 - c_i. \end{aligned} \tag{5}$$

By the multiplication rule, $P(u_{ij}, v_{ij}|\theta_j, \xi_i) = P(u_{ij}|v_{ij}, \theta_j, \xi_i)P(v_{ij})$, the joint distribution of u_{ij} and v_{ij} can be calculated as:

$$\begin{aligned} P(u_{ij} = 1, v_{ij} = 1|\theta_j, \xi_i) &= P_i^*(\theta_j), \\ P(u_{ij} = 1, v_{ij} = 0|\theta_j, \xi_i) &= c_i(1 - P_i^*(\theta_j)), \\ P(u_{ij} = 0, v_{ij} = 1|\theta_j, \xi_i) &= 0, \\ P(u_{ij} = 0, v_{ij} = 0|\theta_j, \xi_i) &= (1 - c_i)(1 - P_i^*(\theta_j)). \end{aligned} \tag{6}$$

Note that $P_i^*(\theta_j) + c_i(1 - P_i^*(\theta_j)) + (1 - c_i)(1 - P_i^*(\theta_j)) = 1$, so $P(u_{ij} = 0, v_{ij} = 1|\theta_j, \xi_i) = 0$ is actually redundant and can be omitted from the probability density function. From this distribution law of (u_{ij}, v_{ij}) conditional on (θ_j, ξ_i) , the marginal likelihood function of the EMM is

$$L(\mathbf{U}, \mathbf{V}|\xi) = \prod_{j=1}^N \int_{\theta_j} \left\{ \prod_{i=1}^n P_i^*(\theta_j)^{u_{ij}v_{ij}} \times [c_i(1 - P_i^*(\theta_j))]^{u_{ij}(1-v_{ij})} \right.$$

$$\left. \times [(1 - c_i)(1 - P_i^*(\theta_j))]^{(1-u_{ij})(1-v_{ij})} \right\} g(\theta_j|\tau) d\theta_j. \tag{7}$$

where \mathbf{U} is the $n \times N$ response matrix with u_{ij} as its elements; \mathbf{V} is defined, with respect to v_{ij} , in analogy to \mathbf{U} and u_{ij} ; $\xi = (\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_n)$ is the matrix for item parameters; $g(\theta_j|\tau)$ is a density function for examinees' ability, and τ is the vector containing the parameters of the examinee population ability distribution.

Zheng et al. (2017) proposed the EMM algorithm based on the new likelihood function derived from the reformulation with the g-process coming first (see **Appendix A**). The EMM can solve the convergence problem of MMLE/EM in a modest sample size of about 1,000 examinees, but it cannot eliminate occasional improbably large estimates. In fact, this is still a symptom of the 3PLM being possibly unidentifiable, though much more alleviated than in the original MMLE/EM, and this can be resolved by additional prior information which leads to the Bayesian EMM algorithms.

Please note that Béguin and Glas (2001) did not develop an EM algorithm similar to the EMM for the 3PNO because the integral in the ogive model introduces additional difficulty for the E-step which can be conveniently addressed by a MCMC algorithm. An EM algorithm and its variants are a much more natural alternative for the 3PLM which is the main topic of the current paper.

THE BAYESIAN EMM ALGORITHMS

Take the first reformulation as an example to illustrate how to derive the BEMM algorithm. Following Mislevy (1986)'s parameterization to take logarithmic form of a_i , the 3PLM can be rewritten as:

$$p_i(\theta_j) = c_i + (1 - c_i) \times P_i^*(\theta_j), \tag{8}$$

with

$$P_i^*(\theta_j) = \frac{1}{1 + \exp(-De^{\ln a_i}(\theta_j - b_i))}. \tag{9}$$

Mislevy (1986) also has given a general Bayesian formulation for the 3PLM which we may apply to the BEMM as well. Let ψ_i represents any item parameter for item i in ξ_i , and then the first derivative of the general Bayesian formulation for each item parameter can be obtained as:

$$\frac{\partial \ln L}{\partial \psi_i} + \frac{\partial \ln g(\psi_i|\eta)}{\partial \psi_i}, \tag{10}$$

with

$$\begin{aligned} L(\mathbf{U}, \mathbf{V}|\xi) &= \prod_{j=1}^N \int_{\theta_j} \left\{ \prod_{i=1}^n P_i^*(\theta_j)^{u_{ij}v_{ij}} \times [c_i(1 - P_i^*(\theta_j))]^{u_{ij}(1-v_{ij})} \right. \\ &\quad \left. \times [(1 - c_i)(1 - P_i^*(\theta_j))]^{(1-u_{ij})(1-v_{ij})} \right\} g(\theta_j|\tau) d\theta_j \end{aligned} \tag{11}$$

where L is the likelihood of the EMM (Equation 7) with the logarithmic form of a_i , and $g(\psi_i|\eta)$ is the item parameter prior distribution for item i . Mislevy (1986) suggested that $\ln a_i$ and b_i follow a normal distribution and c_i a beta distribution, specifically,

$$\ln a_i \sim N(\mu_{\ln a_i}, \sigma_{\ln a_i}^2), b_i \sim N(\mu_{b_i}, \sigma_{b_i}^2), c_i \sim \text{Beta}(\alpha_i, \beta_i) \quad (12)$$

in which $\mu_{\ln a_i}, \mu_{b_i}, \sigma_{\ln a_i}^2$, and $\sigma_{b_i}^2$ are the means and variances for the corresponding normal distribution and α_i, β_i are the parameters for the beta distribution for the guessing parameter. They may be specified as in the BILOG-MG default setting (Du Toit, 2003):

$$\ln a \sim N(0, 0.5^2), b \sim N(0, 2^2), c \sim \text{Beta}(4, 16). \quad (13)$$

The first and second derivatives for the three priors are given by Mislevy (1986) as:

$$\begin{aligned} \frac{\partial \ln g(\ln a_i | \mu_{\ln a_i}, \sigma_{\ln a_i}^2)}{\partial \ln a_i} &= -\frac{\ln a_i - \mu_{\ln a_i}}{\sigma_{\ln a_i}^2}, \quad \frac{\partial^2 \ln g(\ln a_i | \mu_{\ln a_i}, \sigma_{\ln a_i}^2)}{\partial \ln a_i \partial \ln a_i} = -\frac{1}{\sigma_{\ln a_i}^2}, \\ \frac{\partial \ln g(b_i | \mu_{b_i}, \sigma_{b_i}^2)}{\partial b_i} &= -\frac{b_i - \mu_{b_i}}{\sigma_{b_i}^2}, \quad \frac{\partial^2 \ln g(b_i | \mu_{b_i}, \sigma_{b_i}^2)}{\partial b_i \partial b_i} = -\frac{1}{\sigma_{b_i}^2}, \quad (14) \\ \frac{\partial \ln g(c_i | \alpha_i, \beta_i)}{\partial c_i} &= \frac{\alpha_i - 1}{c_i} - \frac{\beta_i - 1}{1 - c_i}, \quad \frac{\partial^2 \ln g(c_i | \alpha_i, \beta_i)}{\partial c_i \partial c_i} = -\frac{\alpha_i - 1}{c_i^2} - \frac{\beta_i - 1}{(1 - c_i)^2}. \end{aligned}$$

With the prior distribution component explained, the next will describe the Bayesian EMM method in which the likelihood component will be carefully delineated.

