Expectation-Maximization-Maximization: A Feasible MLE Algorithm for the Three-Parameter Logistic Model Based on a Mixture Modeling Reformulation

Chanjin Zheng[1], Xiangbin Meng[2,3]*, Shaoyang Guo[1] and Zhengguang Liu[4]

[1] School of Psychology, Jiangxi Normal University, Nanchang, China, [2] Faculty of Education, Northeast Normal University, Changchun, China, [3] Key Laboratory of Applied Statistics of Moe, Northeast Normal University, Changchun, China, [4] National Cooperative Innovation Center for Assessment and Improvement of Basic Education Quality, Beijing Normal University, Beijing, China

Stable maximum likelihood estimation (MLE) of item parameters in 3PLM with a modest sample size remains a challenge. The current study presents a mixture-modeling approach to 3PLM based on which a feasible Expectation-Maximization-Maximization (EMM) MLE algorithm is proposed. The simulation study indicates that EMM is comparable to the Bayesian EM in terms of bias and RMSE. EMM also produces smaller standard errors (SEs) than MMLE/EM. In order to further demonstrate the feasibility, the method has also been applied to two real-world data sets. The point estimates in EMM are close to those from the commercial programs, BILOG-MG and flexMIRT, but the SEs are smaller.

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

1. INTRODUCTION

In the field of educational measurement, item response theory (IRT) models are a powerful tool aimed at providing an appropriate representation of students' test-taking behavior, and produce accurate estimates of students' ability. IRT models are expected to capture the underlying response processes such as students' ability and other strategies students might take. One of the most common strategies is guessing behavior when students cannot solve a problem correctly. The guessing strategy is prevalent particularly for multiple-choice questions in a low-stakes test (Lord, 1980; Baker and Kim, 2004; Cao and Stokes, 2008; Woods, 2008). To count for guessing, the three-parameter logistic model (3PLM; Birnbaum, 1968) has been commonly used in many applications of IRT in the measurement industry.

Despite of its popularity, however, several studies have pointed out technical and theoretical issues regarding the *c*-parameter (the guessing parameter) and its interpretation (Lord, 1974, 1980; Kolen, 1981; Thissen and Wainer, 1982; Holland, 1990; Hambleton et al., 1991; Yen et al., 1991; San Martín et al., 2006, 2013, 2015; Woods, 2008; Maris and Bechger, 2009; McDonald, 2013). The current paper focuses on one of long-standing issues for the 3PLM, the item parameter estimation, which has proved to be challenging (Thissen and Wainer, 1982). Mislevy (1986) pointed out that the essential difficulty is the sparse data for the guessing parameter, yielding unstable maximum likelihood estimates (MLEs). The well-established marginal maximum likelihood estimation via expectation-maximization (MMLE/EM) algorithm (Bock and Aitkin, 1981) is not feasible in this

case. According to Thissen and Wainer (1982), the sample size required for obtaining MLE for 3PLM with acceptable standard errors using MMLE/EM is about 10,000 and, for some items, as large as 1,000,000 which seems prohibitively expensive or impractical even now. A most recent empirical study testifying to this claim is Tay et al. (2016) in which the researchers recommended that a sample of 20,000 is desirable when fitting a 3PLM with covariate (Tay et al., 2013). In fact, Thissen and Wainer (1982) even concluded that "naked maximum likelihood estimation for the three-parameter model is not a technique that is likely to give happy results." San Martín et al. (2015) showed that the 3PLM is not even identifiable if a fixed-effects approach is adopted and the question of identifiability of the random-effects 3PLM is still open.

By now only some decent Bayesian methods have been developed, such as Bayesian EM (Mislevy, 1986), Bayesian joint estimation (Swaminathan and Gifford, 1986), and MCMC (Patz and Junker, 1999) MLE, however, enjoys favorable statistical properties (Casella and Berger, 2002). It may avoid some issues in Bayesian methods, such as assigning priors, convergence checking in MCMC, etc. and, thus, usually is preferred for parameter estimation. It would be a valuable addition to existing methods if that a practical feasible MLE algorithm could be developed for the 3PLM. The current study intends to make some contribution in this respect. The authors propose a feasible EM algorithm for the 3PLM, namely expectation-maximization-maximization (EMM). EMM can be considered as a modified version of the MMLE/EM algorithm (Bock and Aitkin, 1981) and the extra maximization step is especially designed for the guessing parameter due to a different setup for the complete data based on a mixture-modeling reformulation of the 3PLM.

This mixture-modeling approach to the 3PLM is not entirely new in the IRT literature. Mixture modeling is a well-established tradition in IRT, especially for Rasch models (von Davier and Rost, 2006). As for the 3PLM in particular, Hutchinson (1991) first discussed the two underlying processes of guessing and answering based on ability which points to the idea of mixture modeling for the 3PLM. San Martín et al. (2006) proposed an ability-based guessing model to describe the interaction between guessing behavior and examinee's ability based on this perspective. von Davier (2009) further presented two different interpretations of the 3PLM from this standpoint. One may even easily notice that the reformulation of the 3PLM developed by the current study can be considered as a special case of the HTBRID model (Yamamoto, 1982) and the general diagnostic model (von Davier, 2008), both of which are of a strong mixture modeling flavor. Motivated by this tradition, especially the work by von Davier (2009) and Hutchinson (1991), the current study presents a new mixture-modeling reformulation of the 3PLM by introducing an extra latent indicator for the guessing behavior and develops a feasible MLE algorithm, although this concept of mixture-modeling approach for the 3PLM has recurred in the IRT literature.

More specifically, a conceptual summary of developing such a new algorithm goes as: (a) introducing a new latent variable to construct a space one dimension higher than the old one, which appears to be unwise because the reformulation makes the original 3PLM estimation problem more difficult by adding one more dimension. (b) invoking the independent assumption of guessing and problem-solving process to approximate the joint distribution of the response and the newly introduced latent variable. It is worth noting that this practice is very common in statistics to approximate a high-dimension space. (c) using the approximation as a surrogate of the original 3PLM likelihood function to obtain item parameter estimate. Since the goal is to calibrate items with the traditional 3PLM, the convergence criterion is still calculated through the likelihood function of the original 3PLM, though the E and M steps are involved with the approximation of the likelihood function of the reformulated model.

The remaining is organized as the follows: Section 1 will present the reformulation of 3PLM based on the mixture modeling (McLachlan and Peel, 2004). In next section, EMM is developed and derivation of the estimation standard errors (SEs) for EMM is provided. In section 3, a comprehensive simulation study is conducted and the EMM is also applied to two empirical examples from BILOG-MG (Zimowski et al., 2003) and flexMIRT (Houts and Cai, 2015). The last section gives a brief discussion and future directions.

