FROM EXPECTATION-3-MAXIMIZATION TO
BAYESIAN EXPECTATION-3-MAXIMIZATION:
A LATENT MIXTURE MODELING-BASED BAYESIAN ALGORITHM
FOR THE 4-PARAMETER LOGISTIC MODEL

BY

CI ZHANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Master's Committee:

Associate Professor Jinming Zhang, Chair
Professor Hua-Hua Chang
Professor Carolyn J. Anderson

ABSTRACT

There is renewed interest in the four-parameter logistic model (4PLM), but the lack of a user-friendly calibration method constitutes a major barrier to its widespread application. In the present study, this researcher reformulated the 4PLM from a latent mixture modeling view and developed the Expectation-Maximization-Maximization-Maximization (EMMM) method. Combining the EMMM with the Bayesian approach, allowed the Bayesian Expectation-Maximization-Maximization-Maximization (BEMMM) algorithm to be proposed. First, the author compared the EMMM with BEMMM to confirm that the BEMMM method reduced the number of implausible estimates in EMMM. Next, when comparing the BEMMM with the Markov Chain Monte Carlo method (Culpepper, 2016) and Bayesian Modal Estimation (Waller & Feuerstahler, 2017), the results from a simulation study and a real-world data calibration indicated that the BEMMM and the MCMC are more accurate than the BME, while the BEMMM is much faster than the MCMC.

ACKNOWLEDGEMENTS

First, I would like to express my great thanks to my advisor, Dr. Jinming Zhang, for his support and guidance. And I thank all of the members of my committee, Professor Hua-Hua Chang and Professor Carolyn J. Anderson, for the suggestions in the academic studies and research processes.

I thank my friends and teachers at the Department of Educational Psychology at the University of Illinois at Urbana-Champaign and Jiangxi Normal University, particularly Dr. Chanjin Zheng and Mr. Shaoyang Guo for their generous help with my research.

Finally, I would like to extend my gratitude to my parents for their financial support and selfless love.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

The four-parameter logistic item response model (4PLM) was first mentioned by McDonald (1967) and was formally proposed by Barton and Lord (1981). It received little attention due to doubts about its utility and technical difficulties related to parameter calibration. After three decades of neglect, the psychometrics community has developed a rekindled interest in 4PL due to several applications (Cheng & Liu, 2015; Liao, Ho, Yen, & Cheng, 2012; Loken & Rulison, 2010; Rulison & Loken, 2009; Waller & Reise, 2010; Yen, Ho, Liao, Chen, & Kuo, 2012). Currently, however, the calibration method continues to present challenges to methodologists and practitioners which has hindered its widespread application. There are two kinds of major approaches to item calibration for the 4PLM. One type is the MCMC methods, including Metropolis-Hasting (Loken & Rulison, 2010) and the Gibbs sampler (Culpepper, 2016). The major problem with these estimation means is that they are computationally intensive and time-consuming. The other sort of approach is to apply the Bayesian method for the 3PLM to the 4PLM (Waller & Feuerstahler, 2017). Compared to the MCMC methods, this kind of approach is faster, but is not as accurate as the first methods.

In this study, an Expectation-Maximization-Maximization-Maximization (EMMM) algorithm and a Bayesian Expectation-Maximization-Maximization-Maximization (BEMMM) algorithm based on latent-mixture-modeling reformulation are proposed. In section 1, the expression and item characteristic curve of the 4PLM, the brief history of 4PLM and several previous algorithms for the 4PLM are introduced. The author then illustrates the derivatives and algorithms of EMMM and Bayesian EMMM step-by-step. In section 3, two simulation studies are conducted. One is designed for the purpose of comparing the EMMM with the BEMMM, the results of which indicate that the Bayesian method helps reduce implausible estimates in the EMMM. The second simulation study focuses on the item recovery of the BEMMM, BME, and MCMC. The comparison indicates the conclusion such that the BEMMM is as accurate as MCMC and is more precise than BME. Finally, we apply the 4PLM to bullying item responses of 7491 adolescents from the 2005-2006 Health Behavior in School-Aged Children (HBSC) study (see Culpepper, 2016, section 4). The results support the contention that using the BEMMM can estimate lower and upper asymptotes in large-scale surveys as MCMC does and do so much quicker than MCMC.

1.1 The Four-Parameter Logistic Model

The 4PLM extends the three-parameter model by adding an upper asymptote parameter. The probability of correct response for examinee i on item j is:

$$P(u_{ij}=1 \mid \theta_i, a_j, b_j, c_j, d_j) = c_j + \frac{d_j - c_j}{1+\exp(-Da_j(\theta_i - b_j))}. \qquad (1.1)$$

where a_j and b_j are the discrimination and difficulty parameters, c_j and d_j are the lower and upper asymptotes, θ_i is the latent trait score, and D is 1.702. In Figure 1.1, two curves with both $a_j = 1.4$ and $b_j = -0.15$ are illustrated. For the darker curve, $c_j = 0.2$, which means that the probability of lower-ability examinees correctly answering is 0.2; $d_j = 0.8$, which indicates that the probability of higher ability examinees correctly answering is 0.8. For the lighter curve, $c_j = 0.1$, $d_j = 0.9$. When

$\theta_i = b_j = -0.15$, the probability of answering correctly for both functions is

$\dfrac{c_j + d_j}{2} = 0.5$.

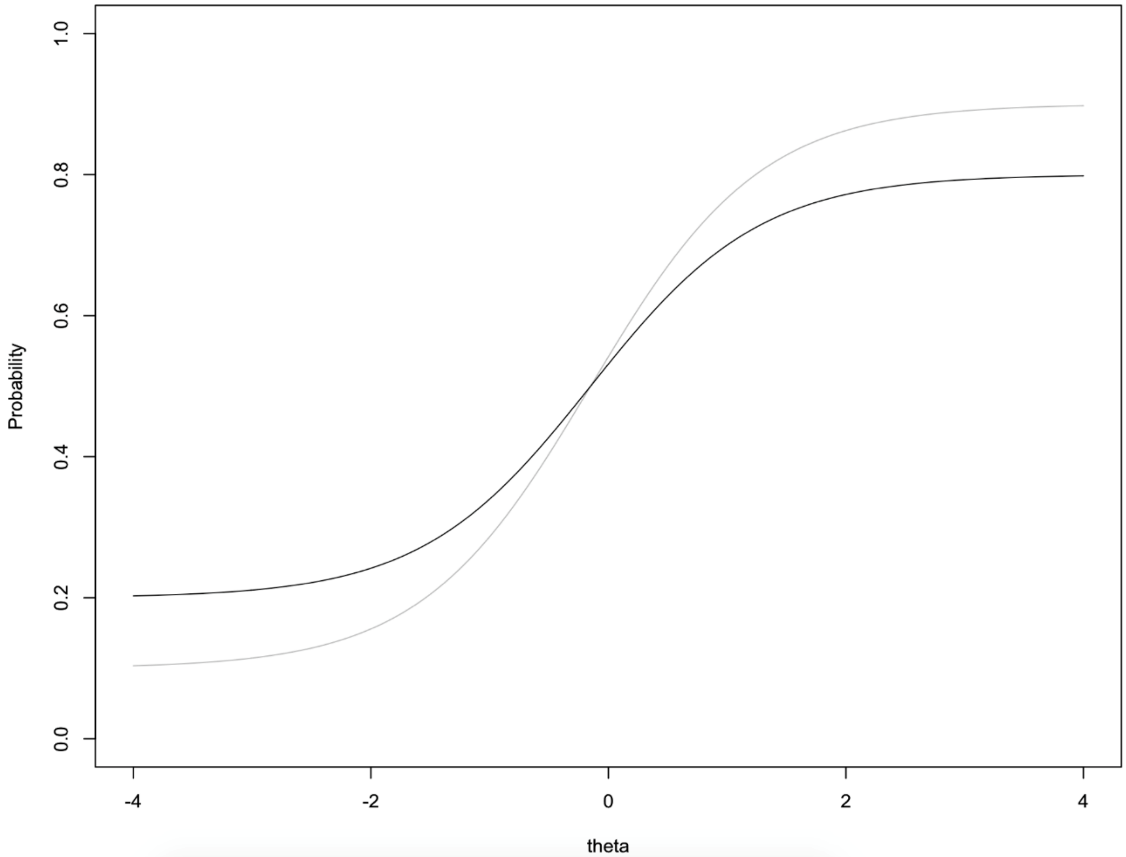

Figure 1.1 Item character curve for 4-parameter logistic model

1.2 The history of the four-parameter logistic model

The concept of the 4PLM was firstly mentioned by McDonald in 1967:

> "It would be convenient for some applications to relax these restrictions on the model, by introducing upper and lower asymptotes which can be determined from the data, and which are free to take values other than zero and unity" (p. 67).

Unlike the lower asymptote, which was popularly used by psychometric researchers, the upper asymptote suffered 15 years of neglect before being formally proposed by Barton and Lord (1981). Motivated by the concern that severe penalizations might occur when high-ability examinees make clerical errors on easy items if using a 3PLM (1981, p. 2), they introduced “an upper asymptote with a value of slightly less than 1” (1981, p. 2) and compared the new model with the 3PL model. Without really estimating the parameters, Barton and Lord considered the effect of the upper asymptote by changing the values of the upper asymptotes (0.98, 0.99 or 1.00) to determine the changes in log-likelihoods and ability estimates. However, the use of the four-parameter model did not “consistently improve the likelihood or significantly change any ability estimates,” so they concluded that the need for 4PLM is neither compelling nor urgent (1981, p. 6). Similar negative opinion on the 4PLM was raised by Hambleton and Swaminathan in 1985.

After about two decades of silence, there was renewed interest in the application of the 4PLM (e.g., Loken & Rulison, 2010; Magis, 2013; Reise &Waller, 2003; Waller & Reise, 2010; Yen, et al., 2012). Reise and Waller (2003) fitted the 2PLM and 3PLM, respectively, to 15 unidimensional factor scales (p.164), and in the last part of the article, they pointed out that a four-parameter model estimation program is needed to characterize the functioning of psychopathology items completely (p.182). They did not fit the 4PLM to the Minnesota Multiphasic Personality Inventory (MMPI) scales at that time because they did not at the time know of there is any software that could estimate the 4PLM (Waller & Reise, 2010, p.151). In 2010, Waller and Reise found that “it is now possible to estimate IRT models via a Gibbs sampler.” Using an open-source R package “BRUGS” (Thomas, 2006), which is based on OpenBUGS (http://mathstat.helsinki.filopenbugs /) architecture (p.157), researchers fit the 4PLM to real data. In that same year, Loken and Rulison also estimated the 4PLM with the

Bayesian method where Markov Chain Monte Carlo (MCMC) approach was used to simulate the posterior full joint distribution and marginal distribution of the parameters (p.513). Comparing the model fit of the 4PLM, 3PLM, and 2PLM for data generated using the 4PLM reflected that the results showed that the 4PLM provided the best fit (p.521). They believed that the probability of correct answers could not reach 1 in practical measurements even when examinees' ability level is extremely high, so it is necessary to use a model with an upper asymptote. Furthermore, they mentioned the possibility of future work involving applying an ML approach with some constraints to the 4PLM (p.523). The need of 4PLM also showed in computerized adaptive test (CAT) as the estimation error was profoundly influenced by some aberrant responses such as careless errors and lucky guesses (Yen, et al., 2012, p.75). Yen, et al., compared the accuracy and efficiency of the 3PLM and 4PLM based CAT with items drawn from the English Ability Test for college entrance in Taiwan (Ho & Yen, 2005), the results showed that the issues of ability underestimation were decreased and the efficiency of measurement was increased when the 4PLM was used (p.85). Liao, et al. (2012) conducted a simulation study to investigate the robustness of the 4PLM compared to the 3PLM in CAT under two conditions (normal and poor-start test) and obtained the same conclusion as Yen, et al. (2012) in the empirical experiment. Magis (2013) rewrote the item information function given by Lord (1980, p. 72) and derived the value of the ability level that maximizes the item information function (p.312).