Expectation Step and Artificial Data

The expectation step boils down to the calculation of the conditional expectations of \mathbf{V} and θ . From the joint distribution in Equation (5), one can calculate the expectation of v_{ij} conditional on u_{ij} and the marginal distribution of v_{ij} . By the Bayesian rule,

$$\begin{aligned} P(v_{ij} = 1 | u_{ij} = 1, \theta_j, \xi_i) &= \frac{P_i^*(\theta_j)}{P_i(\theta_j)}, \\ P(v_{ij} = 1 | u_{ij} = 0, \theta_j, \xi_i) &= 0 \end{aligned} \quad (15)$$

can be yielded from Equations (5, 6). Then, the conditional expectation of v_{ij} is

$$E(v_{ij} | u_{ij}, \theta_j, \xi_i) = \frac{P_i^*(\theta_j)}{P_i(\theta_j)} \times u_{ij} + 0 \times (1 - u_{ij}). \quad (16)$$

As for θ , by using summation over a fixed grid of equally-spaced quadrature points X_k ($k = 1, 2, \dots, q$) with an associated weight $A(X_k)$ to approximate integration, one can have the quadrature form of the first derivative of the expected log-likelihood function for each item parameter (see Appendix B for detail):

$$\frac{\partial \ln E[L]}{\partial \psi_i} \approx \sum_{j=1}^N \sum_{k=1}^q \begin{bmatrix} u_{ij} E(v_{ij} | u_{ij}, X_k, \xi_i) \frac{\partial P_i^*(X_k)}{\partial \psi_i} \\ - \frac{P_i^*(X_k)}{1 - E(v_{ij} | u_{ij}, X_k, \xi_i)} \frac{\partial P_i^*(X_k)}{\partial \psi_i} \\ + \frac{u_{ij}(1 - E(v_{ij} | u_{ij}, X_k, \xi_i))}{c_i} \frac{\partial c_i}{\partial \psi_i} \\ - \frac{(1 - u_{ij})(1 - E(v_{ij} | u_{ij}, X_k, \xi_i))}{1 - c_i} \frac{\partial c_i}{\partial \psi_i} \end{bmatrix} P(X_k | \mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}, \boldsymbol{\xi}) \quad (17)$$

with

$$\begin{aligned} P(X_k | \mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}, \boldsymbol{\xi}) &= P(X_k | \mathbf{u}_j, \boldsymbol{\tau}, \boldsymbol{\xi}) = \frac{P(\mathbf{u}_j | X_k, \boldsymbol{\xi}) \times A(X_k)}{\sum_{k=1}^q P(\mathbf{u}_j | X_k, \boldsymbol{\xi}) \times A(X_k)}, \quad (18) \\ P(\mathbf{u}_j | X_k, \boldsymbol{\xi}) &= \prod_{i=1}^n P_i(X_k)^{u_{ij}} \times (1 - P_i(X_k))^{1 - u_{ij}}, \end{aligned}$$

where $P(X_k | \mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}, \boldsymbol{\xi})$ is the posterior probability of θ_j evaluated at X_k given $\mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}$, and $\boldsymbol{\xi}$. Then, $P(X_k | \mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}, \boldsymbol{\xi})$ equals $P(X_k | \mathbf{u}_j, \boldsymbol{\tau}, \boldsymbol{\xi})$ because $P(u_{ij} = 1 | \xi_i, \theta_j) = P(v_{ij} = 1, u_{ij} = 1 | \xi_i, \theta_j) = P_i^*(\theta_j)$. Furthermore, $P(X_k | \mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}, \boldsymbol{\xi})$ can be used to compute the ‘‘artificial data’’. The ‘‘artificial data’’ is essentially various expected frequencies of examinees under the posterior probability of θ_j and can be expressed as different linear combinations of the posterior probability of θ_j . Since it is ‘‘created’’ from the posterior probability, the IRT literature terms them as the ‘‘artificial data’’. For instance, Bock and Aitkin (1981) has provided two fundamental artificial data for traditional EM algorithm as:

$$\begin{aligned} \bar{f}_k &= \sum_{j=1}^N P(X_k | \mathbf{u}_j, \boldsymbol{\tau}, \boldsymbol{\xi}) = \sum_{j=1}^N P(X_k | \mathbf{u}_j, \mathbf{z}_j, \boldsymbol{\tau}, \boldsymbol{\xi}), \\ \bar{r}_{ik} &= \sum_{j=1}^N u_{ij} \times P(X_k | \mathbf{u}_j, \boldsymbol{\tau}, \boldsymbol{\xi}) = \sum_{j=1}^N u_{ij} \times P(X_k | \mathbf{u}_j, \mathbf{z}_j, \boldsymbol{\tau}, \boldsymbol{\xi}), \end{aligned} \quad (19)$$

in which \bar{f}_k is the expected number of examinees with ability X_k . Thus, the sum of \bar{f}_k for every ability point X_k equals the total number of examinees N , and \bar{r}_{ik} is the expected number of examinees with ability X_k answering item i correctly.

Then, as can be seen from Table 1, the EMM algorithm introduced a new latent variable \mathbf{V} , so there are two new artificial data as

$$\begin{aligned} \bar{f}_{ik}^{(V)} &= \bar{r}_{ik}^{(V)} \\ &= \sum_{j=1}^N E(v_{ij} | u_{ij}, X_k, \boldsymbol{\xi}) P(X_k | \mathbf{u}_j, \mathbf{v}_j, \boldsymbol{\tau}, \boldsymbol{\xi}), \end{aligned} \quad (20)$$

in which $\bar{f}_{ik}^{(V)}$ is the expected number of examinees with ability X_k who have employed their ability to answer item i and $\bar{r}_{ik}^{(V)}$ is the expected number of examinees with ability X_k who are able to answer it correctly. Please note that the expected number of examinees with ability X_k who have answered the item i based on their ability but incorrectly is zero, so $\bar{f}_{ik}^{(V)} = \bar{r}_{ik}^{(V)}$. Moreover, it is easy to obtain the expected number of examinees with ability X_k who have not employed their ability (in another words, used the guessing strategy) to answer item i , $\bar{f}_k - \bar{f}_{ik}^{(V)}$, and that of examinees among them who are able to answer it correctly, $\bar{r}_{ik} - \bar{r}_{ik}^{(V)}$.

After the E-step and calculation of the artificial data, the next steps are to compute the first and second derivatives of Equation (17) with respect to each item parameter.

Maximization Step-1 for c Parameter

From Equation (17), the first derivative for the guessing parameter is

$$\begin{aligned} \lambda_{c_i} &= \frac{\partial \ln E[L]}{\partial c_i} + \frac{\partial \ln g(c_i | \alpha_i, \beta_i)}{\partial c_i} \\ &\approx \frac{\alpha_i - 1}{c_i} - \frac{\beta_i - 1}{1 - c_i} + \sum_{j=1}^N \sum_{k=1}^q \left[\frac{u_{ij}(1 - E(v_{ij} | u_{ij}, X_k, \xi_i))}{c_i} - \frac{(1 - u_{ij})(1 - E(v_{ij} | u_{ij}, X_k, \xi_i))}{1 - c_i} \right] P(X_k | \mathbf{u}_j, \mathbf{z}_j, \boldsymbol{\tau}, \boldsymbol{\xi}) \\ &= \frac{\alpha_i - 1}{c_i} - \frac{\beta_i - 1}{1 - c_i} + \sum_{k=1}^q \frac{(\bar{r}_{ik} - \bar{r}_{ik}^{(V)})}{c_i} - \sum_{k=1}^q \frac{(\bar{f}_k - \bar{r}_{ik})}{1 - c_i} \end{aligned} \quad (21)$$