2. A MIXTURE-MODELING APPROACH TO 3PLM

Mixture modeling is a powerful statistical tool for representing the presence of heterogeneous subpopulations within an overall population. The generic density form of mixture modeling for a random vector Y can be written as

$$f(y) = \sum_{i=1}^{g} \pi_i f_i(y), \tag{1}$$

in which the quantities $\pi_1, \ldots, \pi_g$ are called the mixing proportions or weights for the g groups and the functions $f_1(y), \ldots, f_g(y)$ are called the component densities of the mixture. Obviously, mixture modeling overcomes the limitation of traditional modeling approach using one single density and attempts to approximate data by a linear combination of possibly various different densities. The idea of mixture modeling has been applied in psychometrics. Various Rasch mixture IRT models have been proposed to model different response styles and test taking strategies (Rost, 1997; von Davier and Rost, 2006). This paper takes advantage of this idea to reformulate 3PLM. If guessing was considered as a test-taking strategy, with introduction of a latent indicator variable for guessing, the 3PLM could be reformulated as a mixture model for two heterogeneous subpopulations: those who guess and those who do not, within each item. The detail derivation is presented below.

The 3PLM is formulated as the follow:

$$P\left(u_{ij} = 1|\theta_j, a_i, b_i, c_i\right) \equiv P_i(\theta_j) = c_i \times [1] + (1 - c_i) \times \left[P_i^*(\theta_j)\right], \tag{2}$$

with

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

in which $u_{ij} \in \mathbf{U}$ is examinee j $(j = 1, 2, \ldots, N)$'s response to item i $(i = 1, 2, \ldots, n)$, a_i, b_i, and c_i are the item parameters; $\theta_j \in \theta$ is the ability parameter of the examinee j; D is the scaling constant 1.702. The 3PLM can be considered as a mixture of a degenerate distribution (in which all the probability mass is on a single point) with a probability of c_i and a 2PLM with a probability of $1 - c_i$ within each item. Mixture models with a degenerate model is termed as nonstandard mixture models and has been studied intensively in statistics (Cornell University Library, 1989).

The observed data in a mixture problem is usually viewed as incomplete. A latent indicator variable $z_{ij} \in \mathbf{Z}$ for guessing is introduced:

$$z_{ij} = \begin{cases} 1 \text{ if examinee } j \text{ does not guess on item } i; \\ 0 \text{ if examinee } j \text{ does guess on item } i. \end{cases}$$

Thus, the marginal of z_{ij} follows a Bernoulli $(1 - c_i)$. And the conditional distribution of u_{ij} on z_{ij} is as the follow (let $\xi_i = \{a_i, b_i, c_i\} \in \xi$ represents the item parameter vector for the ith item):

$$P(u_{ij} = 1 | z_{ij} = 1, \theta_j, \xi_i) = P_i^*(\theta_j), \quad P(u_{ij} = 1 | z_{ij} = 0, \theta_j, \xi_i) = 1,$$
$$P(u_{ij} = 0 | z_{ij} = 1, \theta_j, \xi_i) = 1 - P_i^*(\theta_j), \quad P(u_{ij} = 0 | z_{ij} = 0, \theta_j, \xi_i) = 0.$$
(4)

To simplify the derivation of the joint distribution of u_{ij} and z_{ij}, z_{ij} and θ_j are assumed to be independent which leads to:

$$P(u_{ij}, z_{ij} | \theta_j, \xi_i) = P(u_{ij} | z_{ij}, \theta_j, \xi_i) P(z_{ij}). \tag{5}$$

More specifically,

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

Please note that the case of $u_{ij} = 0$ and $z_{ij} = 0$, whose probability is zero, is redundant, so the other three cases consist of the three elements of the joint distribution. The assumption of independence means that, for each item, the examinee decides randomly whether guessing or ability-based responding is chosen first (San Martín et al., 2006; von Davier, 2009). Let $\mathbf{u}_j$ and $\mathbf{z}_j$ be the response vector and the latent indicator vector for examinee j, respectively, and then the joint distribution for the new augmented complete data $(\mathbf{U}, \mathbf{Z}, \theta)$ is

$$P(\mathbf{u}_j, \mathbf{z}_j, \theta_j | \xi, \tau) = P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi) g(\theta_j | \tau), \tag{7}$$

where

$$P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi) = \prod_{i=1}^{n} \left[(1 - c_i) P_i^*(\theta_j) \right]^{u_{ij} z_{ij}} \times c_i^{u_{ij}(1 - z_{ij})}$$
$$\times \left[(1 - c_i)(1 - P_i^*(\theta_j)) \right]^{(1 - u_{ij}) z_{ij}}, \tag{8}$$

and $g(\theta_j | \tau)$ is a density function for θ, where τ is the vector containing the parameters of the examinee population ability

distribution. Following Bock and Lieberman (1970), the marginal distribution for a single examinee j is:

$$P(\mathbf{u}_j, \mathbf{z}_j | \xi) = \int_{\theta_j} P(\mathbf{u}_j, \mathbf{z}_j | \xi, \theta_j) g(\theta_j | \tau) d\theta_j. \tag{9}$$

So, the likelihood function of the EMM is:

$$L(\mathbf{U}, \mathbf{Z} | \xi) = \prod_{j=1}^{N} P(\mathbf{u}_j, \mathbf{z}_j | \xi)$$
$$= \prod_{j=1}^{N} \int_{\theta_j} P(\mathbf{u}_j, \mathbf{z}_j | \xi, \theta_j) g(\theta_j | \tau) d\theta_j. \tag{10}$$

and the log-likelihood $\ln L = \ln \left(L(\mathbf{U}, \mathbf{Z} | \xi) \right)$.

3. THE EXPECTATION-MAXIMIZATION-MAXIMIZATION (EMM) ALGORITHM

The reformulation of 3PLM points to the possibility that the guessing parameter can be estimated as a mixing parameter in mixture models, separated from the item discriminatory and difficulty parameters which will be estimated as the unknown parameters in the component density (the 2PLM). Put this in the language of the EM for IRT models and one has the Expectation (with respect to the latent variables $\Theta, \mathbf{Z}$)—Maximization (with respect to the guessing parameter (c_i)—Maximization (with respect to the item discrimination and difficulty parameter a_i and b_i). The first expectation step follows the same idea as the EM for IRT (Bock and Aitkin, 1981) with slight modification due to the introduction of the extra latent variable $\mathbf{Z}$; the two maximization steps consist of the EM algorithm for the mixture models. In order to facilitate understanding, the description and mathematical notations of EMM in the sequel will follow Baker and Kim (2004). It is worthwhile noting that there is a tremendous similarity in the derivation of EMM and MMLE/EM with the only exception of the joint distribution of the complete data.

3.1. Expectation Step and Artificial Data
Let ψ_i denote any item parameter for item i in ξ_i, and very similar to the mathematical derivation for MMEL/EM in Baker and Kim (2004), the first derivative of the log-likelihood function in EMM for each item parameter can be obtained as (see Appendix A in Supplementary Material for details):

$$\frac{\partial \ln L}{\partial \psi_i} = \sum_{j=1}^{N} \frac{\partial \ln P(\mathbf{u}_j, \mathbf{z}_j | \xi_i)}{\partial \psi_i}$$
$$= \sum_{j=1}^{N} \frac{1}{P(\mathbf{u}_j, \mathbf{z}_j | \xi_i)} \frac{\partial P(\mathbf{u}_j, \mathbf{z}_j | \xi_i)}{\partial \psi_i}$$
$$= \sum_{j=1}^{N} \frac{1}{P(\mathbf{u}_j, \mathbf{z}_j | \xi_i)} \int_{\theta_j} \frac{\partial P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi_i) g(\theta_j | \tau) d\theta_j}{\partial \psi_i}$$