Via a recently developed *mirt* (Chalmers, 2012) package, Feuerstahler and Waller (2014) estimated the 4PLM using marginal maximum likelihood (MML) method. By fitting the 4PLM to the MMPI-A factor scale (Butcher, Dahlstrom, Graham, Tellegen, &

Kaemmer, 1992) and comparing the 4PLM with 3PLM and 2PLM, researchers found that although reasonably accurate estimates were obtained only when the sample size was large (N=10,000), the 4PLM significantly improved model fit. Following Béguin and Glas (2001), Culpepper (2016) remodeled the 4-parameter normal ogive model (4PNO) by introducing a discrete augmented variable. Through Monte Carlo simulations and a real data sample, the results showed that the 4PNO model provided the best model fit when comparing to the 2PNO and the 3PNO models, and the sample size which is needed to obtain accurate estimates is 2,500. Waller and Feuerstahler (2017) fitted the 4PLM, 3PLMu (a submodel of the 4PLM where the lower asymptote is equal to 0), 3PLM and 2PLM to MMPI-A factor scales (Butcher, et al., 1992) using Bayesian modal estimation (BME). The results of the comparison indicated that models with non-constrained upper asymptotes (4PM and 3PMu) are more suitable for some psychopathology scales (p.18). They also explored the minimum sample size needed for accurate item parameters estimation is larger than 5,000.

1.3 Algorithms for 4PLM

The MMLE/EM algorithm. The marginal maximum likelihood estimation with the Expectation-Maximization algorithm (MMLE/EM) for the 3PLM, proposed by Bock and Aitkin (1981), can be modified to estimate the 4PLM because the 4PLM is a generalization of the 3PLM. Zhang (2005, 2012) uses a modified MMLE/EM to estimate item parameters for multi-dimensional compensatory three-parameter logistic models. The underlying assumptions of the MMLE are the independence between each item, the independence between each examinee and the independence between items and examinees so that researchers can separately estimate the item parameters and ability parameters. By maximizing the marginal log-likelihood function, the best estimates of the item parameters can be obtained. During the process, numerical integration and artificial data are used to simplify the calculation. This method was commonly accepted in the estimation of the 3PLM, although it often yields infinite or implausible parameter estimates in small samples (Mislevy, 1986). When applied to 4PLM, the algorithm has difficulties with convergence due to the complexity of the derivatives. Feuerstahler and Waller (2014) used the MMLE to estimate 4PLM item parameters. The sample size which is needed for reasonably accurate estimation is large (N=10,000).

The Bayesian EM algorithm. To eliminate the implausible estimates in MMLE, researchers introduced a Bayesian method. Mislevy (1986) considered the prior distribution of item parameters, and proposed a general formula:

$$0 = \frac{\partial \ln M(U \mid \xi)}{\partial \psi_j} + \frac{\partial \ln g(\psi_j \mid \eta)}{\partial \psi_j}. \tag{1.2}$$

where $\ln M(U \mid \xi)$ is the log-likelihood function in the MMLE/EM, and $g(\psi_j \mid \eta)$ is the prior distribution for item parameters. The priors provide more information which helps avoid the unstable estimation caused by using uninformative data. This is the reason why the Bayesian method can solve the problem of infinite or implausible estimates in traditional MMLE/EM. Waller and Feuerstahler (2017) estimated the item parameters of the MMPI-A scale (Butcher, et al., 1992) using BME (Bayesian model estimation). The results showed that at least 66% of the items require the upper asymptotes, which helps them assert that the 4PM is needed when modeling such self-report data (p.6). However, the defect of this method is that the quality of results is highly dependent on the properness of priors. If the priors are deflective, the estimates will be inaccurate.

The MCMC algorithm. Culpepper (2016) presented the Bayesian formulation for 4PNO by defining a binary augmented variable W_{ij}. The 4PNO is assumed to be expressed as:

$$P(Y_{ij}=1|\theta_i,\xi_j,\gamma_j,\varsigma_j)=\gamma_j+(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij}). \tag{1.3}$$

where $P(Y_{ij}=1)$ denotes the probability of correct response for examinee i on item j, $\xi_j=(\alpha_j,\beta_j)$ are the slope and threshold parameters for item j, γ_j and $1-\varsigma_j$ are the lower and upper asymptotes, θ_i is the latent trait score, $\eta_{ij}=\alpha_j\theta_i-\beta_j$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Following Béguin and Glas (2001), W_{ij} was related to θ_i through a two-parameter normal ogive model as $P(W_{ij}=1|\eta_{ij})=\Phi(\eta_{ij})$. The relationship between ability and response can then be modeled through Y_{ij} conditional on W_{ij}:

$$P(Y_{ij}|W_{ij})=\begin{cases}\gamma_j^{Y_{ij}}(1-\gamma_j)^{1-Y_{ij}},\ W_{ij}=0\\(1-\varsigma_j)^{Y_{ij}}\varsigma_j^{1-Y_{ij}},\ W_{ij}=1\end{cases}. \tag{1.4}$$

The conditional probability $P(Y_{ij}=1|W_{ij}=0)=\gamma_j$ is the probability that student i who does not know the answer to item j correctly guesses, $P(Y_{ij}=1|W_{ij}=1)=\varsigma_j$ is the probability that student i knows the correct answer, but "slips" and provides an incorrect response (Culpepper, 2016). Through a series of derivations, the probabilities of W_{ij} condition on $Y_{ij},\eta_{ij},\gamma_j,\varsigma_j$ are attained:

$$
\begin{aligned}
P(W_{ij}=1 \mid Y_{ij}=1,\eta_{ij},\gamma_j,\varsigma_j) &== \frac{(1-\varsigma_j)\Phi(\eta_{ij})}{\gamma_j+(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij})} \\
P(W_{ij}=0 \mid Y_{ij}=1,\eta_{ij},\gamma_j,\varsigma_j) &== \frac{\gamma_j(1-\Phi(\eta_{ij}))}{\gamma_j+(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij})} \\
P(W_{ij}=1 \mid Y_{ij}=0,\eta_{ij},\gamma_j,\varsigma_j) &== \frac{\varsigma_j\Phi(\eta_{ij})}{1-\gamma_j-(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij})} \\
P(W_{ij}=0 \mid Y_{ij}=0,\eta_{ij},\gamma_j,\varsigma_j) &== \frac{(1-\gamma_j)(1-\Phi(\eta_{ij}))}{1-\gamma_j-(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij})}.
\end{aligned} \tag{1.5}
$$

Given the data and model parameters, W_{ij} can be sampled. Albert (1992) derived the full conditional for θ_i:

$$
p(\theta \mid Z,\xi) = \prod_{i=1}^{N}\prod_{j=1}^{J}\phi(Z_{ij} \mid \theta_i,\xi_j)\phi(\theta_i;\mu_\theta,\sigma_\theta^2), \tag{1.6}
$$

where Z_{ij} a continuous, normally distributed random variable. Fox (2010) showed the way to sample the item threshold and slope parameters:

$$
p(\xi \mid Z,\theta) = \prod_{i=1}^{N}\prod_{j=1}^{J}\phi(Z_{ij} \mid \theta_i,\xi_j)\phi(\xi_j;\mu_\xi,\Sigma_\xi)\mathrm{I}(\alpha_j>0). \tag{1.7}
$$

The full conditional distribution for γ_j and ς_j is:

$$
p(\gamma_j,\varsigma_j \mid Y_j,W_j) \propto p(Y_j \mid W_j,\gamma_j,\varsigma_j)\,p(\gamma_j,\varsigma_j). \tag{1.8}
$$

Let $f_{\gamma\varsigma}$ denote the joint probability function distribution, and assume:

$$
f_{\gamma\varsigma} \propto f_\gamma f_\varsigma \mathrm{I}((\gamma,\varsigma)\in\Omega). \tag{1.9}
$$

The marginal distribution f_γ and conditional distribution $f_{\varsigma|\gamma}$ can then be obtained. Consequently, γ_j can be sampled from f_γ through a Gibbs-within-Gibbs sampler, ς_j can be sampled as $\varsigma_j \mid \gamma_j \sim Beta(\tilde{a}_\varsigma,\tilde{b}_\varsigma)\mathrm{I}(0\le\varsigma_j\le 1-\gamma_j)$.

CHAPTER 2: THE EMMM AND THE BAYESIAN EMMM ALGORITHM FOR THE 4PLM

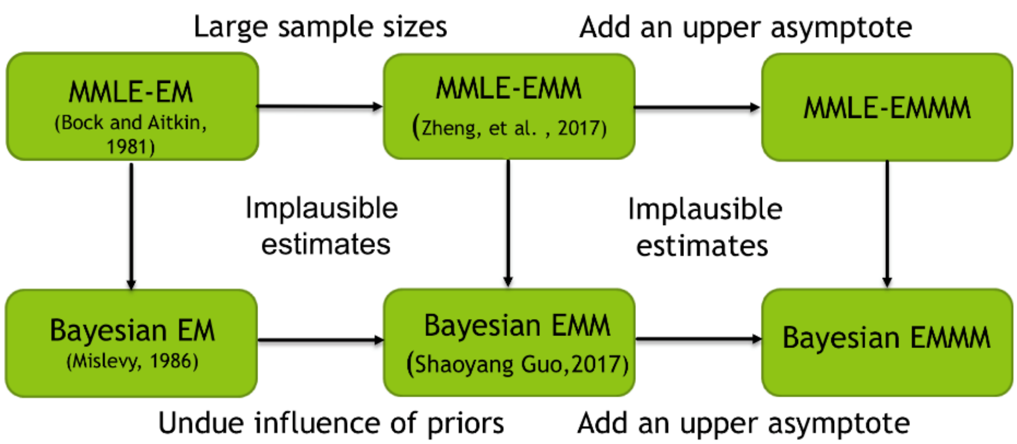

Figure 2.1 The development process of the EMMM and the BEMMM

2.1 The EMMM algorithm

The EMM algorithm. Before introducing the Expectation-Maximization-Maximization-Maximization (EMMM) algorithm, the author reviewed the Expectation-Maximization-Maximization (EMM) proposed by Zheng, Meng, Guo and Liu (2017). They rewrote the 3PLM as follows:

$$P(u_{ij}=1|\theta_i,a_j,b_j,c_j)=c_j\times\left[1\right]+(1-c_j)\times\left[P_j^*(\theta_i)\right], \qquad (2.1)$$

with $P_j^*(\theta_i)=\dfrac{1}{1+\exp\left(-Da_j(\theta_i-b_j)\right)}$.