Set Equation (21) to 0 and solve for the estimate of c_i which leads to a closed solution. The derivation is as follows:

$$\begin{aligned} &\Rightarrow \frac{\alpha_i - 1 + \sum_{k=1}^q (\bar{r}_{ik} - \bar{r}_{ik}^{(V)})}{c_i} = \frac{\beta_i - 1 + \sum_{k=1}^q (\bar{f}_k - \bar{r}_{ik})}{1 - c_i} \\ &\Rightarrow c_i = \frac{\alpha_i - 1 + \sum_{k=1}^q (\bar{r}_{ik} - \bar{r}_{ik}^{(V)})}{\alpha_i + \beta_i - 2 + \sum_{k=1}^q (\bar{r}_{ik} - \bar{r}_{ik}^{(V)} + \bar{f}_k - \bar{r}_{ik})} \quad (22) \\ &= \frac{\alpha_i - 1 + \sum_{k=1}^q (\bar{r}_{ik} - \bar{r}_{ik}^{(V)})}{\alpha_i + \beta_i - 2 + \sum_{k=1}^q (\bar{f}_k - \bar{r}_{ik}^{(V)})} \end{aligned}$$

The estimate for the guessing parameter is contributed by two components: the prior and the data. The magnitude of the prior parameters α_i and β_i determines the influence of the prior through the two terms $(\alpha_i - 1)$ and $(\beta_i - 1)$. By ignoring the prior terms $(\alpha_i - 1)$ and $(\beta_i - 1)$, the data component of the estimate offers a very intuitive interpretation of the guessing parameter: It is calculated as the proportion of examinees who answer item i correctly using the guessing strategy in the total sample. This interpretation nicely fits into general philosophy of mixture modeling, drastically different from the traditional interpretation which is defined as the lower bound for the probability with which an examinee answers an item correctly.

Since an analytical solution can be easily obtained, the calculation of the corresponding second derivative and implementation of Newton-Raphson or Fisher-scoring algorithm, as in the traditional EM algorithm, are unnecessary. However, the second derivative is still useful for estimating SEs and is given below:

$$\begin{aligned} \lambda_{cc_i} &= E\left[\frac{\partial^2 \ln E[L]}{\partial c_i^2} | u_{ij}\right] + \frac{\partial^2 \ln g(c_i | \alpha_i, \beta_i)}{\partial c_i^2} \\ &\approx E\left[\frac{\partial}{\partial c_i} \sum_{k=1}^q \left[\frac{(\bar{r}_{ik} - \bar{r}_{ik}^{(V)})}{c_i} - \frac{(\bar{f}_k - \bar{r}_{ik})}{1 - c_i} \right] | u_{ij}\right] - \frac{\alpha_i - 1}{c_i^2} - \frac{\beta_i - 1}{(1 - c_i)^2} \quad (23) \\ &= E\left[-\sum_{k=1}^q \frac{(\bar{r}_{ik} - \bar{r}_{ik}^{(V)})}{c_i^2} - \frac{\sum_{k=1}^q (\bar{f}_k - \bar{r}_{ik})}{(1 - c_i)^2} | u_{ij}\right] - \frac{\alpha_i - 1}{c_i^2} - \frac{\beta_i - 1}{(1 - c_i)^2} \\ &= -\frac{\sum_{k=1}^q (1 - p_i^*(X_k)) \bar{f}_k}{c_i(1 - c_i)} - \frac{\alpha_i - 1}{c_i^2} - \frac{\beta_i - 1}{(1 - c_i)^2} \end{aligned}$$

since

$$\begin{aligned} E[u_{ij} | X_k] &= p_i(X_k) \\ E[E(v_{ij} | u_{ij}, X_k, \xi) | u_{ij}] &= E(v_{ij}) = p_i^*(X_k) \end{aligned} \quad (24)$$

and

$$\begin{aligned} E[\bar{f}_k | u_{ij}] &= E[\sum_{j=1}^N P(X_k | u_j, v_j, \tau, \xi) | u_{ij}] = \sum_{j=1}^N P(X_k | u_j, v_j, \tau, \xi) \\ &= \bar{f}_k, \\ E[\bar{r}_{ik} | u_{ij}] &= E[\sum_{j=1}^N u_{ij} P(X_k | u_j, v_j, \tau, \xi) | u_{ij}] \\ &= p_i(X_k) \sum_{j=1}^N P(X_k | u_j, v_j, \tau, \xi) \\ &= p_i(X_k) \bar{f}_k, \\ E[\bar{r}_{ik}^{(V)} | u_{ij}] &= E[\bar{r}_{ik}^{(V)} | u_{ij}] = E[\sum_{j=1}^N E(v_{ij} | u_{ij}, X_k, \xi) P(X_k | u_j, v_j, \tau, \xi) | u_{ij}] \\ &= p_i^*(X_k) \sum_{j=1}^N P(X_k | u_j, v_j, \tau, \xi) \\ &= p_i^*(X_k) \bar{f}_k. \end{aligned} \quad (25)$$

TABLE 1 | The definition of four kinds of artificial data.

Item i	$v_i = 1$	$v_i = 0$	Marginal of v_i
$u_i = 1$	$\bar{r}_{ik}^{(V)}$	$\bar{r}_{ik} - \bar{r}_{ik}^{(V)}$	\bar{r}_{ik}
$u_i = 0$	0	$\bar{f}_k - \bar{r}_{ik}$	
Marginal of u_i	$\bar{f}_k^{(V)}$	$\bar{f}_k - \bar{f}_k^{(V)}$	\bar{f}_k

Artificial data: The expected number of examinees.

Maximization Step-2 for a and b Parameters

The second Maximization step is to execute the Fisher-scoring procedure to obtain estimates for $\ln a_i$ and b_i . The required first derivatives for $\ln a_i$ and b_i are

$$\begin{aligned} \lambda_{a_i} &= D e^{\ln a_i} \sum_{k=1}^q [(\bar{r}_{ik}^{(V)} - \bar{f}_k \times p_i^*(X_k))(X_k - b_i)] - \frac{\ln a_i - \mu \ln a_i}{\sigma_{\ln a_i}^2}, \\ \lambda_{b_i} &= -D e^{\ln a_i} \sum_{k=1}^q [\bar{r}_{ik}^{(V)} - \bar{f}_k \times p_i^*(X_k)] - \frac{b_i - \mu_i}{\sigma_{b_i}^2}. \end{aligned} \quad (26)$$

The corresponding expectation of second derivatives are:

$$\begin{aligned} \lambda_{aa_i} &= -D^2 e^{2 \ln a_i} \sum_{k=1}^q (X_k - b_i)^2 w_{ik}^* \times \bar{f}_k - \frac{1}{\sigma_{\ln a_i}^2}, \\ \lambda_{bb_i} &= -D^2 e^{2 \ln a_i} \sum_{k=1}^q w_{ik}^* \times \bar{f}_k - \frac{1}{\sigma_{b_i}^2}, \\ \lambda_{ab_i} &= D^2 e^{2 \ln a_i} \sum_{k=1}^q [(X_k - b_i) w_{ik}^* \times \bar{f}_k], \end{aligned} \quad (27)$$

where

$$w_{ik}^* = p_i^*(X_k) \times (1 - p_i^*(X_k)), \quad (28)$$

which lead to the Fisher-scoring algorithm for the BEMM:

$$\begin{bmatrix} \ln a_i^{(t+1)} \\ b_i^{(t+1)} \end{bmatrix} = \begin{bmatrix} \ln a_i^{(t)} \\ b_i^{(t)} \end{bmatrix} - \begin{bmatrix} \lambda_{aa_i} & \lambda_{ab_i} \\ \lambda_{ab_i} & \lambda_{bb_i} \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{a_i} \\ \lambda_{b_i} \end{bmatrix}. \quad (29)$$

In this case, the estimation of the c parameter is separated from that of a and b , so the BEMM has a simplified 2-by-2 Hessian matrix (negative information matrix) in the iteration formulation.