$$= \sum_{j=1}^{N} \frac{1}{P(\mathbf{u}_j, \mathbf{z}_j | \xi_i)} \int_{\theta_j} \left[\frac{\partial \ln P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi_i)}{\partial \psi_i} \right]$$
$$P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi_i) g(\theta_j | \tau) d\theta_j$$

$$= \sum_{j=1}^{N} \int_{\theta_j} \left[\frac{\partial \ln P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi_i)}{\partial \psi_i} \right] \left[\frac{P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi_i) g(\theta_j | \tau)}{P(\mathbf{u}_j, \mathbf{z}_j | \xi_i)} \right] d\theta_j$$

$$= \sum_{j=1}^{N} \int_{\theta_j} \left[\frac{\partial \ln P(\mathbf{u}_j, \mathbf{z}_j | \theta_j, \xi_i)}{\partial \psi_i} \right] P\left(\theta_j | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi_i\right) d\theta_j$$

$$= \sum_{j=1}^{N} \int_{\theta_j} \left[\frac{\partial}{\partial \psi_i} \ln \prod_{i=1}^{n} [(1-c_i) P_i^*(\theta_j)]^{u_{ij} z_{ij}} \times c_i^{u_{ij}(1-z_{ij})} \right.$$
$$\left. \times [(1-c_i)(1-P_i^*(\theta_j))]^{(1-u_{ij}) z_{ij}} \right] P\left(\theta_j | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right) d\theta_j$$

$$= \sum_{j=1}^{N} \int_{\theta_j} \left[\frac{[u_{ij} - P_i^*(\theta_j)] z_{ij}}{P_i^*(\theta_j)[1-P_i^*(\theta_j)]} \frac{\partial P_i^*(\theta_j)}{\partial \psi_i} \right.$$
$$\left. + \frac{u_{ij}(1-z_{ij})}{c_i} \frac{\partial c_i}{\partial \psi_i} - \frac{z_{ij}}{1-c_i} \frac{\partial c_i}{\partial \psi_i} \right] P\left(\theta_j | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right) d\theta_j \tag{11}$$

with

$$P\left(\theta_j | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right) = P\left(\theta_j | \mathbf{u}_j, \tau, \xi\right) = \frac{P(\mathbf{u}_j | \theta_j, \xi) g(\theta_j | \tau)}{\int_{\theta_j} P(\mathbf{u}_j | \theta_j, \xi) g(\theta_j | \tau) d\theta_j},$$
$$P(\mathbf{u}_j | \theta_j, \xi) = \prod_{i=1}^{n} P_i(\theta_j)^{u_{ij}} \times (1-P_i(\theta_j))^{1-u_{ij}}, \tag{12}$$

where $P\left(\theta_j | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right)$ is the posterior probability of θ_j given $\mathbf{u}_j, \mathbf{z}_j, \tau$, and ξ, and $P\left(\theta_j | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right)$ always equals $P\left(\theta_j | \mathbf{u}_j, \tau, \xi\right)$ because of θ_j and $\mathbf{z}_j$ are assumed to be independent. The conditional expectation of Z is also essential for the E step. From the joint distribution in Equation (6), one can calculate the expectation of z_{ij} conditional on u_{ij} and the marginal distribution of z_{ij}. By using the Bayesian rule, one has the conditional distribution of z_{ij} on u_{ij} as:

$$P(z_{ij} = 1 | u_{ij} = 1, \theta_j, \xi_i) = \frac{(1-c_i) P_i^*(\theta_j)}{P_i(\theta_j)},$$
$$P(z_{ij} = 1 | u_{ij} = 0, \theta_j, \xi_i) = 1. \tag{13}$$

Then the conditional expectation of z_{ij} is

$$E(z_{ij} | u_{ij}, \theta_j, \xi_i) = \frac{(1-c_i) P_i^*(\theta_j)}{P_i(\theta_j)} \times u_{ij} + 1 \times (1-u_{ij}). \tag{14}$$

The expectation step is to obtain the conditional expectation of the complete-data log likelihood with respect to the observation data, namely the response matrix $\mathbf{U}$. In EMM it essentially amounts to plugging the conditional expectation of the indicator variable $\mathbf{Z}$ and integrating over the latent ability variable as in the original MMLE/EM. Let $X_k(k = 1, 2, \ldots, q)$ be nodes on the

ability scale with an associated weight $A(X_k)$, and the E step is as the follow:

$$\frac{\partial \ln E[L]}{\partial \psi_i}$$
$$\approx \sum_{j=1}^{N} \sum_{k=1}^{q} \left[\frac{[u_{ij} - P_i^*(\theta_j)] z_{ij}}{P_i^*(\theta_j)[1-P_i^*(\theta_j)]} \frac{\partial P_i^*(\theta_j)}{\partial \psi_i} + \frac{u_{ij}(1-E(z_{ij} | u_{ij}, X_k, \xi_i))}{c_i} \frac{\partial c_i}{\partial \psi_i} \right.$$
$$\left. - \frac{E(z_{ij} | u_{ij}, X_k, \xi_i)}{1-c_i} \frac{\partial c_i}{\partial \psi_i} \right] P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right) \tag{15}$$

with

$$P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right) = P\left(X_k | \mathbf{u}_j, \tau, \xi\right)$$
$$= \frac{P(\mathbf{u}_j | X_k, \xi) A(X_k)}{\sum_{k=1}^{q} P(\mathbf{u}_j | X_k, \xi) A(X_k)}, \tag{16}$$
$$P(\mathbf{u}_j | X_k, \xi) = \prod_{i=1}^{n} P_i(X_k)^{u_{ij}} \times (1-P_i(X_k))^{1-u_{ij}},$$

where $P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right)$ is the posterior probability of θ_j at X_k given $\mathbf{u}_j, \mathbf{z}_j, \tau$, and ξ, and $P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right)$ always equals $P\left(X_k | \mathbf{u}_j, \tau, \xi\right)$ because of the independence between $X_k |$ and $\mathbf{z}_j$. Furthermore, $P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right)$ can be used to compute the "artificial data". For instance, Bock and Aitkin (1981) has provided two fundamental artificial data for the traditional EM algorithm as:

$$\bar{f}_{ik} = \sum_{j=1}^{N} P\left(X_k | \mathbf{u}_j, \tau, \xi\right) = \sum_{j=1}^{N} P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right),$$
$$\bar{r}_{ik} = \sum_{j=1}^{N} u_{ij} \times P\left(X_k | \mathbf{u}_j, \tau, \xi\right) = \sum_{j=1}^{N} u_{ij} \times P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right), \tag{17}$$

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

As can be seen from **Table 1**, the EMM algorithm introduced a new latent variable Z, so there are two new artificial data:

$$\bar{f}_{ik}^{(\mathbf{Z})} = \sum_{j=1}^{N} E(z_{ij} | u_{ij}, X_k, \xi) P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right),$$
$$\bar{r}_{ik}^{(\mathbf{Z})} = \sum_{j=1}^{N} u_{ij} \times E(z_{ij} | u_{ij}, X_k, \xi) P\left(X_k | \mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right), \tag{18}$$

in which $\bar{f}_{ik}^{(\mathbf{Z})}$ is the expected number of examinees with ability X_k without using guessing strategy; $\bar{r}_{ik}^{(\mathbf{Z})}$ is the expected number of examinees with ability X_k answering item i correctly without using the guessing strategy. Thus, $\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})}$ is the expected number of examinees with ability X_k who can answer item i correctly using the guessing strategy. The expected number of examinees with ability X_k who can answer item i incorrectly by using the guessing strategy is zero which can be inferred from $P(u_{ij} = 0, z_{ij} = 0 | \theta_j, \xi_i) = 0$. Putting these facts together, one

TABLE 1 | The expected frequencies among examinees with ability X_k for item i.