A latent indicator variable is thus defined as:

$$z_{ij}=\begin{cases}1 \text{ if examinee } i \text{ does not guess on item } j;\\ 0 \text{ if examinee } i \text{ does guess on item } j.\end{cases} \qquad (2.2)$$

Here, $z_{ij} \sim Bernoulli\left(1-c_j\right)$. The probability of an examinee using a guessing strategy is $P\left(z_{ij}=0\right)=c_j$. The conditional possibilities of responses to items conditional on $z_{ij},\ \theta_i,\ \boldsymbol{\xi}_j$ are:

$$\begin{aligned}
&P(u_{ij}=1 \mid z_{ij}=1,\theta_i,\boldsymbol{\xi}_j)=P_j^*(\theta_i)\\
&P(u_{ij}=1 \mid z_{ij}=0,\theta_i,\boldsymbol{\xi}_j)=1\\
&P(u_{ij}=0 \mid z_{ij}=1,\theta_i,\boldsymbol{\xi}_j)=1-P_j^*(\theta_i)\\
&P(u_{ij}=0 \mid z_{ij}=0,\theta_i,\boldsymbol{\xi}_j)=0.
\end{aligned} \tag{2.3}$$

Since the joint distribution of u_{ij} and z_{ij} can be expressed as $P(u_{ij}, z_{ij} \mid \theta_i, \boldsymbol{\xi}_j) = P(u_{ij} \mid z_{ij}, \theta_i, \boldsymbol{\xi}_j) P(z_{ij})$:

$$\begin{aligned}
&P(u_{ij}=1, z_{ij}=1 \mid \theta_i,\boldsymbol{\xi}_j)=(1-c_j)P_j^*(\theta_i)\\
&P(u_{ij}=1, z_{ij}=0 \mid \theta_i,\boldsymbol{\xi}_j)=c_j\\
&P(u_{ij}=0, z_{ij}=1 \mid \theta_i,\boldsymbol{\xi}_j)=(1-c_j)(1-P_j^*(\theta_i))\\
&P(u_{ij}=0, z_{ij}=0 \mid \theta_i,\boldsymbol{\xi}_j)=0.
\end{aligned} \tag{2.4}$$

Let u_i and z_i denote the response and the latent indicator vector, $\boldsymbol{\tau}$ is the vector containing the parameters of the examinee population ability distribution, so that the joint distribution for complete data is:

$$P(\mathbf{u}_i, \mathbf{z}_i, \theta_i \mid \boldsymbol{\xi}, \boldsymbol{\tau}) = P(\mathbf{u}_i, \mathbf{z}_i \mid \theta_i, \boldsymbol{\xi}) g(\theta_i \mid \boldsymbol{\tau}), \tag{2.5}$$

where

$$P(\mathbf{u}_i, \mathbf{z}_i \mid \theta_i, \boldsymbol{\xi}) = \prod_{j=1}^{n} [(1-c_j)P_j^*(\theta_i)]^{u_{ij}z_{ij}} \times c_j^{u_{ij}(1-z_{ij})} \times [(1-c_j)\left(1-P_j^*(\theta_i)\right)]^{(1-u_{ij})z_{ij}}, \tag{2.6}$$

$g\left(\theta_i \mid \boldsymbol{\tau}\right)$ is a density function of θ_i. By integrating to θ_i, the marginal distribution becomes

$$P(\mathbf{u}_i, \mathbf{z}_i \mid \xi) = \int_{\theta_i} P(\mathbf{u}_i, \mathbf{z}_i \mid \xi, \theta_i) g(\theta_i \mid \boldsymbol{\tau}) d\theta_i. \tag{2.7}$$

For all examinees, the likelihood function is:

$$L(\mathbf{U}, Z \mid \xi) = \prod_{i=1}^{N} P(\mathbf{u}_i, \mathbf{z}_i \mid \xi)$$

$$= \prod_{i=1}^{N} \int_{\theta_i} P(\mathbf{u}_i, \mathbf{z}_i \mid \xi, \theta_i) g(\theta_i \mid \boldsymbol{\tau}) d\theta_i. \tag{2.8}$$

The difference between the EMM algorithm and the EM algorithm is such that the maximization step for c_j is separated from that for a_j and b_j, so the estimations of c_j and (a_j , b_j) are independent. Zheng, et al. (2017) compared the EMM with Bayesian EM in BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 1996). The results verified the feasibility of the new algorithm with a small sample size (N=1000).

The EMMM algorithm. Continuing with the introduction of a latent variable which was inspired by Culpepper (2016), the author introduces a discrete augmented variable W_{ij} to the 4PLM and proposes the Expectation-Maximization-Maximization-Maximization (EMMM) algorithm. The expression of 4PLM is:

$$P(Y_{ij} = 1 \mid \theta_i, a_j, b_j, \gamma_j, \varsigma_j) = \gamma_j + \frac{(1-\varsigma_j) - \gamma_j}{1 + \exp(-Da_j(\theta_i - b_j))}, \tag{2.9}$$

which can be rewritten as:

$$P(Y_{ij} = 1 \mid \theta_i, a_j, b_j, \gamma_j, \varsigma_j) = \gamma_j \left(1 - P_j^*(\theta_i)\right) + (1-\varsigma_j) P_j^*(\theta_i), \tag{2.10}$$

where $P_j^*(\theta_i) = \dfrac{1}{1 + \exp(-Da_j(\theta_i - b_j))}$.

Let $P(Y_{ij}=1)$ denotes the probability of response for examinee i on item j. The a_j and b_j are the discrimination and difficulty parameters, γ_j and $1-\varsigma_j$ are the lower and upper asymptotes, θ_i is the latent trait score, and D is 1.702.

Using the definition of W_{ij} given by Culpepper (2016),

$$W_{ij}=\begin{cases}1 \text{ if examinee i knows the correct answer to item j} \\ 0 \text{ if examinee i does not know the answer to item j}\end{cases}, \tag{2.11}$$

the distribution of W_{ij} is

$$\begin{aligned} P\left(W_{ij}=1\right) &= P_j^*\left(\theta_i\right) \\ P\left(W_{ij}=0\right) &= 1-P_j^*\left(\theta_i\right), \end{aligned} \tag{2.12}$$

that is, $W_{ij} \sim Bernoulli(P_j^*(\theta_i))$. The probability of an examinee's response Y_{ij} condition on W_{ij} are then (let ψ_j denote the item parameters vector $\left\{a_j, b_j, \gamma_j, \varsigma_j\right\}$):

$$\begin{aligned} P\left(Y_{ij}=1 \mid W_{ij}=1,\theta_i,\psi_j\right) &= 1-\varsigma_j \\ P\left(Y_{ij}=1 \mid W_{ij}=0,\theta_i,\psi_j\right) &= \gamma_j \\ P\left(Y_{ij}=0 \mid W_{ij}=1,\theta_i,\psi_j\right) &= \varsigma_j \\ P\left(Y_{ij}=0 \mid W_{ij}=0,\theta_i,\psi_j\right) &= 1-\gamma_j. \end{aligned} \tag{2.13}$$

The interpretation of the probabilities is comprehensible. When $W_{ij}=1$, the examinee who knows the correct answer still have a probability of ς_j making mistakes. When an examinee does not know the correct answer, the probability of correctly guessing is γ_j. By multiplying the conditional probability $P\left(Y_{ij} \mid W_{ij},\theta_i,\psi_j\right)$ by $P\left(W_{ij}\right)$, the joint distribution is

$$P\left(Y_{ij}=1,W_{ij}=1 \mid \theta_i,\psi_j\right)=P\left(Y_{ij}=1 \mid W_{ij}=1,\theta_i,\psi_j\right)P\left(W_{ij}=1\right)=(1-\varsigma_j)P_j^*\left(\theta_i\right)$$

$$
\begin{aligned}
&P\left(Y_{ij}=1,W_{ij}=0\mid\theta_i,\psi_j\right)=P\left(Y_{ij}=1\mid W_{ij}=0,\theta_i,\psi_j\right)P\left(W_{ij}=0\right)=\gamma_j\left(1-P_j^*\left(\theta_i\right)\right)\\
&P\left(Y_{ij}=0,W_{ij}=1\mid\theta_i,\psi_j\right)=P\left(Y_{ij}=0\mid W_{ij}=1,\theta_i,\psi_j\right)P\left(W_{ij}=1\right)=\varsigma_jP_j^*\left(\theta_i\right)\\
&P\left(Y_{ij}=0,W_{ij}=0\mid\theta_i,\psi_j\right)=P\left(Y_{ij}=0\mid W_{ij}=0,\theta_i,\psi_j\right)P\left(W_{ij}=0\right)=(1-\gamma_j)\left(1-P_j^*\left(\theta_i\right)\right).
\end{aligned}\tag{2.14}
$$

Using the Bayesian rule, the probability of the W_{ij} condition on Y_{ij} is

$$
\begin{aligned}
P(W_{ij}=1\mid Y_{ij}=1,\theta_i,\psi_j)&=\frac{P(W_{ij}=1,Y_{ij}=1\mid\theta_i,\psi_j)}{P(Y_{ij}=1\mid\theta_i,\psi_j)}=\frac{(1-\varsigma_j)P_j^*\left(\theta_i\right)}{P(\theta_i)}\\
P(W_{ij}=1\mid Y_{ij}=0,\theta_i,\psi_j)&=\frac{P(W_{ij}=1,Y_{ij}=0\mid\theta_i,\psi_j)}{P(Y_{ij}=0\mid\theta_i,\psi_j)}=\frac{\varsigma_jP_j^*\left(\theta_i\right)}{1-P(\theta_i)}.
\end{aligned}\tag{2.15}
$$

In the previous section, for the 4PNO model, gave the conditional probability, that is

$$
\begin{aligned}
P(W_{ij}=1\mid Y_{ij}=1,\eta_{ij},\gamma_j,\varsigma_j)&==\frac{(1-\varsigma_j)\Phi(\eta_{ij})}{\gamma_j+(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij})}=\frac{(1-\varsigma_j)\Phi(\eta_{ij})}{P(Y_{ij}=1)}\\
P(W_{ij}=1\mid Y_{ij}=0,\eta_{ij},\gamma_j,\varsigma_j)&==\frac{\varsigma_j\Phi(\eta_{ij})}{1-\gamma_j-(1-\varsigma_j-\gamma_j)\Phi(\eta_{ij})}=\frac{\varsigma_j\Phi(\eta_{ij})}{1-P(Y_{ij}=1)},
\end{aligned}\tag{1.5}
$$

note the similar format for the conditional probability in the BEMMM.