To summarize, the flow chart of the BEMM has been given in **Figure 1**.

The same line of reasoning can be used to develop a BEMM for the second reformulation and the details are presented in **Appendix A**.

Standard Errors (SEs) of Parameter Estimation

Both the BEMM and the Bayesian EM are members of the EM family and it is well recognized that one drawback of the EM algorithms is that estimation SEs are not the natural products of their implementation, so we still need a practical method to obtain SEs (McLachlan and Krishnan, 2007). In detail, SEs can be calculated via the Fisher information matrix (Thissen and

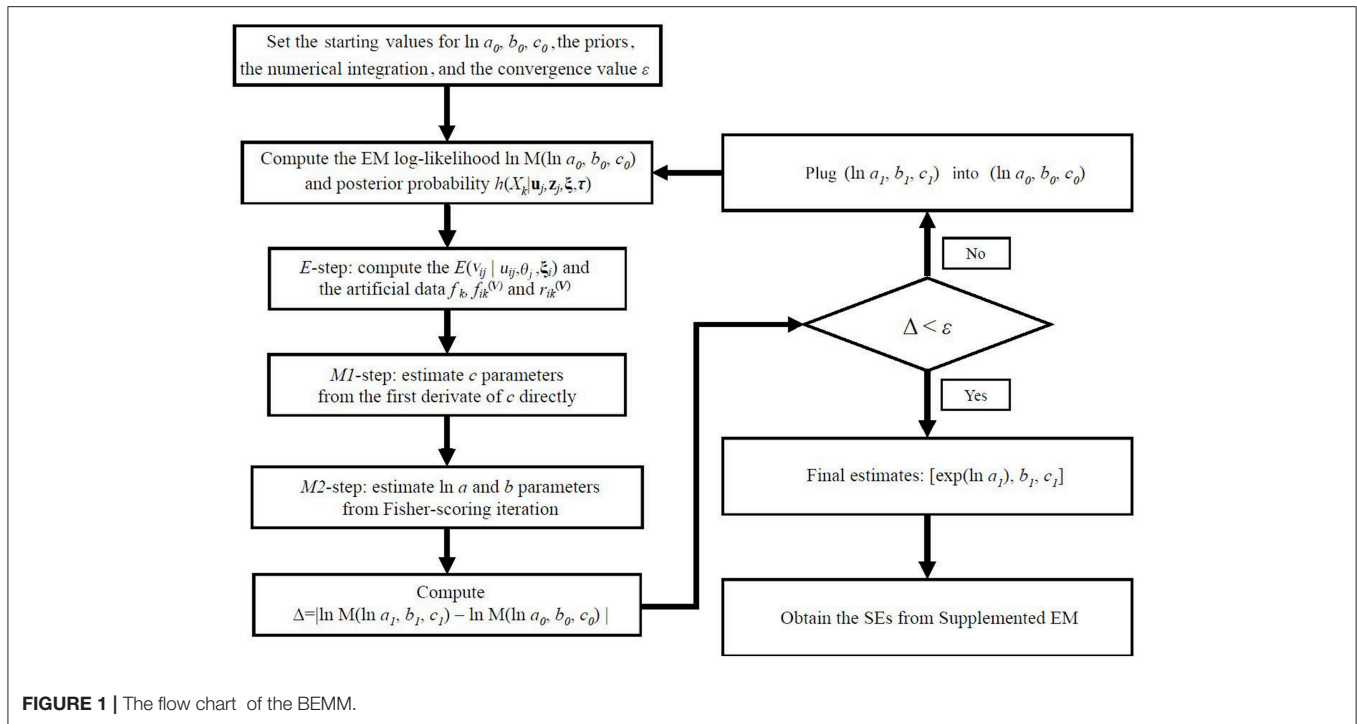


FIGURE 1 | The flow chart of the BEMM.

Wainer, 1982), empirical cross-product approximation (Jones and Geoffrey, 1992), the supplemented EM (SEM) method (Meng and Rubin, 1991; Cai, 2008), the forward difference method and the Richardson extrapolation method (Jamshidian and Jennrich, 2000), or sandwich covariance matrix (Kauermann and Carroll, 2001). Recently, the SEM method has been extended to various IRT models (Cai, 2008; Cai and Lee, 2009; Tian et al., 2013) and proved to be flexible enough to handle complex models in IRT. The current study, therefore, will focus on how to apply the SEM to the BEMM.

Cai and Lee (2009) has given a general SEM formulation for the large-sample covariance matrix as:

$$V(\hat{\xi}_i|Y) = \Phi^{-1}(\hat{\xi}_i|Y) = \Phi_c^{-1}(\hat{\xi}_i|Y)\{E_d - \Delta(\hat{\xi}_i)\}^{-1}, \quad (30)$$

where $\Phi^{-1}(\hat{\xi}_i|Y)$ is the inverse of item information matrix, E_d is the identity matrix with 3 dimensions, and $\Delta(\hat{\xi}_i)$ can be calculated from the Fisher-scoring execution. Please note that from Equation (10), in Bayesian approach (both Bayesian EM and Bayesian EMM), $\Phi^{-1}(\hat{\xi}_i|Y)$ involves not only the likelihood component but also the prior component. Refer to Cai and Lee (2009) for additional details. It is worth noting that the likelihood function for the BEMM algorithms are different the one for the 3PLM due to mixture modeling reformulation, so the SEs for the BEMM are different from those for the BME.

Firstly, the item information matrix of the BEMM can be obtained from Equations (23, 27) as:

$$\Phi_c(\hat{\xi}_i|Y) = \begin{bmatrix} -\lambda_{aa_i} & -\lambda_{ab_i} & 0 \\ -\lambda_{ab_i} & -\lambda_{bb_i} & 0 \\ 0 & 0 & -\lambda_{cc_i} \end{bmatrix}. \quad (31)$$

Then, the Δ matrix was also simplified due to the covariance between c and (a, b) equaling 0:

$$\Delta(\hat{\xi}_i) = \begin{bmatrix} \delta_{aa_i} & \delta_{ab_i} & 0 \\ \delta_{ab_i} & \delta_{bb_i} & 0 \\ 0 & 0 & \delta_{cc_i} \end{bmatrix}. \quad (32)$$

Finally, SEs of the Bayesian EMM can be obtained from:

$$(SE_{a_i}, SE_{b_i}, SE_{c_i}) = \sqrt{\text{diag}\{\Phi_c^{-1}(\hat{\xi}_i|Y)[E_3 - \Delta(\hat{\xi}_i)]^{-1}\}}. \quad (33)$$

Obviously, the covariance elements between the guessing parameter and other two parameters are not zero in the traditional MLE/EM and BEM, but due to the setup of two maximization steps in EMM, these elements can be legitimately set to zero. The zero covariance removes undesirable fluctuation in the item parameter estimation and thus makes the estimated SEs smaller than the counterparts, especially the guessing parameter, than in the Bayesian EM.