Item i	$z_j = 1$	$z_j = 0$	Marginal of z_j
$\mathbf{u}_j = 1$	$\bar{r}_{ik}^{(\mathbf{z})}$	$\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{z})}$	$\bar{r}_{ik}$
$\mathbf{u}_j = 0$	$\bar{f}_{ik}^{(\mathbf{z})} - \bar{r}_{ik}^{(\mathbf{z})}$	0	$\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{z})}$
Marginal of $\mathbf{u}_j$	$\bar{f}_{ik}^{(\mathbf{z})}$	$\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{z})}$	$\bar{f}_{ik}$

can infer that $\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})} + \bar{f}_{ik}^{(\mathbf{Z})}$ across all is equivalent to the total number of examinees N, namely,

$$\sum_{k=1}^{q} \bar{f}_{ik} = \sum_{k=1}^{q} (\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})} + \bar{f}_{ik}^{(\mathbf{Z})}) = N. \quad (19)$$

After the E-step and calculation of the artificial data, the next steps are computing the first and second derivatives of Equation (15) with respect to each item parameter.

3.2. Maximization Step-1 for the *c* Parameter

Setting the first derivative equals to 0 and solving for c_i:

$$0 = \lambda_{c_i} = \frac{\partial \ln E[L]}{\partial c_i}$$

$$\approx \sum_{j=1}^{N} \sum_{k=1}^{q} \left[\frac{u_{ij}(1 - E(z_{ij}|u_{ij}, X_k, \xi))}{c_i} - \frac{E(z_{ij}|u_{ij}, X_k, \xi)}{1 - c_i} \right] P\left(X_k|\mathbf{u}_j, \mathbf{z}_j, \tau, \xi\right) \quad (20)$$

$$= \frac{\sum_{k=1}^{q} \left(\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})}\right)}{c_i} - \frac{\sum_{k=1}^{q} \bar{f}_{ik}^{(\mathbf{Z})}}{1 - c_i},$$

and surprisingly, a closed solution for c_i can be available as:

$$\Rightarrow \frac{\sum_{k=1}^{q} \left(\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})}\right)}{c_i} = \frac{\sum_{k=1}^{q} \bar{f}_{ik}^{(\mathbf{Z})}}{1 - c_i}$$

$$\Rightarrow c_i = \frac{\sum_{k=1}^{q} \left(\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})}\right)}{\sum_{k=1}^{q} \left(\bar{f}_{ik}^{(\mathbf{Z})} + \bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})}\right)} = \frac{\sum_{k=1}^{q} \left(\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})}\right)}{N} \quad (21)$$

$$\text{or} \quad c_i = 1 - \frac{\sum_{k=1}^{q} \bar{f}_{ik}^{(\mathbf{Z})}}{N}.$$

There is a nice interpretation for the estimate of the guessing parameter. From the description of the artificial data, it is easy to note that $\sum_{k=1}^{q} (\bar{r}_{ik} - \bar{r}_{ik}^{(\mathbf{Z})})$ is the expected number of examinees who can answer item i correctly using the guessing strategy and thus, the estimate is exactly the proportion of these examinees in the total sample. This interpretation presents a strong mixture modeling flavor, drastically different from the traditional interpretation, the lower bound for the probability with which a person solves the item correctly.

3.3. Maximization Step-2 for *a* and *b* Parameters

The second Maximization step is to execute the Newton-Raphson procedure to obtain estimates for a_i and b_i. The required first derivatives for a_i and b_i are

$$\lambda_{a_i} = \frac{\partial \ln E[L]}{\partial a_i} = D \sum_{k=1}^{q} [(\bar{r}_{ik}^{(\mathbf{Z})} - \bar{f}_{ik}^{(\mathbf{Z})} \times P_i^*(X_k))(X_k - b_i)],$$

$$\lambda_{b_i} = \frac{\partial \ln E[L]}{\partial b_i} = -D a_i \sum_{k=1}^{q} [\bar{r}_{ik}^{(\mathbf{Z})} - \bar{f}_{ik}^{(\mathbf{Z})} \times P_i^*(X_k)]. \quad (22)$$

One may also set the derivatives to zero, but there is no analytical solution to them and thus Newton-Raphson method have to be used. In order to implement Newton-Raphson, the second derivatives are derived as

$$\lambda_{aa_i} = \frac{\partial^2 \ln E[L]}{\partial a_i \partial a_i} = -D^2 \sum_{k=1}^{q} (X_k - b_i)^2 W_{ik}^* \times \bar{f}_{ik}^{(Z)},$$

$$\lambda_{bb_i} = \frac{\partial^2 \ln E[L]}{\partial b_i \partial b_i} = -D^2 a_i^2 \sum_{k=1}^{q} W_{ik}^* \times \bar{f}_{ik}^{(Z)},$$

$$\lambda_{ab_i} = \frac{\partial^2 \ln E[L]}{\partial a_i \partial b_i} = \sum_{k=1}^{q} \left[\begin{array}{c} D^2 a_i(X_k - b_i) W_{ik}^* \times \bar{f}_{ik}^{(Z)} \\ -D\left(\bar{r}_{ik}^{(Z)} - \bar{f}_{ik}^{(Z)} \times P_i^*(\theta_j)\right) \end{array} \right], \quad (23)$$

where

$$W_{ik}^* = P_i^*(X_k) \times (1 - P_i^*(X_k)), \quad (24)$$

So, the estimates can be obtained using the Newton-Raphson iteratively:

$$\begin{bmatrix} a_i^{(t+1)} \\ b_i^{(t+1)} \end{bmatrix} = \begin{bmatrix} a_i^{(t)} \\ b_i^{(t)} \end{bmatrix} - \begin{bmatrix} \lambda_{aa_i} & \lambda_{ab_i} \\ \lambda_{ab_i} & \lambda_{bb_i} \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{a_i} \\ \lambda_{b_i} \end{bmatrix}. \quad (25)$$