The expectation is:

$$
E(W_{ij}\mid Y_{ij},\theta_i,\psi_j)=Y_{ij}\left[\frac{(1-\varsigma_j)P_j^*\left(\theta_i\right)}{P(\theta_i)}\right]+(1-Y_{ij})\left[\frac{\varsigma_jP_j^*\left(\theta_i\right)}{1-P(\theta_i)}\right].\tag{2.16}
$$

The joint distribution of $\left(Y_{ij},W_{ij}\mid\theta_i,\psi_j\right)$ has been obtained, but the ability parameters are latent. As in the case of MMLE/EM, for examinee i, the researcher calculates the joint distribution by using:

$$
P\left(Y_i,W_i,\theta_i\mid\psi,\tau\right)=P(Y_i,W_i\mid\theta_i,\psi)g(\theta_i\mid\tau),\tag{2.17}
$$

where

$$
\begin{aligned}
P(Y_i,W_i\mid\theta_i,\psi)=\prod_{j=1}^{m}\Big\{&\left[(1-\varsigma_j)P_j^*\left(\theta_i\right)\right]^{Y_{ij}W_{ij}}\times\left[\gamma_j\left(1-P_j^*\left(\theta_i\right)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\\
&\times\left[\varsigma_jP_j^*\left(\theta_i\right)\right]^{\left(1-Y_{ij}\right)W_{ij}}\times\left[\left(1-\gamma_j\right)\left(1-P_j^*\left(\theta_i\right)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}\Big\},
\end{aligned}\tag{2.18}
$$

and $g(\theta_i \mid \tau)$ is a density function of θ and τ containing the parameters of the examinee population ability distribution. Following Bock and Lieberman (1970), the marginal distribution for a single examinee i by integrating over the ability parameters is

$$P(Y_i, W_i \mid \psi) = \int_{\theta_i} P(Y_i, W_i \mid \psi, \theta_i) g(\theta_i \mid \boldsymbol{\tau}) d\theta_i. \tag{2.19}$$

The likelihood function for all of the examinees is

$$\begin{aligned} L\left(Y, W \mid \psi\right) &= \prod_{i=1}^{N} P\left(Y_i, W_i \mid \psi\right) \\ &= \prod_{i=1}^{N} \int_{\theta_i} P\left(Y_i, W_i \mid \psi, \theta_i\right) g\left(\theta_i \mid \tau\right) d\theta_i. \end{aligned} \tag{2.20}$$

Plugging equation (2.17) into (2.19), the likelihood function of EMMM becomes:

$$\begin{aligned} L\left(Y, W \mid \psi\right) = \prod_{i=1}^{N} \Bigg\{ \int_{\theta_i} \prod_{j=1}^{m} & \left[(1-\varsigma_j) P_j^*\left(\theta_i\right)\right]^{Y_{ij} W_{ij}} \times \left[\gamma_j \left(1 - P_j^*\left(\theta_i\right)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)} \\ & \times \left[\varsigma_j P_j^*\left(\theta_i\right)\right]^{\left(1-Y_{ij}\right) W_{ij}} \times \left[\left(1-\gamma_j\right)\left(1 - P_j^*\left(\theta_i\right)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)} g\left(\theta_i \mid \tau\right) d\theta \Bigg\}. \end{aligned} \tag{2.21}$$

2.2 The Bayesian EMMM algorithm

Mislevy (1986) used Bayesian method to the MMLE/EM algorithm and proposed the BME approach, which estimated the 3PL model accurately and solved the implausible issue in EM. Shaoyang Guo (2017) applied the Bayesian method to EMM algorithm (Zheng et al., 2017) and proposed the BEMM. The current study combines the Bayesian method with the EMMM and proposes the BEMMM algorithm.

Using the general Bayesian formulation (Mislevy, 1986), the author obtains the best estimates of the item parameters when the following equation holds:

$$0 = \frac{\partial \ln L(U \mid \xi)}{\partial \psi_j} + \frac{\partial \ln g(\psi_j \mid \eta)}{\partial \psi_j}. \tag{2.22}$$

The $L(U \mid \xi)$ is similar to the likelihood function for EMMM in the previous section but using the logarithmic form of a_j, $g(\psi_j \mid \eta)$ are the prior distributions for item parameters. Their first and second derivations are:

$$
\begin{aligned}
&\frac{\partial \ln g\left(\ln a_j \mid \mu_{\ln a_j}, \sigma^2_{\ln a_j}\right)}{\partial \ln a_j} = -\frac{\ln a_j - \mu_{\ln a_j}}{\sigma^2_{\ln a_j}}, \frac{\partial^2 \ln g\left(\ln a_j \mid \mu_{\ln a_j}, \sigma^2_{\ln a_j}\right)}{\partial \ln a_j \partial \ln a_j} = -\frac{1}{\sigma^2_{\ln a_j}} \\
&\frac{\partial \ln g\left(b_j \mid \mu_{b_j}, \sigma^2_{b_j}\right)}{\partial b_j} = -\frac{b_j - \mu_{b_j}}{\sigma^2_{b_j}}, \qquad \frac{\partial^2 \ln g\left(b_j \mid \mu_{b_j}, \sigma^2_{b_j}\right)}{\partial b_j \partial b_j} = -\frac{1}{\sigma^2_{b_j}} \\
&\frac{\partial \ln g\left(\gamma_j \mid \alpha_j^\gamma, \beta_j^\gamma\right)}{\partial \gamma_j} = \frac{\alpha_j^\gamma - 1}{\gamma_j} - \frac{\beta_j^\gamma - 1}{1-\gamma_j}, \quad \frac{\partial^2 \ln g\left(\gamma_j \mid \alpha_j^\gamma, \beta_j^\gamma\right)}{\partial \gamma_j \partial \gamma_j} = -\frac{\alpha_j^\gamma - 1}{\gamma_j^2} - \frac{\beta_j^\gamma - 1}{(1-\gamma_j)^2} \\
&\frac{\partial \ln g\left(\varsigma_j \mid \alpha_j^\varsigma, \beta_j^\varsigma\right)}{\partial \varsigma_j} = \frac{\alpha_j^\varsigma - 1}{\varsigma_j} - \frac{\beta_j^\varsigma - 1}{1-\varsigma_j}, \quad \frac{\partial^2 \ln g\left(\varsigma_j \mid \alpha_j^\varsigma, \beta_j^\varsigma\right)}{\partial \varsigma_j \partial \varsigma_j} = -\frac{\alpha_j^\varsigma - 1}{\varsigma_j^2} - \frac{\beta_j^\varsigma - 1}{(1-\varsigma_j)^2}.
\end{aligned}
\tag{2.23}
$$

In general, the BEMMM algorithm is illustrated in the flow chart below:

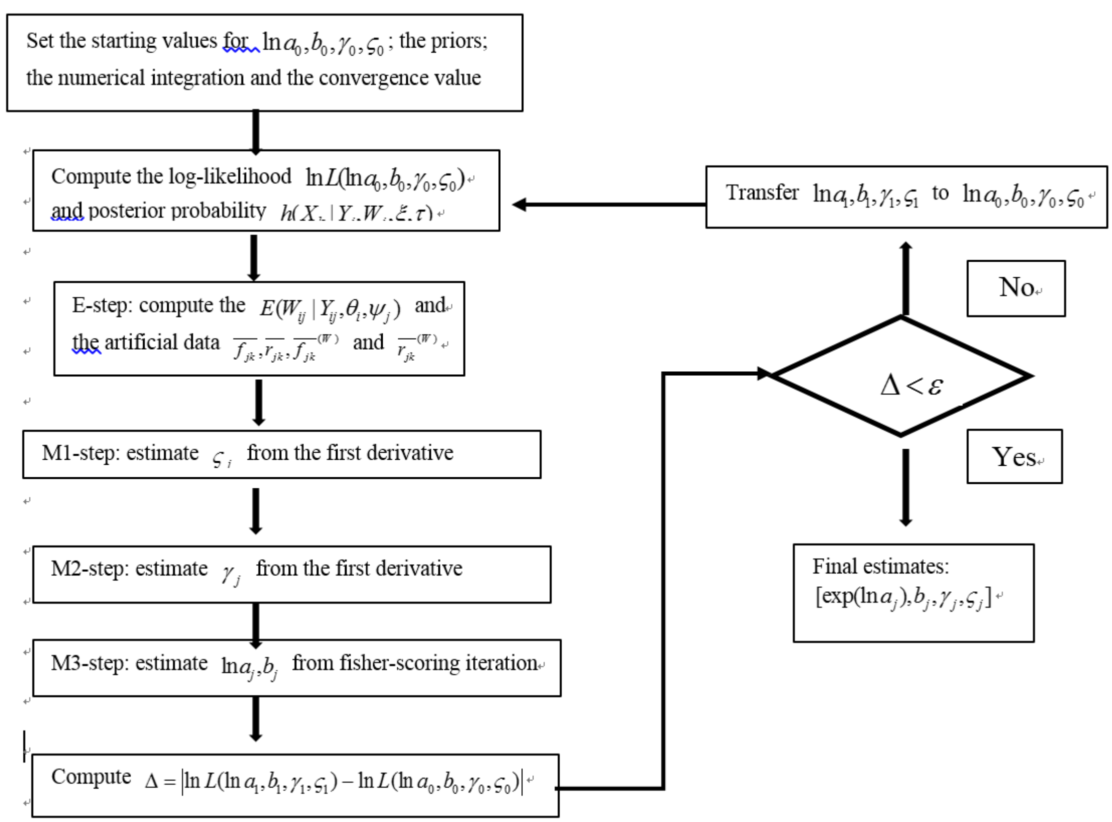

Figure 2.2 The Flow Chart of the Bayesian EMMM

Expectation step and artificial data. Following the E-step in the EM algorithm (Bock & Aitkin, 1981), the best estimates are attained when the first derivation of the logarithmic likelihood function of the Bayesian EMMM is equal to 0.