SIMULATION STUDY

The simulation study intends to demonstrate how different priors impact the Bayesian EMM algorithms, and compares it with the Bayesian EM in BILOG-MG under two different priors for c parameters. One prior in this study comes from BILOG-MG default setting, $c \sim \text{Beta}(4, 16)$ with $\mu = 0.2$, $\sigma^2 = 0.008$, the other is a more non-informative prior from the flexMIRT (Houts and Cai, 2015) default setting, $c \sim \text{Beta}(1, 4)$ with $\mu = 0.2$, $\sigma^2 = 0.027$.

To implement these algorithms, we developed a MATLAB toolbox, IRTEMM, to obtain the BEMM estimates. IRTEMM also offers several different options for estimating SEs including SEM.

Data Generation

Following Mislevy (1986)'s setting for data generation, the current study simulated the item parameters a , b , c for 10 and 20 items from an independent normal distribution follows $\ln a \sim N(0, 0.5)$ where $(0.3 \leq a \leq 2.5)$; $b \sim N(0, 1)$ where $(-3 \leq b \leq 3)$; and $\logit c \sim N(-1.39, 0.16)$. Three sample sizes of examinees (1,000, 1,500 and 2,000) were simulated from the standard normal distribution. For each condition, we ran 50 replications for each condition in the fully crossed 4 (two BEMM methods vs. two BILOG-MG methods) \times 3(1,000 vs. 1,500 vs. 2,000) \times 2(10 vs. 20) design.

Evaluation Criteria

The evaluation criteria for item parameter recovery are bias and the root mean squared error (RMSE), which are calculated as:

$$bias = \frac{\sum_{s=1}^{S=50} (\hat{\psi}_{is} - \psi_i)}{S}, RMSE = \sqrt{\frac{\sum_{s=1}^{S=50} (\hat{\psi}_{is} - \psi_i)^2}{S}}$$

Due to space constraint, the complete results have been summarized in **Appendix C** and only that for the condition of 1,000 examinees and 20 items here will be shown here since the others conditions presented a very similar pattern.

Results

As can be seen from **Table 2**: The biases across the four conditions for a and b parameters seem very similar, but for c parameters, these biases are highly influenced by priors. Both the absolute values of biases from the Bayesian EMM for c parameters are smaller than the Bayesian EM in BILOG-MG when changing priors. A more intuitive conclusion can be drawn from **Figure 2**: The RMSEs of c parameters from the right plot shows the Bayesian EMM has lower RMSE than the Bayesian EM in BILOG-MG. Furthermore, the difference of RMSEs produced by the Bayesian EMM between two prior conditions are much smaller than the Bayesian EM, which means the Bayesian EMM tends to be less affected by priors and yields more stable estimates than the Bayesian EM.

With the increasing of examinees from 1,000 to 2,000, the item parameter recovery of the Bayesian EMM and the Bayesian EM in BILOG-MG is obviously improved, and the difference between two priors for both methods is also decreased. However, some relatively large biases and RMSEs of the Bayesian EM still exist due to changing priors even in the largest sample size of the current study (e.g., 2,000 examinees and 20 items), while those from the Bayesian EMM are more stable (see **Appendix C** for details). In addition, the increase of the item number from 10 to 20 has no obvious influence for majority of item parameters.

Thus, it can be concluded that the Bayesian EMM inherits the advantages of both the EMM and the Bayesian method, and yields better estimates than the Bayesian EM. In other words, the Bayesian EMM has the most stable solutions among the two methods.

TWO EMPIRICAL EXAMPLES

Two empirical examples of different sample sizes and item numbers are given here to demonstrate feasibility of the BEMM in practice: The first dataset represents a case where the numbers of items and examinees are relatively large while the other small. The estimates of the BEMM and the EMM can be obtained from the MATLAB toolbox, IRTEMM. As for the Bayesian EM, in addition to BILOG-MG, we also use two of the most recent IRT programs flexMIRT (Houts and Cai, 2015) and IRTPRO (Cai et al., 2011) to carry out a cross-implementation validation. The complete results of two examples are summarized in the tables and figures in **Appendix D**. Only two figures (**Figure 3** for flexMIRT data and **Figure 4** for IRTPRO data) are presented here to compare with the results of the BEMM and BILOG-MG analyses.

The Dataset From flexMIRT

The first dataset is the flexMIRT example "g341-19.txt," which consists of the responses to 12 items from 2,844 examinees. This example aimed at demonstrating whether the BEMM is more robust against the differential effects of priors than the Bayesian EM. The default settings for guessing priors in BILOG-MG and flexMIRT [$c \sim Beta(4, 16)$ vs. $c \sim Beta(1, 4)$] have been applied to the BEMM and both software implementations.

The results are presented in **Figure 3**. The estimates for both a and b parameters are in each other's proximity for all different implementations. The estimates for c parameters present non-negligible divergence for different guessing priors for the BILOG-MG and flexMIRT: the estimates of these implementations cluster in two groups based on the guessing prior setting. The BEMM, in contrast, produced almost identical estimates under different prior setting. This cross-software validation shows that (1) the divergence in point estimates is possibly inherent in the Bayesian EM algorithm, but not due to different software executions; (2) the BEMM algorithm can provide stable point estimates that are robust against change in priors; and (3) correspondingly, there is noticeable difference in SEs for different priors for BILOG-MG and flexMIRT while there is no such difference for the BEMM.