3.4. Standard Errors (SEs) of Parameter Estimation in EMM

One of the important indices to assess the estimation quality is the standard errors. One of the major criticisms of EM is, however, that parameter estimate SEs are not a natural product of the algorithm and some other methods have to be devised (McLachlan and Krishnan, 2007). EMM, as a member of EM algorithms, does not provide parameter estimate SEs either. In general, the inverse of the negative expected value of the matrix of second derivatives of a log likelihood is the asymptotic variance-covariance matrix of the estimates (McLachlan and Krishnan, 1968). The square roots of the diagonal elements of the inverse are the asymptotic standard errors of the parameters. This part will present a theoretical argument on why EMM works better than MMLE/EM for 3PLM. In the original MMLE/EM, the expected second derivative matrix can generically be written as

$$\begin{bmatrix} l_{aa} & l_{ab} & l_{ac} \\ l_{ab} & l_{bb} & l_{bc} \\ l_{ac} & l_{bc} & l_{cc} \end{bmatrix},$$

in which l_{cc} is the expected second derivatives of the log likelihood with respect to the guessing parameter, l_{ac} and l_{bc} are

the elements for two expected second partial derivatives. The convergence issue of the MMLE/EM for 3PLM is caused by this matrix being non-positive-definite or ill-conditioned (Baker and Kim, 2004). To solve this problem, several researchers suggested that the guessing parameter can usually set to be reciprocal of the number of the alternatives in the item (Hutchinson, 1991; Han, 2012; McDonald, 2013), which essentially mounted to set l_{ac} and l_{bc} to be zeros and then the 3PLM could be estimated. This option is available in several IRT programs, such as NOHARM (Fraser and Mcdonald, 1988), but an obvious limitation of this practice is that the guessing parameter is, in fact, not estimated, but a rather rough "guess" and does not seem to be a reasonable assumption (San Martín et al., 2006). In contrast, this matrix is different in EMM because EMM amounts to a divide-and-conquer strategy. The challenge of estimating item parameters is intelligently partitioned into two smaller estimation problems: estimation of the guessing parameter as the mixing proportion parameter in mixing modeling and that of the 2PL model. Interestingly, the two smaller problems happen to have been fully investigated in statistics and IRT. Conceptually, EMM can be perceived as a combination of these two mature techniques in statics and IRT. The prominent advantage of EMM comes from the separation of estimation of the guessing parameter from the other two and, thus, any instability in estimating the guessing parameter will not negatively affect that of the other two parameters, or otherwise. Such a setup statistically implies that the covariance between the guessing estimate and the other two are zeros. One may ensemble the expected second derivative matrix in generic form as

$$
\begin{bmatrix}
E\left(-\lambda_{aa}|u_{ij}\right) & E\left(-\lambda_{ab}|u_{ij}\right) & 0 \\
E\left(-\lambda_{ab}|u_{ij}\right) & E\left(-\lambda_{bb}|u_{ij}\right) & 0 \\
0 & 0 & E\left(-\lambda_{cc}|u_{ij}\right)
\end{bmatrix}.
$$

EMM provides a scientific alternative to the practice of fixing the guessing parameter estimate as the reciprocal of the number of the alternatives in the item. It can not only eliminate undesirable fluctuation in estimating the guessing parameter, but also make estimation of the guessing parameter possible. For the sake of completeness, the second derivatives of the log likelihood with respect to the guessing parameter is given as

$$
\lambda_{cc_i} = \frac{\partial^2 \ln E[L]}{\partial^2 c_i^2} = -\sum_{k=1}^{q} \left(\frac{\bar{r}_{ik} - \bar{r}_{ik}^{(Z)}}{c_i^2} + \frac{\bar{f}_{ik}^{(Z)}}{(1-c_i)^2} \right) \quad (26)
$$

As a consequence of the separation setup, the calculation does not involve any terms related to the difficulty and discrimination parameters. The detailed derivation of the SEs for EMM is provided in Appendix B (Supplementary Material).
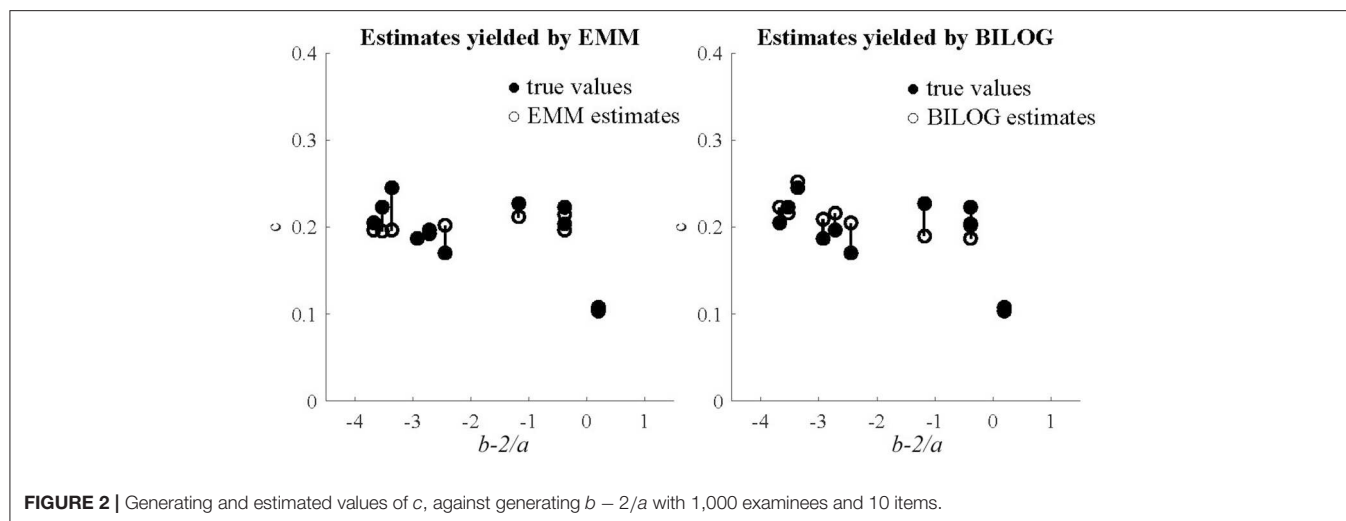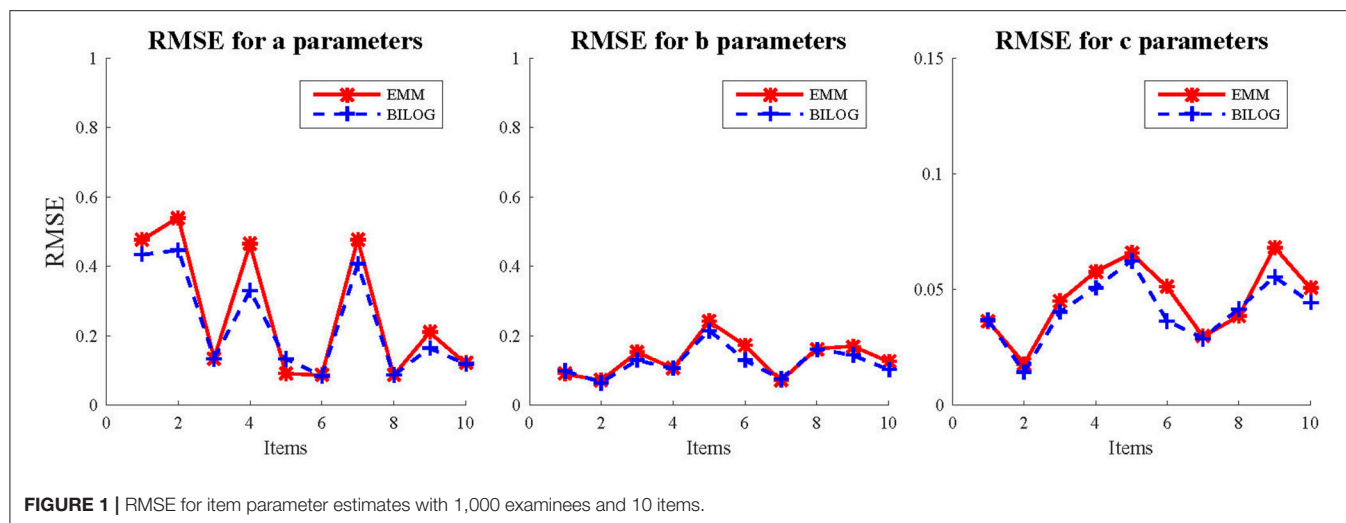
4. SIMULATION STUDY

A simulation study was done to demonstrate the feasibility of the new method, compared to the Bayesian EM in BILOG-MG (Zimowski et al., 2003). The parameters for 10 or 20 items were generated from independent normal distribution as in Mislevy (1986): for $\ln a$, the mean and variance were 0 and 0.5; for b, 0.5 and 1.0; for logit c, -1.39 and 0.16.