$$
0=\frac{\partial \ln L}{\partial \psi_j}=\frac{\partial \ln\left(\prod_{i=1}^{N} P\left(Y_i,W_i \mid \psi\right)\right)}{\partial \psi_j}=\sum_{i=1}^{N}\frac{\ln P\left(Y_i,W_i \mid \psi\right)}{\partial \psi_j}=\sum_{i=1}^{N}\frac{1}{P\left(Y_i,W_i \mid \psi\right)}\frac{\partial P\left(Y_i,W_i \mid \psi\right)}{\partial \psi_j}
$$

$$
=\sum_{i=1}^{N}\frac{1}{P\left(Y_i,W_i \mid \psi\right)}\frac{\partial\left(\int_{\theta_i} P(Y_i,W_i \mid \psi,\theta_i)g(\theta_i \mid \boldsymbol{\tau})d\theta_i\right)}{\partial \psi_j}
$$

$$
=\sum_{i=1}^{N}\frac{1}{P\left(Y_i,W_i \mid \psi\right)}\int_{\theta_i}\left[\frac{\partial \ln P(Y_i,W_i \mid \psi,\theta_i)}{\partial \psi_j}\right]P(Y_i,W_i \mid \psi,\theta_i)g(\theta_i \mid \boldsymbol{\tau})d\theta_i
$$

$$
=\sum_{i=1}^{N}\int_{\theta_i}\left[\frac{\partial \ln P(Y_i,W_i \mid \psi,\theta_i)}{\partial \psi_j}\right]\left[\frac{P(Y_i,W_i \mid \psi,\theta_i)g(\theta_i \mid \boldsymbol{\tau})}{P\left(Y_i,W_i \mid \psi\right)}\right]d\theta_i
$$

$$
=\sum_{i=1}^{N}\int_{\theta_i}\left[\frac{\partial \ln P(Y_i,W_i \mid \psi,\theta_i)}{\partial \psi_j}\right]P(\theta_i \mid Y_i,W_i,\psi,\boldsymbol{\tau})d\theta_i
$$

$$
=\sum_{i=1}^{N}\int_{\theta_i}\left\{\frac{\partial}{\partial \psi_j}\ln\left(\prod_{j=1}^{m}\left[(1-\varsigma_j)P_j^*\left(\theta_i\right)\right]^{Y_{ij}W_{ij}}\left[\gamma_j\left(1-P_j^*\left(\theta_i\right)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\left[\varsigma_j P_j^*\left(\theta_i\right)\right]^{\left(1-Y_{ij}\right)W_{ij}}\left[\left(1-\gamma_j\right)\left(1-P_j^*\left(\theta_i\right)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}\right)\right\}P(\theta_i \mid Y_i,W_i,\psi,\boldsymbol{\tau})d\theta_i
$$

$$=\sum_{i=1}^{N}\int_{\theta_i}\left\{\frac{\dfrac{\partial}{\partial\psi_j}\prod_{j=1}^{m}\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}}\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\left[\varsigma_jP_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}}\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}}{\prod_{j=1}^{m}\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}}\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\left[\varsigma_jP_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}}\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}}\right\}P(\theta_i\,|\,Y_i,W_i,\psi,\boldsymbol{\tau})d\theta_i$$

$$=\sum_{i=1}^{N}\int_{\theta_i}\left\{\frac{\dfrac{\partial}{\partial\psi_j}\begin{bmatrix}\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}}\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\left[\varsigma_jP_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}}\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}\\ \times\prod_{h\neq j}^{m}\left[(1-\varsigma_h)P_h^*(\theta_i)\right]^{Y_{ih}W_{ih}}\left[\gamma_h\left(1-P_h^*(\theta_i)\right)\right]^{Y_{ih}\left(1-W_{ih}\right)}\left[\varsigma_hP_h^*(\theta_i)\right]^{\left(1-Y_{ih}\right)W_{ih}}\left[\left(1-\gamma_h\right)\left(1-P_h^*(\theta_i)\right)\right]^{\left(1-Y_{ih}\right)\left(1-W_{ih}\right)}\end{bmatrix}}{\begin{bmatrix}\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}}\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\left[\varsigma_jP_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}}\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}\\ \times\prod_{h\neq j}^{m}\left[(1-\varsigma_h)P_h^*(\theta_i)\right]^{Y_{ih}W_{ih}}\left[\gamma_h\left(1-P_h^*(\theta_i)\right)\right]^{Y_{ih}\left(1-W_{ih}\right)}\left[\varsigma_hP_h^*(\theta_i)\right]^{\left(1-Y_{ih}\right)W_{ih}}\left[\left(1-\gamma_h\right)\left(1-P_h^*(\theta_i)\right)\right]^{\left(1-Y_{ih}\right)\left(1-W_{ih}\right)}\end{bmatrix}}\right\}P(\theta_i\,|\,Y_i,W_i,\psi,\boldsymbol{\tau})d\theta_i$$

Let $R_h=\left[(1-\varsigma_h)P_h^*(\theta_i)\right]^{Y_{ih}W_{ih}}\left[\gamma_h\left(1-P_h^*(\theta_i)\right)\right]^{Y_{ih}\left(1-W_{ih}\right)}\left[\varsigma_hP_h^*(\theta_i)\right]^{\left(1-Y_{ih}\right)W_{ih}}\left[\left(1-\gamma_h\right)\left(1-P_h^*(\theta_i)\right)\right]^{\left(1-Y_{ih}\right)\left(1-W_{ih}\right)}$

$$R_j=\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}}\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\left[\varsigma_jP_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}}\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}$$

$$RHS=\sum_{i=1}^{N}\int_{\theta_i}\left\{\frac{\dfrac{\partial}{\partial\psi_j}\left(R_j\times\prod_{h\neq j}^{m}R_h\right)}{R_j\times\prod_{h\neq j}^{m}R_h}\right\}P(\theta_i\,|\,Y_i,W_i,\psi,\boldsymbol{\tau})d\theta_i$$

$$= \sum_{i=1}^{N} \int_{\theta_i} \left[\frac{\left(\frac{\partial R_j}{\partial \psi_j} \times \prod_{h \neq j}^{m} R_h \right) + \left(R_j \times \frac{\partial}{\partial \psi_j} \prod_{h \neq j}^{m} R_h \right)}{R_j \times \prod_{h \neq j}^{m} R_h} \right] P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

$$= \sum_{i=1}^{N} \int_{\theta_i} \left[\frac{\left(\frac{\partial R_j}{\partial \psi_j} \times \prod_{h \neq j}^{m} R_h \right) + \left(R_j \times 0 \right)}{R_j \times \prod_{h \neq j}^{m} R_h} \right] P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

$$= \sum_{i=1}^{N} \int_{\theta_i} \left(\frac{1}{R_j} \times \frac{\partial R_j}{\partial \psi_j} \right) P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

$$= \sum_{i=1}^{N} \int_{\theta_i} \left\{ \frac{1}{R_j} \frac{\partial}{\partial \psi_j} \left[(1-\varsigma_j) P_j^*(\theta_i) \right]^{Y_{ij} W_{ij}} \left[\gamma_j \left(1 - P_j^*(\theta_i) \right) \right]^{Y_{ij}(1-W_{ij})} \left[\varsigma_j P_j^*(\theta_i) \right]^{(1-Y_{ij}) W_{ij}} \left[\left(1 - \gamma_j \right) \left(1 - P_j^*(\theta_i) \right) \right]^{(1-Y_{ij})(1-W_{ij})} \right\} P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

Let $A = \left[(1-\varsigma_j) P_j^*(\theta_i) \right]^{Y_{ij} W_{ij}}$, $B = \left[\gamma_j \left(1 - P_j^*(\theta_i) \right) \right]^{Y_{ij}(1-W_{ij})}$, $C = \left[\varsigma_j P_j^*(\theta_i) \right]^{(1-Y_{ij}) W_{ij}}$, $D = \left[\left(1 - \gamma_j \right) \left(1 - P_j^*(\theta_i) \right) \right]^{(1-Y_{ij})(1-W_{ij})}$,

$$RHS = \sum_{i=1}^{N} \int_{\theta_i} \left\{ \frac{1}{R_j} \left[\frac{\partial (1-\varsigma_j) P_j^*(\theta_i)}{\partial \psi_j} Y_{ij} W_{ij} \left[(1-\varsigma_j) P_j^*(\theta_i) \right]^{Y_{ij} W_{ij} - 1} BCD \right. \right.$$

$$\left. + \frac{\partial \gamma_j \left(1 - P_j^*(\theta_i) \right)}{\partial \psi_j} Y_{ij} \left(1 - W_{ij} \right) \left[\gamma_j \left(1 - P_j^*(\theta_i) \right) \right]^{Y_{ij}(1-W_{ij}) - 1} ACD \right.$$

$$+\frac{\partial \varsigma_j P_j^*(\theta_i)}{\partial \psi_j}\left(1-Y_{ij}\right)W_{ij}\left[\varsigma_j P_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}-1} ABD$$

$$\left.\left.+\frac{\partial\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)}{\partial \psi_j}\left(1-Y_{ij}\right)\left(1-W_{ij}\right)\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)-1} BCD\right]\right\} P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

$$=\sum_{i=1}^{N}\int_{\theta_i}\left\{\frac{1}{R_j}\left(\frac{\partial(1-\varsigma_j)P_j^*(\theta_i)}{\partial \psi_j}\frac{Y_{ij}W_{ij}}{(1-\varsigma_j)P_j^*(\theta_i)}\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}} BCD\right.\right.$$

$$+\frac{\partial \gamma_j\left(1-P_j^*(\theta_i)\right)}{\partial \psi_j}\frac{Y_{ij}\left(1-W_{ij}\right)}{\gamma_j\left(1-P_j^*(\theta_i)\right)}\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)} ACD$$

$$+\frac{\partial \varsigma_j P_j^*(\theta_i)}{\partial \psi_j}\frac{\left(1-Y_{ij}\right)W_{ij}}{\varsigma_j P_j^*(\theta_i)}\left[\varsigma_j P_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}} ABD$$

$$\left.\left.+\frac{\partial\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)}{\partial \psi_j}\frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)}\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)-1} BCD\right)\right\} P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

Where $R_j=\left[(1-\varsigma_j)P_j^*(\theta_i)\right]^{Y_{ij}W_{ij}}\times\left[\gamma_j\left(1-P_j^*(\theta_i)\right)\right]^{Y_{ij}\left(1-W_{ij}\right)}\times\left[\varsigma_j P_j^*(\theta_i)\right]^{\left(1-Y_{ij}\right)W_{ij}}\times\left[\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)\right]^{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}=ABCD$.

$$RHS=\sum_{i=1}^{N}\int_{\theta_i}\left\{\frac{1}{R_j}\times ABCD\times\left[\frac{\partial(1-\varsigma_j)P_j^*(\theta_i)}{\partial \psi_j}\frac{Y_{ij}W_{ij}}{(1-\varsigma_j)P_j^*(\theta_i)}+\frac{\partial \gamma_j\left(1-P_j^*(\theta_i)\right)}{\partial \psi_j}\frac{Y_{ij}\left(1-W_{ij}\right)}{\gamma_j\left(1-P_j^*(\theta_i)\right)}+\frac{\partial \varsigma_j P_j^*(\theta_i)}{\partial \psi_j}\frac{\left(1-Y_{ij}\right)W_{ij}}{\varsigma_j P_j^*(\theta_i)}\right.\right.$$