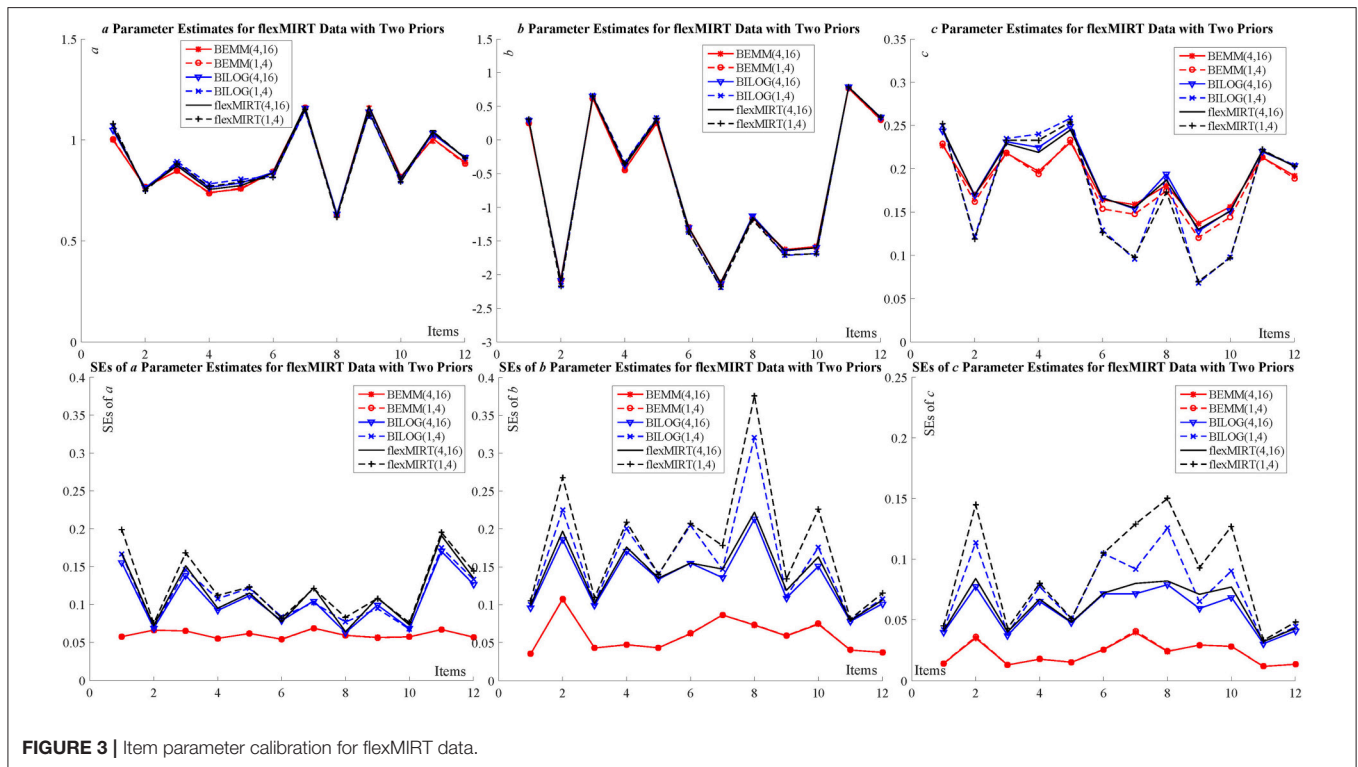
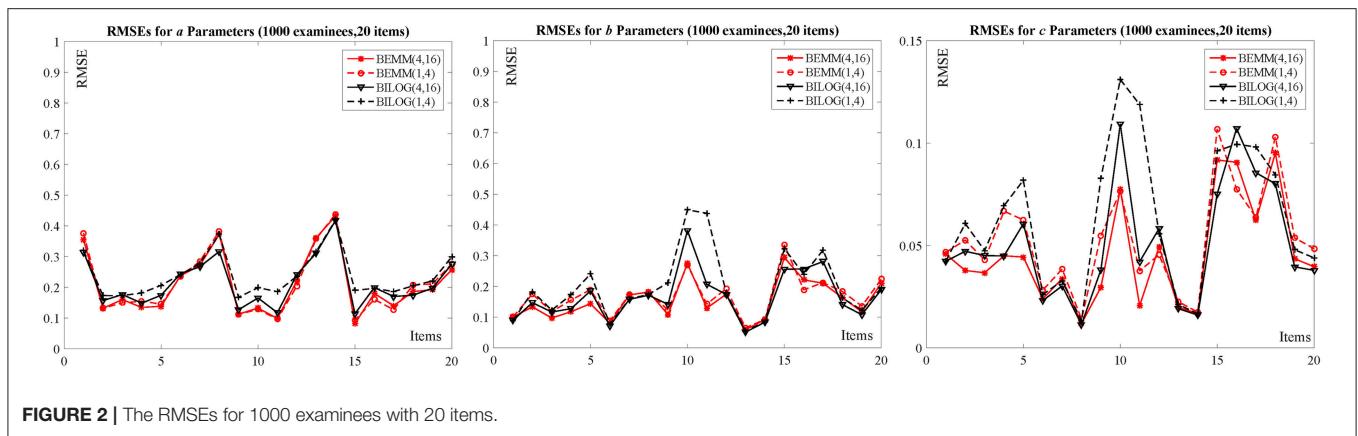
The Dataset From IRTPRO

The dataset of the second example is from the IRTPRO example "lsat6.csv," which consists of responses to 5 items from 1,000 examinees. The dataset originally came from the Law School Admissions Test Section 6 (LSAT6) and has been widely used as an example in the item response theory (Bock and Lieberman, 1970; McDonald, 1999; Du Toit, 2003; Chalmers, 2012). More importantly, this dataset presents a case with realistically small number of examinees in educational testing scenarios and Bayesian EMM's performance with this dataset testify its applicability in practice. Following the IRTPRO default setting, the priors for the a and b parameters as: $\ln a \sim N(0, 1)$ and $b \sim N(0, 3)$.

Figure 4 shows that, all of the item point estimates for a and b parameters among seven conditions which are very similar and relatively low ($0.40 \leq a \leq 0.77$; $-3.18 \leq b \leq 0.37$). In this adverse situation, the c parameters estimated by the Bayesian EM

TABLE 2 | The Biases for item parameter estimates with 1000 examinees and 20 items.

Item	Generating			Biases for a			Biases for b			Biases for c				
	Bayesian EMM			BILOG-MG			Bayesian EMM			BILOG-MG				
	a	b	c	Beta(4, 16)	Beta(1, 4)	Beta(4, 16)	Beta(1, 4)	Beta(4, 16)	Beta(1, 4)	Beta(4, 16)	Beta(1, 4)	Beta(4, 16)	Beta(1, 4)	
1	1.547	0.540	0.292	0.051	0.092	0.051	0.089	-0.044	-0.033	-0.012	-0.019	-0.012	-0.015	-0.008
2	0.605	0.705	0.182	0.089	0.074	0.120	0.128	-0.017	-0.055	0.041	0.009	0.041	0.028	0.026
3	0.692	0.592	0.143	0.113	0.081	0.135	0.117	0.036	-0.015	0.045	0.023	0.045	0.036	0.023
4	0.958	-0.160	0.216	0.044	0.043	0.068	0.086	0.002	-0.012	0.051	-0.012	0.051	0.004	0.007
5	0.581	0.393	0.189	0.089	0.082	0.127	0.148	0.016	-0.018	0.096	0.011	0.110	0.035	0.040
6	1.477	0.909	0.151	0.131	0.101	0.144	0.121	-0.045	-0.053	-0.035	-0.002	-0.035	0.001	-0.003
7	1.076	1.634	0.232	0.040	0.039	0.055	0.062	-0.132	-0.133	-0.111	-0.018	-0.111	-0.014	-0.014
8	1.548	2.107	0.106	0.018	-0.028	-0.008	-0.028	-0.121	-0.116	-0.099	-0.002	-0.099	-0.001	-0.004
9	0.911	-0.791	0.189	0.061	0.050	0.079	0.098	0.058	0.027	0.098	-0.006	0.106	0.014	0.014
10	0.480	0.064	0.123	0.107	0.096	0.142	0.165	0.227	0.178	0.329	0.068	0.351	0.098	0.104
11	0.480	-0.547	0.195	0.071	0.065	0.090	0.144	0.055	0.015	0.153	-0.002	0.331	0.029	0.081
12	0.509	1.533	0.126	0.182	0.148	0.207	0.192	-0.042	-0.077	0.002	0.042	-0.017	0.053	0.045
13	2.081	0.886	0.122	0.093	0.042	0.093	0.054	-0.036	-0.044	-0.020	-0.001	-0.027	0.001	-0.003
14	2.234	1.458	0.198	-0.269	-0.276	-0.336	-0.335	-0.057	-0.057	-0.040	-0.005	-0.042	-0.004	-0.005
15	0.651	0.229	0.294	-0.020	-0.017	0.036	0.107	-0.246	-0.254	-0.107	-0.082	-0.001	-0.039	-0.006
16	0.872	-0.539	0.075	0.143	0.112	0.161	0.140	0.203	0.146	0.237	0.082	0.190	0.098	0.070
17	0.532	0.152	0.128	0.122	0.104	0.152	0.157	0.148	0.089	0.221	0.049	0.207	0.072	0.066
18	1.319	-0.277	0.310	-0.105	-0.080	-0.068	0.002	-0.119	-0.106	-0.076	-0.084	-0.024	-0.063	-0.039
19	1.141	0.558	0.227	0.062	0.063	0.081	0.092	-0.048	-0.054	-0.024	-0.017	-0.023	-0.010	-0.009
20	0.666	1.510	0.233	0.133	0.137	0.173	0.195	-0.123	-0.130	-0.078	-0.003	-0.074	0.009	0.013



are obviously unstable and they are highly affected by priors. However, the Bayesian EMM estimates under two priors are comparatively accordant and very close to the MLE solutions. As regards SEs, the four lines of BILOG-MG and IRTPRO can be divided into two groups according to their priors, while there is no obvious difference in both results of the Bayesian EMM.