Three sample sizes of examinees (1,000, 1,500, and 2,000) were generated from standard normal distribution. 50 replications were run for each condition in the fully crossed 2(EMM vs. BILOG-MG) $\times$ 3(1,000 vs. 1,500 vs. 2,000) $\times$ 2(10 vs. 20) design. The evaluation criteria are the bias and RMSE. In order to further demonstrate the performance of EMM, the EMM and Bayes solutions respectively, against generating values of the quantity $b - 2/a$, a heuristic index based on the observation that less information is obtained about c as items become easier or less reliable (Lord, 1975). Results. The detailed results for the simulation study are presented in tables and figures in Appendix C (Supplementary Material). Only the results for the condition of 1,000 examinees and 10 items are summarized and presented here (**Figures 1**, **2**) since others are very similar. This is the most unfavorable condition for MLE since the sample size is moderately small and the test length is relatively short. But even under this condition, the EMM is comparable to or better than Bayesian EM in terms of bias and RMSE. In contrast, the traditional MMLE EM usually fails to converge with such a small sample size. As a MLE algorithm EMM is as flexible as the Bayesian EM implemented in BILOG-MG to deal with 3PL modeling in most practical testing situations.

The plots for the heuristic index comparing the EMM and Bayes solutions respectively, against generating values of the quantity $b - 2/a$, echo the results in terms of bias and RMSE. The figures for the 1,000-examinee-10-item condition are presented here. In general, both solutions for items with high values in the index are satisfactory, but the Bayes estimates for some items with low values were rather biased while the EMM estimates were very stable.

5. TWO EMPIRICAL EXAMPLES

In order to further demonstrate the performance of the new algorithm, the authors apply EMM to two real data sets from BILOG-MG (Zimowski et al., 2003) and flexMIRT (Houts and Cai, 2015), and compare the item estimates to those from the two commercial programs. Specifically, the two data sets are the Example 1 in the BILOG-MG and Example 2 in flexMIRT. The BILOG-MG data set consists of 1,000 examinees' responses to 15 items and the one for flexMIRT consists of 2,844 examinees' responses to 12 items. Please refer to the manuals for the details. Both programs adopt different priors for the guessing parameter: BILOG-MG uses the default setting, a beta distribution with parameters of 4 and 16, namely, $c \sim Beta(4, 16)$, and flexMIRT, $c \sim Beta(1, 4)$. Please note that due to different parameterization in BILOG-MG technical document, the specification for the prior parameters are 5 and 17 which correspond to 4 and 16 in the standard beta distribution parameterization. The two priors have identical means, but the variance for flexMIRT prior is smaller which indicates that it is less informative. An additional analysis of flexMIRT data using the BILOG-MG default setting for the guessing parameter in flexMIRT [the $c \sim Beta(1, 4)$ solution] is also run. In order to further facilitate the comparison of SEs, the authors employ supplemented EM (SEM) to obtain SEs for EMM (Meng and Rubin, 1991) in the real-data analysis below and SEM

FIGURE 1 | RMSE for item parameter estimates with 1,000 examinees and 10 items.

FIGURE 2 | Generating and estimated values of c, against generating $b - 2/a$ with 1,000 examinees and 10 items.