$$\left.\left.+\frac{\partial\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)}{\partial \psi_j}\frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)}\right]\right\} P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} \int_{\theta_i} \left[\frac{\partial (1-\varsigma_j) P_j^*(\theta_i)}{\partial \psi_j} \frac{Y_{ij} W_{ij}}{(1-\varsigma_j) P_j^*(\theta_i)} + \frac{\partial \gamma_j \left(1 - P_j^*(\theta_i)\right)}{\partial \psi_j} \frac{Y_{ij}\left(1-W_{ij}\right)}{\gamma_j \left(1-P_j^*(\theta_i)\right)} + \frac{\partial \varsigma_j P_j^*(\theta_i)}{\partial \psi_j} \frac{\left(1-Y_{ij}\right) W_{ij}}{\varsigma_j P_j^*(\theta_i)} \right. \\
&\qquad \left. + \frac{\partial \left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)}{\partial \psi_j} \frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)} \right] P(\theta_i \,|\, Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i \\
&= \sum_{i=1}^{N} \int_{\theta_i} \left\{ \left[P_j^*(\theta_i) \frac{\partial (1-\varsigma_j)}{\partial \psi_j} + (1-\varsigma_j) \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} \right] \times \frac{Y_{ij} W_{ij}}{(1-\varsigma_j) P_j^*(\theta_i)} + \left[\left(1-P_j^*(\theta_i)\right) \frac{\partial \gamma_j}{\partial \psi_j} + \gamma_j \frac{\partial \left(1-P_j^*(\theta_i)\right)}{\partial \psi_j} \right] \times \frac{Y_{ij}\left(1-W_{ij}\right)}{\gamma_j \left(1-P_j^*(\theta_i)\right)} \right. \\
&\qquad + \left[P_j^*(\theta_i) \frac{\partial \varsigma_j}{\partial \psi_j} + \varsigma_j \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} \right] \times \frac{\left(1-Y_{ij}\right) W_{ij}}{\varsigma_j P_j^*(\theta_i)} \\
&\qquad \left. + \left[\left(1-P_j^*(\theta_i)\right) \frac{\partial \left(1-\gamma_j\right)}{\partial \psi_j} + \left(1-\gamma_j\right) \frac{\partial \left(1-P_j^*(\theta_i)\right)}{\partial \psi_j} \right] \times \frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-\gamma_j\right)\left(1-P_j^*(\theta_i)\right)} \right\} P(\theta_i \,|\, Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i \\
&= \sum_{i=1}^{N} \int_{\theta_i} \left[\frac{Y_{ij} W_{ij}}{P_j^*(\theta_i)} \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} - \frac{Y_{ij} W_{ij}}{(1-\varsigma_j)} \frac{\partial \varsigma_j}{\partial \psi_j} + \frac{Y_{ij}\left(1-W_{ij}\right)}{\gamma_j} \frac{\partial \gamma_j}{\partial \psi_j} - \frac{Y_{ij}\left(1-W_{ij}\right)}{\left(1-P_j^*(\theta_i)\right)} \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} \right. \\
&\qquad \left. + \frac{\left(1-Y_{ij}\right) W_{ij}}{\varsigma_j} \frac{\partial \varsigma_j}{\partial \psi_j} + \frac{\left(1-Y_{ij}\right) W_{ij}}{P_j^*(\theta_i)} \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} - \frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-\gamma_j\right)} \frac{\partial \gamma_j}{\partial \psi_j} - \frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-P_j^*(\theta_i)\right)} \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} \right] P(\theta_i \,|\, Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i \\
&= \sum_{i=1}^{N} \int_{\theta_i} \left[\left(\frac{Y_{ij} W_{ij}}{P_j^*(\theta_i)} - \frac{Y_{ij}\left(1-W_{ij}\right)}{\left(1-P_j^*(\theta_i)\right)} + \frac{\left(1-Y_{ij}\right) W_{ij}}{P_j^*(\theta_i)} - \frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-P_j^*(\theta_i)\right)} \right) \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} \right. \\
&\qquad \left. + \left(\frac{\left(1-Y_{ij}\right) W_{ij}}{\varsigma_j} - \frac{Y_{ij} W_{ij}}{(1-\varsigma_j)} \right) \frac{\partial \varsigma_j}{\partial \psi_j} + \left(\frac{Y_{ij}\left(1-W_{ij}\right)}{\gamma_j} - \frac{\left(1-Y_{ij}\right)\left(1-W_{ij}\right)}{\left(1-\gamma_j\right)} \right) \frac{\partial \gamma_j}{\partial \psi_j} \right] P(\theta_i \,|\, Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i
\end{aligned}
$$

$$= \sum_{i=1}^{N} \int_{\theta_i} \left[\left(\frac{W_{ij}}{P_j^*(\theta_i)} - \frac{(1-W_{ij})}{(1-P_j^*(\theta_i))} \right) \frac{\partial P_j^*(\theta_i)}{\partial \psi_j} + \left(\frac{(1-Y_{ij})W_{ij}}{\varsigma_j} - \frac{Y_{ij}W_{ij}}{(1-\varsigma_j)} \right) \frac{\partial \varsigma_j}{\partial \psi_j} + \left(\frac{Y_{ij}(1-W_{ij})}{\gamma_j} - \frac{(1-Y_{ij})(1-W_{ij})}{(1-\gamma_j)} \right) \frac{\partial \gamma_j}{\partial \psi_j} \right] P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) d\theta_i$$

with

$$\begin{aligned} P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau}) &= P(\theta_i \mid Y_i, \psi, \boldsymbol{\tau}) = \frac{P(Y_i \mid \psi, \theta_i) g(\theta_i \mid \boldsymbol{\tau})}{\int_{\theta_i} P(Y_i \mid \psi, \theta_i) g(\theta_i \mid \boldsymbol{\tau}) d\theta_i} \\ P(Y_i \mid \psi, \theta_i) &= \prod_{j=1}^{m} P_j(\theta_i)^{Y_{ij}} \times (1 - P_j(\theta_i))^{1-Y_{ij}}, \end{aligned} \tag{2.24}$$

Where $P(\theta_i \mid Y_i, W_i, \psi, \boldsymbol{\tau})$ is the posterior probability of θ_i condition on $Y_i, W_i, \psi, \boldsymbol{\tau}$.

The expectation of W_i was obtained in the last part as

$$E(W_{ij} \mid Y_{ij}, \theta_i, \psi_j) = Y_{ij} \left[\frac{(1-\varsigma_j) P_j^*(\theta_i)}{P(\theta_i)} \right] + (1 - Y_{ij}) \left[\frac{\varsigma_j P_j^*(\theta_i)}{1 - P(\theta_i)} \right]. \tag{2.25}$$

Then the Hermite-Gauss quadrature method is used to approximate the integral, the expectation of the first derivative of the log-likelihood function is:

$$\frac{\partial \ln E(L)}{\partial \psi_j} = \sum_{i=1}^{N}\sum_{k=1}^{q}\left\{\left[\frac{E(W_{ij})}{P_j^*\left(X_k\right)} - \frac{\left(1-E(W_{ij})\right)}{\left(1-P_j^*\left(X_k\right)\right)}\right]\frac{\partial P_j^*\left(\theta_i\right)}{\partial \psi_j} + \left[\frac{\left(1-Y_{ij}\right)E(W_{ij})}{\varsigma_j} - \frac{Y_{ij}E(W_{ij})}{(1-\varsigma_j)}\right]\frac{\partial \varsigma_j}{\partial \psi_j}\right.$$
$$\left. + \left[\frac{Y_{ij}\left(1-E(W_{ij})\right)}{\gamma_j} - \frac{\left(1-Y_{ij}\right)\left(1-E(W_{ij})\right)}{\left(1-\gamma_j\right)}\right]\frac{\partial \gamma_j}{\partial \psi_j}\right\} P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau})$$

(2.26)

with

$$P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) = P(X_k \mid Y_i, \psi, \boldsymbol{\tau}) = \frac{P(Y_i \mid X_k, \psi)A\left(X_k\right)}{\sum_{k=1}^{q} P(Y_i \mid X_k, \psi)A\left(X_k\right)} \qquad (2.27)$$
$$P(Y_i \mid X_k, \psi) = \prod_{i=1}^{n} P_i(X_k)^{Y_{ij}} \times \left(1-P_i\left(X_k\right)\right)^{1-Y_{ij}},$$

where $X_k (k = 1, 2, ..., q)$ are nodes on the ability scale with an associated weight $A\left(X_k\right)$. $P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau})$ is the posterior probability of θ_i at X_k, which is equal to $P(X_k \mid Y_i, \psi, \boldsymbol{\tau})$ given that X_k is independent with W_i. In MMLE/EM, two artificial data are defined as (Bock & Aitkin, 1981):

$$\overline{f_{jk}} = \sum_{i=1}^{n} P(X_k \mid Y_i, \psi, \boldsymbol{\tau}) = \sum_{i=1}^{n} P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau})$$
$$\overline{r_{jk}} = \sum_{i=1}^{n} Y_{ij} \times P(X_k \mid Y_i, \psi, \boldsymbol{\tau}) = \sum_{i=1}^{n} Y_{ij} \times P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}), \qquad (2.28)$$

where $\overline{f_{jk}}$ stands for the expected number of examinees with ability X_k, $\overline{r_{jk}}$ is the expected number of examinees with ability X_k who will answer item j correctly. So the sum of $\overline{f_{jk}}$ is equal to the total number of examinees.

Table 2.1 The expected frequencies among examinees with ability X_k for item *j*

Item *j*	$W_i = 1$	$W_i = 0$	Marginal of W_i
$Y_i = 1$	$\overline{r_{jk}}^{(W)}$	$\overline{r_{jk}} - \overline{r_{jk}}^{(W)}$	$\overline{r_{jk}}$
$Y_i = 0$	$\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}$	$\overline{f_{jk}} - \overline{f_{jk}}^{(W)} - \overline{r_{jk}} + \overline{r_{jk}}^{(W)}$	$\overline{f_{jk}} - \overline{r_{jk}}$
Marginal of Y_i	$\overline{f_{jk}}^{(W)}$	$\overline{f_{jk}} - \overline{f_{jk}}^{(W)}$	$\overline{f_{jk}}$

Similarly, according to table 2.1, the newly defined artificial data are:

$$\begin{aligned} \overline{f_{jk}}^{(W)} &= \sum_{i=1}^{n} E(W_i \mid Y_i, X_k, \psi) P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) \\ \overline{r_{jk}}^{(W)} &= \sum_{i=1}^{n} Y_{ij} \times E(W_i \mid Y_i, X_k, \psi) \times P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}), \end{aligned} \tag{2.29}$$

where $\overline{f_{jk}}^{(W)}$ is the expected number of examinees with ability X_k who know the answer, $\overline{r_{jk}}^{(W)}$ is the expected number of examinees with ability X_k who know the answer and response correctly. Thus, $\overline{f_{jk}} - \overline{f_{jk}}^{(W)}$ stands for the expected number of examinees with ability X_k who do not know the answer, and $\overline{r_{jk}} - \overline{r_{jk}}^{(W)}$ is the expected number of examinees with ability X_k who do not know the answer but respond correctly. These definitions will be interpreted again after the following maximization steps of ς_j and ς_j.