To summarize, these two examples illustrate that the Bayesian EMM was less affected by priors since they take the full advantage of the EMM and the Bayesian method.

DISCUSSION AND FUTURE DIRECTIONS

Based on the results of the simulation study and real-world examples, the conclusions can be summarized as follows: (1)

the BEMM can yield at least comparable or even better item estimates than the Bayesian EM; (2) the BEMM is less sensitive to change in item priors than the Bayesian EM, despite of the implementations; both point estimates and SEs, especially for the guessing parameters, are subjected to less fluctuation than BILOG-MG, flexMIRT and IRTPRO when different priors are used.

Obviously, the BEMM takes full advantage of the EMM and the Bayesian approach. On the one hand, the EMM itself is a more powerful MMLE method than the EM, so the BEMM can explore the likelihood function as thoroughly as the EMM before turning to priors to “shrink” the estimates; on the other hand, the Bayesian approach can naturally be used to solve the issue of estimate inflation for some troublesome items even when the EMM cannot produce reasonable MLE estimates. The simulation

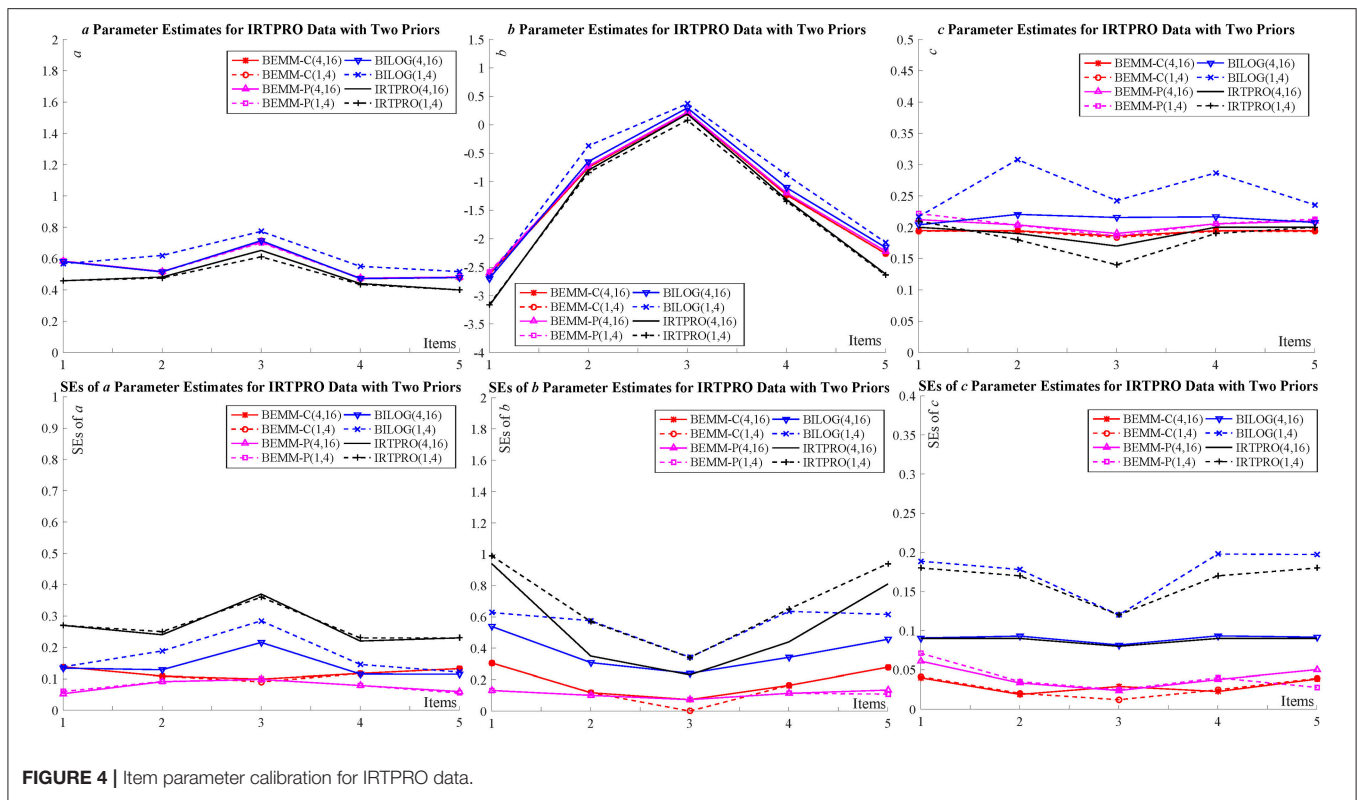


FIGURE 4 | Item parameter calibration for IRTPRO data.

study and the two real dataset examples are of limited scopes, so the conclusions based on their results should be interpreted with caution. It is not the intent of this paper to advocate for the elimination of the usage of other methods. The BEMM can be used to check with the Bayesian EM in the IRT programs in practice. Due to the high complexity in real-world 3PLM data, a combination of the Bayesian EM in different implementations, the BEMM and even the naked MLE solution, the EMM, might lead to a more sophisticated and nuanced understanding of data.

Several questions deserve further attention. Firstly, the BEMM can be readily extended to other IRT models with guessing effect. A case in point is the IRT model with covariates model (Tay et al., 2013). According to Tay et al. (2016), it needs at least a sample of 20,000 examinees to fit a 3PLM with covariates successfully. In this case, the BEMM may offer a better alternative to reducing the required sample sizes for the 3PLM with covariates.

Secondly, the mixture modeling approach and the BEMM can be naturally accommodated for the 4PLM (Barton and Lord, 1981). There is a renewed interest in the 4PLM (Rulison and Loken, 2009; Loken and Rulison, 2010; Liao et al., 2012; Ogasawara, 2012; Feuerstahler and Waller, 2014; Culpepper, 2015) for its usefulness in achievement test and psychological datasets. But, the estimation challenge posed by the 4PLM is greater than the 3PLM due to the additional slipping parameter

in the model (Loken and Rulison, 2010; Culpepper, 2015, 2017; Waller and Feuerstahler, 2017). Adaption of the BEMM for the 4PLM is a promising direction.

AUTHOR CONTRIBUTIONS

SG: software and implementation, mathematical derivation, simulation studies and examples, responding, writing—revision, writing—review, editing, methodology, data curation, and validation. CZ: formal analysis, methodology, conceptualization, funding acquisition, investigation, software, project administration, resources, supervision, writing—original draft, writing—review, and editing.

FUNDING

This research was funded by the Peak Discipline Construction Project of Education at East China Normal University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01175/full#supplementary-material>

REFERENCES

Baker, F. B., and Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques, 2nd Edn.* New York, NY: CRC Press.