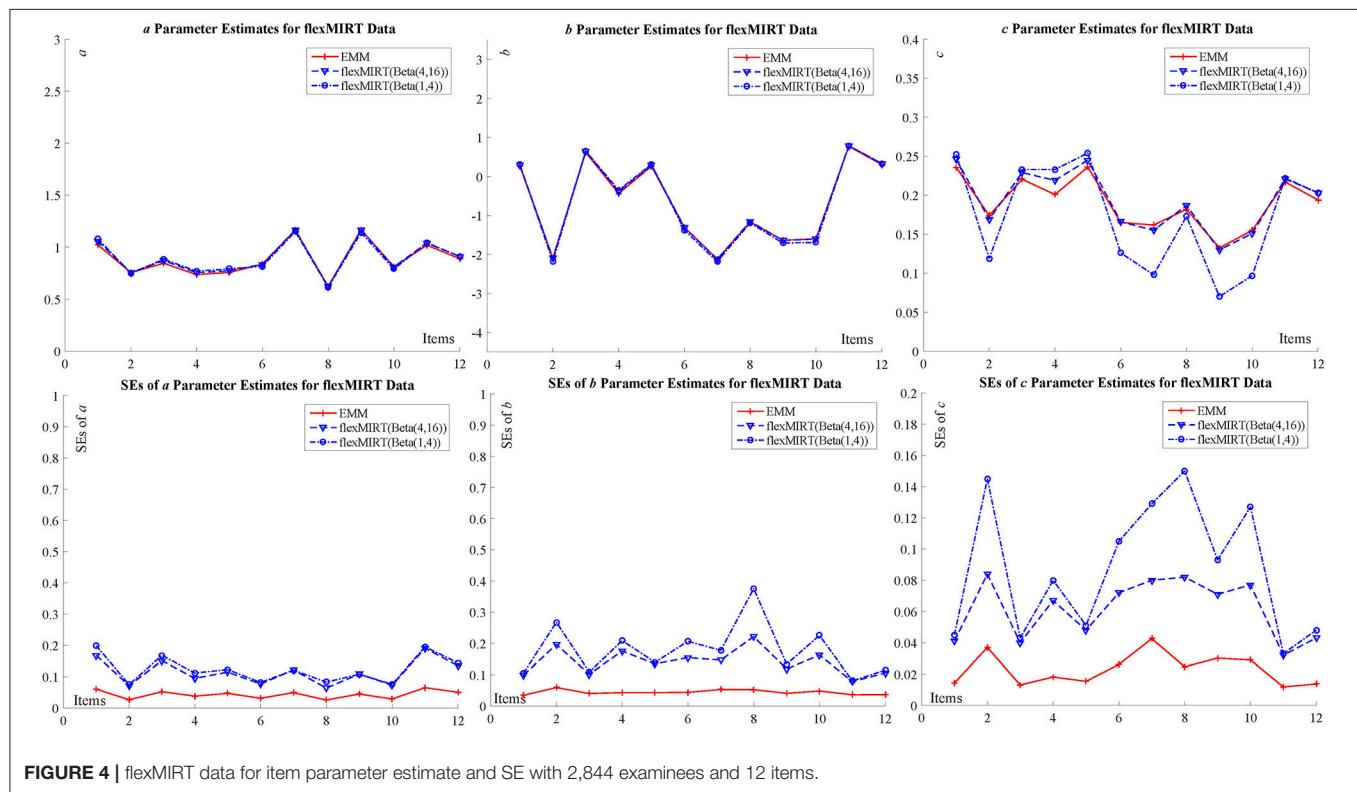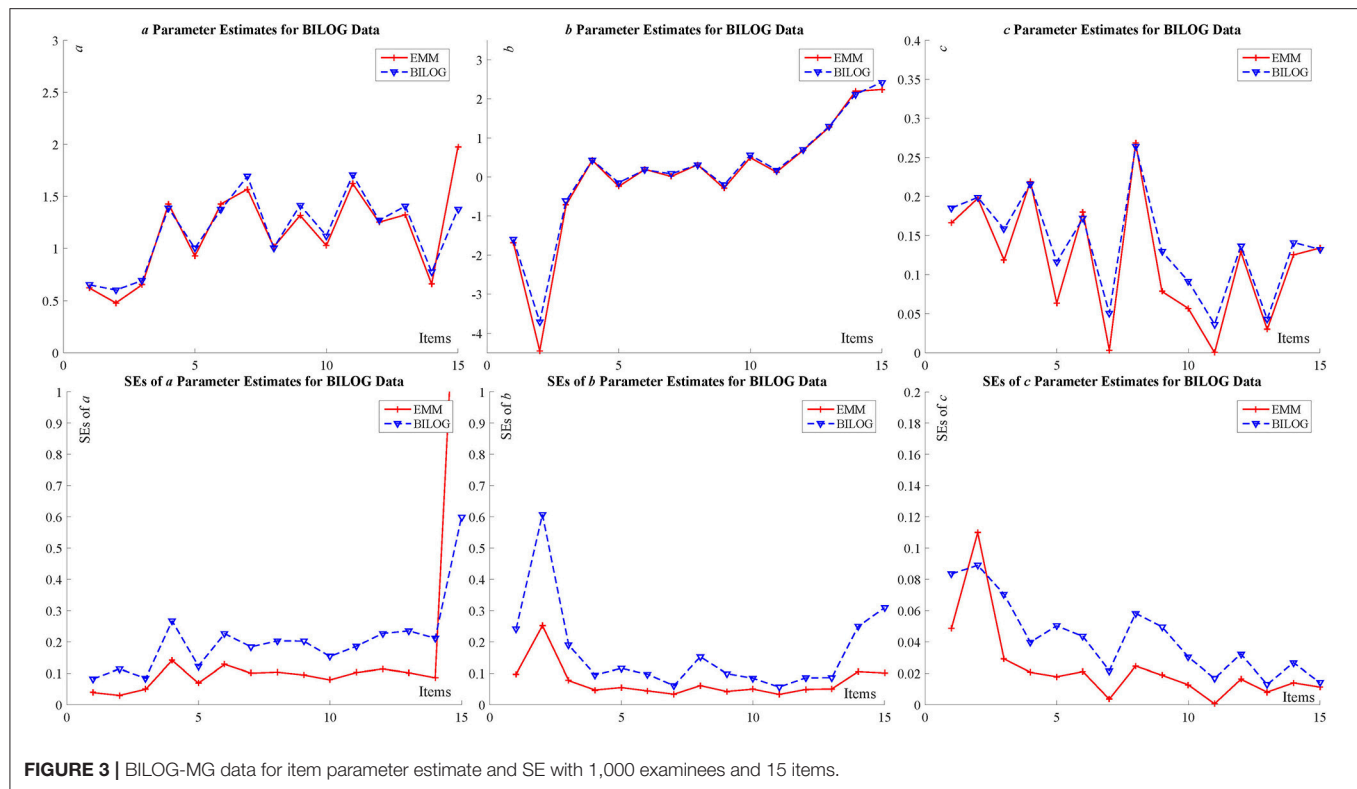has been applied for various IRT models (Cai, 2008; Cai and Lee, 2009; Tian et al., 2013).

The specific results are summarized in the tables and figures in Appendix D (Supplementary Material) and only figures (**Figure 3** for BILOG-MG data and **Figure 4** for flexMIRT data) are presented here. Both **Figures 3**, **4** indicates that, for both real data sets, the point estimates (the upper part of the figures) for item parameters from EMM are comparable to those from BILOG-MG and flexMIRT, but SEs (the lower part of the figures) are smaller. Smaller SEs means that one may ascertain point estimates with greater confidence.

The advantage of EMM can be further confirmed by the difference between the $c \sim Beta(1, 4)$ solution and the $c \sim Beta(4, 16)$ solution to the point estimate for the guessing parameters for flexMIRT data. The EMM point estimates of the guessing parameters coincide with the $c \sim Beta(4, 16)$ solution, with smaller SEs while those from the $c \sim Beta(4, 16)$ solution present nontrivial differences. The authors may reasonably conclude that the difference between the solutions from flexMIRT calibration since the only difference in

the specification is the prior setting for the guessing parameter. This difference is not a surprise at all because assigning proper priors is a common burden shared by the Bayesian approach. In the case of 3PLM calibration, however, different priors exert an undesirable negative effect of the point estimation and SEs. To exclude the possible difference in software implementation in the two programs, the authors also replicated the same analysis in BILOG-MG and the results corroborate the claim. Furthermore, as expected, the SEs in the $c \sim Beta(1, 4)$ solution are bigger than those in the $c \sim Beta(4, 16)$ solution which is another piece of evidence that priors in Bayesian methods lead to different SE estimates.

6. DISCUSSION AND FUTURE DIRECTIONS

In summary, the EMM essentially is a member of EM family for 3PLM. The fundamental difference between the EMM and the original EM is that the old complete data setup $(\mathbf{U}, \theta)$ is expanded

FIGURE 3 | BILOG-MG data for item parameter estimate and SE with 1,000 examinees and 15 items.

FIGURE 4 | flexMIRT data for item parameter estimate and SE with 2,844 examinees and 12 items.

into $(\mathbf{U}, \theta, \mathbf{Z})$. This change enables the algorithm to explore the likelihood function curve more thoroughly. More specifically, the challenge of estimating item parameters is intelligently partitioned into two smaller estimation problems: estimation of the guessing parameter as the mixing proportion parameter in mixing modeling and that of the 2PL model, which happen to be fully investigated in statistics and IRT. The benefit of this strategy is that the estimation error in the two problems does not exert any negative effect on each other and thus both of them are more stable. Moreover, for the guessing parameter, there is an analytical solution for the score function and thus the NR and the second derivatives are not necessary.

The EMM is a feasible algorithm to obtain MLE for 3PLM with modest sample size. It has several theoretical and practical implications. First, the mixture modeling approach to 3PLM is an interesting novel perspective on 3PLM. Second, this new method provides a feasible alternative for practitioners, enabling them to obtain MLE for 3PLM with modest sample size. This paper is not to advocate to eliminate use of Bayesian estimation methods. This can be used to check with the Bayesian solution with other IRT programs in practice. Due to the high complexity in real-world 3PLM data, a combination of EMM and Bayesian methods might lead to a more sophisticated and nuanced understanding of data.