Maximization step for upper asymptote. Specifying the ψ_j to be ς_j, the Bayesian formulation becomes

$$\begin{aligned} \lambda_{\varsigma_i} &= \frac{\partial \ln E(L)}{\partial \varsigma_j} + \frac{\partial \ln g(\varsigma_j \mid \eta)}{\partial \varsigma_j} \\ &= \sum_{i=1}^{N} \sum_{k=1}^{q} \left(\frac{\left(1 - Y_{ij}\right) E(W_{ij})}{\varsigma_j} - \frac{Y_{ij} E(W_{ij})}{(1 - \varsigma_j)} \right) P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) + \frac{\alpha_j^{\varsigma} - 1}{\varsigma_j} - \frac{\beta_j^{\varsigma} - 1}{1 - \varsigma_j} \end{aligned}$$

$$= \frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right)}{\varsigma_j} - \frac{\sum_{k=1}^{q}\overline{r_{jk}}^{(W)}}{(1-\varsigma_j)} + \frac{\alpha_j^{\varsigma}-1}{\varsigma_j} - \frac{\beta_j^{\varsigma}-1}{1-\varsigma_j}. \tag{2.30}$$

When $\lambda_{\varsigma_i} = 0$, the estimation of ς_j is

$$\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\varsigma} - 1}{\varsigma_j} = \frac{\beta_j^{\varsigma} - 1 + \sum_{k=1}^{q}\overline{r_{jk}}^{(W)}}{(1-\varsigma_j)}$$

$$\frac{1-\varsigma_j}{\varsigma_j} = \frac{\beta_j^{\varsigma} - 1 + \sum_{k=1}^{q}\overline{r_{jk}}^{(W)}}{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\varsigma} - 1}$$

$$\frac{1}{\varsigma_j} = \frac{\beta_j^{\varsigma} - 1 + \sum_{k=1}^{q}\overline{r_{jk}}^{(W)} + \sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\varsigma} - 1}{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\varsigma} - 1}$$

$$\varsigma_j = \frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\varsigma} - 1}{\sum_{k=1}^{q}\overline{f_{jk}}^{(W)} + \alpha_j^{\varsigma} - 1 + \beta_j^{\varsigma} - 1}. \tag{2.31}$$

Taking out the priors, the above shows an intuitive interpretation for the “slipping” parameter. Namely, the proportion of examinees who know the answer but respond wrong among all examinees who know the answer. Once the artificial data are obtained, the estimated upper asymptote parameter can be calculated.

Recall that

$$E(W_{ij} \mid Y_{ij}, \theta_i, \psi_j) = Y_{ij}\left[\frac{(1-\varsigma_j)P_j^*(\theta_i)}{P(\theta_i)}\right] + (1-Y_{ij})\left[\frac{\varsigma_j P_j^*(\theta_i)}{1-P(\theta_i)}\right], \tag{2.25}$$

and $Y_{ij} \sim Bernoulli\left(P_j\left(\theta_i\right)\right)$, so

$$
\begin{aligned}
&E\left[E(W_{ij} \mid Y_{ij}, \theta_i, \psi_j) \mid Y_{ij} = 1\right] \\
&= E(Y_{ij})\left[\frac{(1-\varsigma_j)P_j^*(\theta_i)}{P_j(\theta_i)}\right] + (1 - E(Y_{ij}))\left[\frac{\varsigma_j P_j^*(\theta_i)}{1 - P_j(\theta_i)}\right] \\
&= P_j(\theta_i)\left[\frac{(1-\varsigma_j)P_j^*(\theta_i)}{P_j(\theta_i)}\right] + (1 - P_j(\theta_i))\left[\frac{\varsigma_j P_j^*(\theta_i)}{1 - P_j(\theta_i)}\right] \\
&= (1-\varsigma_j)P_j^*(\theta_i) + \varsigma_j P_j^*(\theta_i) \\
&= P_j^*(\theta_i) \\
&= P(W_{ij} = 1).
\end{aligned} \tag{2.32}
$$

So

$$
\begin{aligned}
&E\left[\overline{f_{jk}} \mid Y_{ij} = 1\right] = E\left[\sum_{i=1}^{n} P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) \mid Y_{ij} = 1\right] = \overline{f_{jk}} \\
&E\left[\overline{f_{jk}}^{(W)} \mid Y_{ij} = 1\right] = E\left[\sum_{i=1}^{n} E(W_{ij} \mid Y_i, X_k, \psi) P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) \mid Y_{ij} = 1\right] = P_j^*(\theta_i)\,\overline{f_{jk}} \\
&E\left[\overline{r_{jk}} \mid Y_{ij} = 1\right] = E\left[\sum_{i=1}^{n} Y_{ij} P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) \mid Y_{ij} = 1\right] = P_j(\theta_i)\,\overline{f_{jk}} \\
&E\left[\overline{r_{jk}}^{(W)} \mid Y_{ij} = 1\right] = E\left[\sum_{i=1}^{n} Y_{ij} E(W_{ij} \mid Y_i, X_k, \psi) P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) \mid Y_{ij} = 1\right] = (1-\varsigma_j)P_j^*(\theta_i)\,\overline{f_{jk}}.
\end{aligned} \tag{2.33}
$$

The second derivative, which is necessary for estimating the standard error, is

$$
\begin{aligned}
\lambda_{\varsigma_i \varsigma_i} &= E\left[\frac{\partial^2 \ln E(L)}{\partial^2 \varsigma_j^2} \mid Y_{ij} = 1\right] + \frac{\partial^2 \ln g(\varsigma_j \mid \eta)}{\partial^2 \varsigma_j^2} \\
&= E\left[\frac{\partial}{\partial \varsigma_j}\left(\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right)}{\varsigma_j} - \frac{\sum_{k=1}^{q}\overline{r_{jk}}^{(W)}}{(1-\varsigma_j)}\right) \mid Y_{ij} = 1\right] - \frac{\alpha_j^{\varsigma} - 1}{\varsigma_j^2} - \frac{\beta_j^{\varsigma} - 1}{(1-\varsigma_j)^2}
\end{aligned}
$$

$$= E\left[-\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right)}{\varsigma_j^2} - \frac{\sum_{k=1}^{q}\overline{r_{jk}}^{(W)}}{(1-\varsigma_j)^2} \mid Y_{ij} = 1 \right] - \frac{\alpha_j^{\varsigma}-1}{\varsigma_j^2} - \frac{\beta_j^{\varsigma}-1}{(1-\varsigma_j)^2}$$

$$= -\frac{\sum_{k=1}^{q}\left(P_j^*(\theta_i)\overline{f_{jk}} - (1-\varsigma_j)P_j^*(\theta_i)\overline{f_{jk}}\right)}{\varsigma_j^2} - \frac{\sum_{k=1}^{q}(1-\varsigma_j)P_j^*(\theta_i)\overline{f_{jk}}}{(1-\varsigma_j)^2} - \frac{\alpha_j^{\varsigma}-1}{\varsigma_j^2} - \frac{\beta_j^{\varsigma}-1}{(1-\varsigma_j)^2}$$

$$= -\frac{\sum_{k=1}^{q}P_j^*(\theta_i)\overline{f_{jk}}}{\varsigma_j} - \frac{\sum_{k=1}^{q}P_j^*(\theta_i)\overline{f_{jk}}}{(1-\varsigma_j)} - \frac{\alpha_j^{\varsigma}-1}{\varsigma_j^2} - \frac{\beta_j^{\varsigma}-1}{(1-\varsigma_j)^2}$$

$$= -\frac{\sum_{k=1}^{q}P_j^*(\theta_i)\overline{f_{jk}}}{\varsigma_j(1-\varsigma_j)} - \frac{\alpha_j^{\varsigma}-1}{\varsigma_j^2} - \frac{\beta_j^{\varsigma}-1}{(1-\varsigma_j)^2}. \tag{2.34}$$

Maximization step for lower asymptote. By specifying the ψ_j to be γ_j, then the Bayesian formulation becomes

$$\lambda_{\gamma_j} = \frac{\partial \ln E(L)}{\partial \gamma_j} + \frac{\partial \ln g(\gamma_j \mid \eta)}{\partial \gamma_j}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{q}\left(\frac{Y_{ij}\left(1-E(W_{ij})\right)}{\gamma_j} - \frac{\left(1-Y_{ij}\right)\left(1-E(W_{ij})\right)}{\left(1-\gamma_j\right)}\right)P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) + \frac{\alpha_j^{\gamma}-1}{\gamma_j} - \frac{\beta_j^{\gamma}-1}{1-\gamma_j}$$

$$= \frac{\sum_{k=1}^{q}\left(\overline{r_{jk}} - \overline{r_{jk}}^{(W)}\right)}{\gamma_j} - \frac{\sum_{k=1}^{q}\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)} - \overline{r_{jk}} + \overline{r_{jk}}^{(W)}\right)}{(1-\gamma_j)} + \frac{\alpha_j^{\gamma}-1}{\gamma_j} - \frac{\beta_j^{\gamma}-1}{1-\gamma_j}.$$

(2.35)

When $\lambda_{\gamma_j} = 0$, the estimation of γ_j is

$$\frac{\sum_{k=1}^{q}\left(\overline{r_{jk}} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\gamma} - 1}{\gamma_j} = \frac{\sum_{k=1}^{q}\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)} - \overline{r_{jk}} + \overline{r_{jk}}^{(W)}\right) + \beta_j^{\gamma} - 1}{(1-\gamma_j)}$$

$$\frac{1-\gamma}{\gamma_j}=\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}-\overline{f_{jk}}^{(W)}-\overline{r_{jk}}+\overline{r_{jk}}^{(W)}\right)+\beta_j^{\gamma}-1}{\sum_{k=1}^{q}\left(\overline{r_{jk}}-\overline{r_{jk}}^{(W)}\right)+\alpha_j^{\gamma}-1}$$

$$\frac{1}{\gamma_j}=\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}-\overline{f_{jk}}^{(W)}\right)+\alpha_j^{\gamma}-1+\beta_j^{\gamma}-1}{\sum_{k=1}^{q}\left(\overline{r_{jk}}-\overline{r_{jk}}^{(W)}\right)+\alpha_j^{\gamma}-1}$$

$$\gamma_j=\frac{\sum_{k=1}^{q}\left(\overline{r_{jk}}-\overline{r_{jk}}^{(W)}\right)+\alpha_j^{\gamma}-1}{\sum_{k=1}^{q}\left(\overline{f_{jk}}-\overline{f_{jk}}^{(W)}\right)+\alpha_j^{\gamma}-1+\beta_j^{\gamma}-1}. \tag{2.36}$$

Ignoring the priors, $\overline{r_{jk}}-\overline{r_{jk}}^{(W)}/\overline{f_{jk}}-\overline{f_{jk}}^{(W)}$ is the proportion of examinees who do not know the answer but respond correctly among all examinees who do not know the answer. This interpretation matches the meaning of "guessing."