Barton, M. A., and Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item response model. *ETS Res. Rep. Ser.* 1981, 1–8. doi: 10.1002/j.2333-8504.1981.tb01255.x

- Béguin, A. A., and Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika* 66, 541–561. doi: 10.1007/BF02296195
- Birnbaum, A. (1968). “Some latent trait models and their use in inferring an examinee’s ability,” in *Statistical Theories of Mental Test Scores*, eds F. M. Lord and M. R. Novick (Boston, MA: Adison-Wesley), 395–479.
- Bock, R. D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* 46, 443–459. doi: 10.1007/BF02293801
- Bock, R. D., and Lieberman, M. (1970). Fitting a response model form dichotomously scored items. *Psychometrika* 35, 179–197. doi: 10.1007/BF02291262
- Bock, R. D., and Schilling, S. (2003). “IRT based item factor analysis,” in *IRT From Ssi: Bilog-mg, Multilog, Parscale, Testfact*, ed. M. Du Toit (Skokie, IL: Scientific Software International), 584–591.
- Cai, L. (2008). SEM of another flavour: two new applications of the supplemented EM algorithm. *Br. J. Math. Stat. Psychol.* 61, 309–329. doi: 10.1348/000711007X249603
- Cai, L., and Lee, T. (2009). Covariance structure model fit testing under missing data: an application of the supplemented EM algorithm. *Multivariate Behav. Res.* 44, 281–304. doi: 10.1080/00273170902794255
- Cai, L., Thissen, D., and du Toit, S. (2011). *IRTPRO4 User’s Guide*. Skokie, IL: Scientific Software International. Retrieved from: <http://ssicentral.com/irt/index.html>
- Cao, J., and Stokes, S. L. (2008). Bayesian IRT guessing models for partial guessing behaviors. *Psychometrika* 73, 209–230. doi: 10.1007/s11336-007-9045-9
- Chalmers, R. P. (2012). mirt: a multidimensional item response theory package for the R environment. *J. Stat. Softw.* 48, 1–29. doi: 10.18637/jss.v048.i06
- Cui, S., Pan, K., and Ye, Y. (2018). Language ability or personality works? the return to possessing a global english test certificate for college graduates in China. *ECNU Rev. Educ.* 1, 74–101. doi: 10.30926/ecnuoe2018010204
- Culpepper, S. A. (2015). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika* 81, 1142–1163. doi: 10.1007/s11336-015-9477-6
- Culpepper, S. A. (2017). The prevalence and implications of slipping on low-stakes, large-scale assessments. *J. Educ. Behav. Stat.* 42, 706–725. doi: 10.3102/1076998617705653
- Du Toit, M. (2003). *IRT From Ssi: Bilog-mg, Multilog, Parscale, Testfact*. Skokie, IL: Scientific Software International.
- Feuerstahler, L. M., and Waller, N. G. (2014). Estimation of the 4-parameter model with marginal maximum likelihood. *Multivariate Behav. Res.* 49, 285–285. doi: 10.1080/00273171.2014.912889
- Fox, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. Berlin: Springer Science and Business Media.
- Houts, C. R., and Cai, L. (2015). *JflexMIRT: Flexible Multilevel Multidimensional Item Analysis and Test Scoring User’s Manual Version 3.0 RC*. North Carolina, CA. Retrieved from: <https://www.vpgcentral.com/software/irt-software/>
- Hutchinson, T. P. (1991). *Ability, Partial Information, Guessing: Statistical Modelling Applied to Multiple-Choice Tests*. Rundle Mall, SA: Rumsby Scientific Publishing.
- Jamshidian, M., and Jennrich, R. I. (2000). Standard errors for EM estimation. *J. R. Stat. Soc.* 62, 257–270. doi: 10.1111/1467-9868.00230
- Jones, P. N., and Geoffrey, M. J. (1992). Improving the convergence rate of the EM algorithm for a mixture model fitted to grouped truncated data. *J. Stat. Comput. Simul.* 43, 31–44. doi: 10.1080/00949659208811426
- Kauermann, G., and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *J. Am. Stat. Assoc.* 96, 1387–1396. doi: 10.1198/016214501753382309
- Liao, W.-W., Ho, R.-G., Yen, Y.-C., and Cheng, H.-C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. *Soc. Behav. Pers.* 40, 1679–1694. doi: 10.2224/sbp.2012.40.10.1679
- Loken, E., and Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *Br. J. Math. Stat. Psychol.* 63, 509–525. doi: 10.1348/000711009X474502
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Routledge.
- Maris, G., and Bechger, T. (2009). On interpreting the model parameters for the three parameter logistic model. *Measurement* 7, 75–88. doi: 10.1080/15366360903070385
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. London: Psychology Press.
- McLachlan, G., and Krishnan, T. (2007). *The EM Algorithm and Extensions, 2nd Edn*. New York, NY: John Wiley and Sons.
- Meng, X.-L., and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *J. Am. Stat. Assoc.* 86, 899–909. doi: 10.1080/01621459.1991.10475130
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika* 51, 177–195. doi: 10.1007/BF02293979
- Ogasawara, H. (2012). Asymptotic expansions for the ability estimator in item response theory. *Comput. Stat.* 27, 661–683. doi: 10.1007/s00180-011-0282-0
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer.
- Rulison, K. L., and Loken, E. (2009). I’ve fallen and I can’t get up: can high-ability students recover from early mistakes in CAT? *Appl. Psychol. Meas.* 33, 83–101. doi: 10.1177/0146621608324023
- Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika* 39, 111–121. doi: 10.1007/BF02291580
- San Martín, E., Del Pino, G., and De Boeck, P. (2006). IRT models for ability-based guessing. *Appl. Psychol. Meas.* 30, 183–203. doi: 10.1177/0146621605282773
- Sympson, J. B. (1978). “A model for testing with multidimensional items,” in *Paper Presented at The Proceedings of The 1977 Computerized Adaptive Testing Conference* (Minneapolis, MN: University of Minnesota, Department of Psychology, Psychometrics Methods Program).
- Tay, L., Huang, Q., and Vermunt, J. K. (2016). Item response theory with covariates (IRT-C) assessing item recovery and differential item functioning for the three-parameter logistic model. *Educ. Psychol. Meas.* 76, 22–42. doi: 10.1177/0013164415579488
- Tay, L., Vermunt, J. K., and Wang, C. (2013). Assessing the item response theory with covariate (IRT-C) procedure for ascertaining differential item functioning. *Int. J. Test.* 13, 201–222. doi: 10.1080/15305058.2012.692415
- Thissen, D., and Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika* 47, 397–412. doi: 10.1007/BF02293705
- Tian, W., Cai, L., Thissen, D., and Xin, T. (2013). Numerical differentiation methods for computing error covariance matrices in item response theory modeling: an evaluation and a new proposal. *Educ. Psychol. Meas.* 73, 412–439. doi: 10.1177/0013164412465875
- von Davier, M. (2009). Is there need for the 3PL model? Guess what? *Measurement* 7, 110–114. doi: 10.1080/15366360903117079
- Waller, N. G., and Feuerstahler, L. M. (2017). Bayesian modal estimation of the four-parameter item response model in real, realistic, and idealized data sets. *Multivariate Behav. Res.* 52, 350–370. doi: 10.1080/00273171.2017.1292893
- Woods, C. M. (2008). Consequences of ignoring guessing when estimating the latent density in item response theory. *Appl. Psychol. Meas.* 32, 371–384. doi: 10.1177/0146621607307691
- Zheng, C., Meng, X., Guo, S., and Liu, Z. (2017). Expectation-maximization-maximization: a feasible MLE algorithm for the three-parameter logistic model based on a mixture modeling reformulation. *Front. Psychol.* 8:2302. doi: 10.3389/fpsyg.2017.02302
- Zimowski, M. F., Muraki, E., Mislevy, R. J., and Bock, R. D. (2003). *Bilog-mg3 User’s Guide*. Mooresville, IN: Scientific Software International.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Guo and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.