Another important feature of EMM is that the EMM stopping criterion serves as a self-checking/correcting mechanism for the independence assumption. The stopping criterion is calculated through the original 3PLM likelihood function, although the E and M steps are built with the approximation of the likelihood function of the reformulated model. That is to say, the EMM would not converge until the difference between the values of the original 3PLM likelihood function evaluated at the current and last-cycle item estimates becomes negligible. If the independence assumption is severely violated and thus the approximation is not accurate enough, the convergence of the algorithm will suffer. In this sense, the EMM algorithm has provided a data-driven validation method for this assumption.

Several research questions deserve further attention. Firstly, there are two different interpretation of 3PLM (Hutchinson, 1991; San Martín et al., 2006; von Davier, 2009) and this paper chose one of them. The derivation of an EMM for the other interpretation will be very similar to this paper, but it needs to provoke different assumptions, so it would

be interesting to compare the performance of two versions in terms of estimate and interpretation for the guessing parameter, numerical stability, etc. Secondly, a Bayesian EMM can be investigated. EMM is a more powerful MMLE method comparable to Bayesian EM, but the simulation studies indicate that there might be some inflation in item estimates in EMM which points to the possibility of improvement. A natural alternative is to combining the Bayesian approach and EMM. Bayesian EMM will solve the issue of estimate inflation while taking advantage of the EMM exploring the likelihood function. Thirdly, the mixture modeling approach and EMM can be naturally extended to 4PLM (Barton and Lord, 1981), which is a generalization of 3PLM and includes an upper asymptote for the probability of a correct response. There is a renewed interest in 4PLM (Rulison and Loken, 2009; Loken and Rulison, 2010; Liao et al., 2012; Ogasawara, 2012; Feuerstahler and Waller, 2014; Culpepper, 2016) for its usefulness in measuring psychological constructs. But, just as in the case of 3PLM, one consistent discussion point regarding 4PL pertains to the difficulty in estimating item parameters. The method proposed in this paper is a promising way for 4PLM.

AUTHOR CONTRIBUTIONS

CZ: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing - original draft, Writing - review, and editing. XM: Funding acquisition and Software. SG: Software, Writing - review, and editing. ZL: Data curation, Project Administration, and Validation.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02302/full#supplementary-material

REFERENCES

Baker, F. B., and Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques*. Boca Raton, FL: CRC Press.

Barton, M. A., and Lord, F. M. (1981). An upper asy, ptote for the three-parameter logistic item-response model. *ETS Res. Report Ser.* 1, 1–8.

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 forn dichotomously scored items. *Psychometrika* 35, 179–197. doi: 10.1007/BF02291262

Cai, L. (2008). Sem of another flavour: two new applications of the supplemented em algorithm. *Br. J. Math. Stat. Psychol.* 62, 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. *Multiv. Behav. Res.* 44, 281–304. doi: 10.1080/00273170902794255

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

Casella, G., and Berger, R. (2002). *Statistical Inference*. Pacific Grove, CA: Duxbury.

Cornell University Library (1989). Statistical models and analysis in auditing: panel on nonstandard mixtures of distributions. *Stat. Sci.* 4, 2–33.

Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika* 81, 1142–1163. doi: 10.1007/s11336-015-9477-6

Feuerstahler, L., and Waller, N. G. (2014). Abstract: estimation of the 4-parameter model with marginal maximum likelihood. *Multiv. Behav. Res.* 49, 285–285. doi: 10.1080/00273171.2014.912889

Fraser, C., and Mcdonald, R. P. (1988). Noharm: least squares item factor analysis. *Multiv. Behav. Res.* 23, 267–269. doi: 10.1207/s15327906mbr2302_9

Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory (Measurement Methods for the Social Sciences Series, Vol. 2)*. Newbury Park, CA: SAGE Publications.

Han, K. T. (2012). Fixing the c parameter in the three-parameter logistic model. *Pract. Assessm. Res. Eval.* 17:24. Available online at: https://eric.ed.gov/?id= EJ977575

Holland, P. W. (1990). The dutch identity: a new tool for the study of item response models. *Psychometrika* 55, 5–18. doi: 10.1007/BF02294739

Houts, C. R., and Cai, L. (2015). *flexMIRT R User's Manual Version 3.0: Flexible Multilevel Multidimensional Item Analysis and Test Scoring*. Chapel Hill, NC: Vector Psychometric Group, LLC.

Hutchinson, T. P. (1991). *Ability, Partial Information, Guessing: Statistical Modelling Applied to Multiple-choice Tests*. Rundle Mall, SA: Rumsby Scientific Publishing.

Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *J. Educ. Meas.* 18, 1–11. doi: 10.1111/j.1745-3984. 1981.tb00838.x

Liao, W., Ho, R., Yen, Y., and Cheng, H. (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. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika* 39, 247–264. doi: 10.1007/BF02291471

Lord, F. M. (1975). Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. *ETS Res. Bull. Ser.* 1975, 1–52.

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: 10.1080/15366 360903070385

McDonald, R. P. (2013). *Test Theory: A Unified Treatment*. London: Psychology Press.

McLachlan, G., and Krishnan, T. (1968). *The Advanced Theory of Statistics*. New York, NY: Macmillan.

McLachlan, G., and Krishnan, T. (2007). *The EM Algorithm and Extensions, 2nd Edn*. New York, NY: John Wiley and Sons.

McLachlan, G., and Peel, D. (2004). *Finite Mixture Models*. John Wiley & 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

Patz, R. J., and Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *J. Educ. Behav. Stat.* 24, 342–366. doi: 10.3102/10769986024004342

Rost, J. (1997). "Logistic mixture models," in *Handbook of Modern Item Response Theory*, eds W. J. van der Linden and R. K. Hambleton (Berlin: Springer-Verlag), 449–463.

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

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

San Martín, E., González, J., and Tuerlinckx, F. (2015). On the unidentifiability of the fixed-effects 3pl model. *Psychometrika* 80, 450–467. doi: 10.1007/s11336-014-9404-2

San Martín, E., Rolin, J.-M., and Castro, L. M. (2013). Identification of the 1PL model with guessing parameter: parametric and semi-parametric results. *Psychometrika* 78, 341–379. doi: 10.1007/s11336-013-9322-8

Swaminathan, H., and Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika* 51, 589–601. doi: 10.1007/BF022 95598

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 tesponse 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. (2008). A general diagnostic model applied to language testing data. *Br. J. Math. Stat. Psychol.* 61, 287–307. doi: 10.1348/000711007X193957

von Davier, M. (2009). Is there need for the 3pl model? guess what? *Meas. Interdiscip. Res. Perspect.* 7, 110–114. doi: 10.1080/15366360903117079

von Davier, M., and Rost, J. (2006). "Mixture distribution item response models," in *Handbook of Statistics*, vol. 26, eds C. R. Rao and S. Sinharay (Amsterdam: Elsevier), 643–661.

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

Yamamoto, K. (1982). Hybrid model of IRT and latent class models. *ETS Res. Rep.* 1982, 1–61. doi: 10.1002/j.2333-8504.1982.tb01326.x

Yen, W. M., Burket, G. R., and Sykes, R. C. (1991). Nonunique solutions to the likelihood equation for the three-parameter logistic model. *Psychometrika* 56, 39–54. doi: 10.1007/BF02294584

Zimowski, M., Muraki, E., Mislevy, R., and Bock, R. (2003). *BILOG-MG3 User's Guide*. Chicago, IL: SSI International.