The expectation of the second derivative is:

$$\lambda_{\gamma_i\gamma_i}=E\left[\frac{\partial^2\ln E(L)}{\partial^2\gamma_j^2}\,|\,Y_{ij}=1\right]+\frac{\partial^2\ln g(\gamma_j\,|\,\eta)}{\partial^2\gamma_j^2}$$

$$=E\left[\frac{\partial}{\partial\gamma_j}\left(\frac{\sum_{k=1}^{q}\left(\overline{r_{jk}}-\overline{r_{jk}}^{(W)}\right)}{\gamma_j}-\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}-\overline{f_{jk}}^{(W)}-\overline{r_{jk}}+\overline{r_{jk}}^{(W)}\right)}{(1-\gamma_j)}\right)|\,Y_{ij}=1\right]-\frac{\alpha_j^{\gamma}-1}{\gamma_j^2}-\frac{\beta_j^{\gamma}-1}{(1-\gamma_j)^2}$$

$$=E\left[-\frac{\sum_{k=1}^{q}\left(\overline{r_{jk}}-\overline{r_{jk}}^{(W)}\right)}{\gamma_j^2}-\frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}-\overline{f_{jk}}^{(W)}-\overline{r_{jk}}+\overline{r_{jk}}^{(W)}\right)}{(1-\gamma_j)^2}\,|\,Y_{ij}=1\right]-\frac{\alpha_j^{\gamma}-1}{\gamma_j^2}-\frac{\beta_j^{\gamma}-1}{(1-\gamma_j)^2}$$

$$=-\frac{\sum_{k=1}^{q}\left(1-P_j^*(\theta_i)\right)\overline{f_{jk}}}{\gamma_j}-\frac{\sum_{k=1}^{q}\left(1-P_j^*(\theta_i)\right)\overline{f_{jk}}}{1-\gamma_j}-\frac{\alpha_j^{\gamma}-1}{\gamma_j^2}-\frac{\beta_j^{\gamma}-1}{(1-\gamma_j)^2}$$

$$= -\frac{\sum_{k=1}^{q}\left(1-P_j^*\left(\theta_i\right)\right)\overline{f_{jk}}}{\varsigma_j(1-\varsigma_j)} - \frac{\alpha_j^{\varsigma}-1}{\varsigma_j^2} - \frac{\beta_j^{\varsigma}-1}{(1-\varsigma_j)^2}. \tag{2.37}$$

Maximization step for discrimination and difficulty parameters. The first derivatives for $\ln a_j$ and b_j are:

$$\lambda_{a_j} = \frac{\partial \ln E(L)}{\partial a_j} + \frac{\partial \ln g(a_j \mid \eta)}{\partial a_j}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{q}\left(\left(\frac{E(W_{ij})}{P_j^*\left(X_k\right)} - \frac{\left(1-E(W_{ij})\right)}{\left(1-P_j^*\left(X_k\right)\right)}\right)\frac{\partial P_j^*\left(X_k\right)}{\partial a_j}\right)P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) - \frac{\ln a_j - \mu_{\ln a_j}}{\sigma_{\ln a_j}^2}$$

$$= \left(\sum_{k=1}^{q}\frac{\overline{f_{jk}}^{(W)}}{P_j^*\left(X_k\right)}\frac{\partial P_j^*\left(X_k\right)}{\partial a_j} - \sum_{k=1}^{q}\frac{\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)}\right)}{1-P_j^*\left(X_k\right)}\frac{\partial P_j^*\left(X_k\right)}{\partial a_j}\right) - \frac{\ln a_j - \mu_{\ln a_j}}{\sigma_{\ln a_j}^2}$$

$$= \left(\sum_{k=1}^{q}\frac{\overline{f_{jk}}^{(W)}}{P_j^*\left(X_k\right)}De^{\ln a_j}\left(X_t - b_j^{(k)}\right)w_{jk} - \sum_{k=1}^{q}\frac{\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)}\right)}{1-P_j^*\left(X_k\right)}De^{\ln a_j}\left(X_t - b_j^{(k)}\right)w_{jk}\right) - \frac{\ln a_j - \mu_{\ln a_j}}{\sigma_{\ln a_j}^2}$$

$$= De^{\ln a_j}\sum_{k=1}^{q}\left(X_t - b_j^{(k)}\right)\left[\overline{f_{jk}}^{(W)} - \overline{f_{jk}}P_j^*\left(X_k\right)\right] - \frac{\ln a_j - \mu_{\ln a_j}}{\sigma_{\ln a_j}^2}. \tag{2.38}$$

$$\lambda_{b_j} = \frac{\partial \ln E(L)}{\partial b_j} + \frac{\partial \ln g(b_j \mid \eta)}{\partial b_j}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{q}\left(\left(\frac{E(W_{ij})}{P_j^*\left(X_k\right)} - \frac{\left(1-E(W_{ij})\right)}{\left(1-P_j^*\left(X_k\right)\right)}\right)\frac{\partial P_j^*\left(X_k\right)}{\partial b_j}\right)P(X_k \mid Y_i, W_i, \psi, \boldsymbol{\tau}) - \frac{b_j - \mu_{b_j}}{\sigma_{b_j}^2}$$

$$= \left(\sum_{k=1}^{q}\frac{\overline{f_{jk}}^{(W)}}{P_j^*\left(X_k\right)}\frac{\partial P_j^*\left(\theta_i\right)}{\partial b_j} - \sum_{k=1}^{q}\frac{\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)}\right)}{1-P_j^*\left(X_k\right)}\frac{\partial P_j^*\left(X_k\right)}{\partial b_j}\right) - \frac{b_j - \mu_{b_j}}{\sigma_{b_j}^2}$$

$$= \left(\sum_{k=1}^{q}\frac{\overline{f_{jk}}^{(W)}}{P_j^*\left(X_k\right)}\left(-De^{\ln a_j}w_{jk}\right) - \sum_{k=1}^{q}\frac{\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)}\right)}{1-P_j^*\left(X_k\right)}\left(-De^{\ln a_j}w_{jk}\right)\right) - \frac{b_j - \mu_{b_j}}{\sigma_{b_j}^2}$$

$$= -De^{\ln a_j} \sum_{k=1}^{q} \left(X_t - b_j^{(k)} \right) \left[\overline{f_{jk}}^{(W)} - \overline{f_{jk}} P_j^* \left(X_k \right) \right] - \frac{b_j - \mu_{b_j}}{\sigma_{b_j}^2}. \tag{2.39}$$

where $w_{jk} = P_j^* \left(X_k \right) \times \left(1 - P_j^* \left(X_k \right) \right)$

The corresponding expectation for the second derivatives are:

$$\begin{aligned}
\lambda_{a_j a_j} &= -D^2 e^{2 \ln a_j} \sum_{k=1}^{q} \left[\left(X_k - b_j^{(k)} \right)^2 w_{jk} \overline{f_{jk}}^{(W)} \right] - \frac{1}{\sigma_{\ln a_j}^2} \\
\lambda_{b_j b_j} &= -D^2 e^{2 \ln a_j} \sum_{k=1}^{q} \left[w_{jk} \overline{f_{jk}}^{(W)} \right] - \frac{1}{\sigma_{b_j}^2} \\
\lambda_{a_j b_j} &= D^2 e^{2 \ln a_j} \sum_{k=1}^{q} \left[\left(X_k - b_j^{(k)} \right) w_{jk} \overline{f_{jk}}^{(W)} \right].
\end{aligned} \tag{2.40}$$

Using Fisher-scoring iteration, the estimates for $\ln a_j$ and b_j can be obtained:

$$\begin{bmatrix} \ln a_j^{(t+1)} \\ b_j^{(t+1)} \end{bmatrix} = \begin{bmatrix} \ln a_j^{(t)} \\ b_j^{(t)} \end{bmatrix} - \begin{bmatrix} \lambda_{aa} & \lambda_{ab} \\ \lambda_{ab} & \lambda_{bb} \end{bmatrix}^{-1} \begin{bmatrix} \lambda_a \\ \lambda_b \end{bmatrix}. \tag{2.41}$$

Compare (2.41) to the 4-by-4 matrix in traditional 4PL, which includes all of the parameters,

$$\begin{bmatrix} \ln a_j^{(t+1)} \\ b_j^{(t+1)} \\ \gamma_j^{(t+1)} \\ \varsigma_j^{(t+1)} \end{bmatrix} = \begin{bmatrix} \ln a_j^{(t)} \\ b_j^{(t)} \\ \gamma_j^{(t+1)} \\ \varsigma_j^{(t+1)} \end{bmatrix} - \begin{bmatrix} \lambda_{aa} & \lambda_{ba} & \lambda_{\gamma a} & \lambda_{\varsigma a} \\ \lambda_{ab} & \lambda_{bb} & \lambda_{\gamma b} & \lambda_{\varsigma b} \\ \lambda_{a\gamma} & \lambda_{b\gamma} & \lambda_{\gamma\gamma} & \lambda_{\varsigma\gamma} \\ \lambda_{a\varsigma} & \lambda_{b\varsigma} & \lambda_{\gamma\varsigma} & \lambda_{\varsigma\varsigma} \end{bmatrix}^{-1} \begin{bmatrix} \lambda_a \\ \lambda_b \\ \lambda_\gamma \\ \lambda_\varsigma \end{bmatrix}, \tag{2.42}$$

the proposed method separates the iteration of γ_j and ς_j from the discrimination and difficulty parameters. That might be the reason why the BEMMM algorithm is fast and stable.

CHAPTER 3: SIMULATION STUDIES

Two simulation studies were conducted. The first compares the BEMMM with EMMM to determine the power of the Bayesian method. The second examines the accuracy of BEMMM, BME, and MCMC for 4PNO. The estimates for EMMM and BEMMM were programmed in MATLAB while the BME used the *mirt* package (Chalmers, 2012). To compare the BEMMM and the MCMC, the following parameterization which is same as in Culpepper's paper (2016) is used:

$$\alpha_j = a_j,\ \beta_j = a_j b_j. \tag{3.1}$$

where α_j is item threshold, β_j is the intercept.

3.1 Simulation study 1: BEMMM VS EMMM

To verify that the Bayesian method does reduce the number of implausible estimates in EMMM, both the EMMM and Bayesian EMMM method were used in this study.

Item parameter generation: The starting values of the item parameters are generated as follows (See Culpepper, 2016): $\alpha_j \sim N(2,\ 0.5)\mathrm{I}(\alpha_j > 0)$, $\beta_j \sim N(0,0.5)$, $\gamma_j \sim \mathrm{Beta}(2,8)$, and $\varsigma_j \mid \gamma_j \sim Beta(2,8)\mathrm{I}(\varsigma_j < 1-\gamma_j)$.

Simulation design: Two sample sizes of examinees (2500 and 5000) were generated from the standard normal distribution. Twenty items were generated. It should be noticed that "α_j has a normal distribution truncated at zero" (Culpepper, 2016), the expectation of α_j is $2/\sqrt{\pi}$ instead of 0. Thus, $\alpha_j \sim N(0,2)I(a \ge 0),\ \beta_j \sim N(0,2)$. The priors for γ_j and ς_j are beta (1, 10). Overall, 50 replications for each condition in the fully crossed 2 (EMMM vs. BEMMM) × 2 (2500 vs. 5000) design were generated.

Evaluation criteria: Bias and root mean squared error (RMSE) for item parameter recovery; that are:

$$bias = \frac{\sum_{s=1}^{S=50}(\hat{\psi}_{is} - \psi_i)}{S}, \quad RMSE = \sqrt{\frac{\sum_{s=1}^{S=50}(\hat{\psi}_{is} - \psi_i)^2}{S}}. \tag{3.2}$$

Item parameter recovery: The bias and RMSE results for cognitive testing with a sample size of 2500 are presented in Figure 3.1 and Table 3.1. The results for the 5000 examinees condition shows a similar pattern and are summarized in Appendix A.

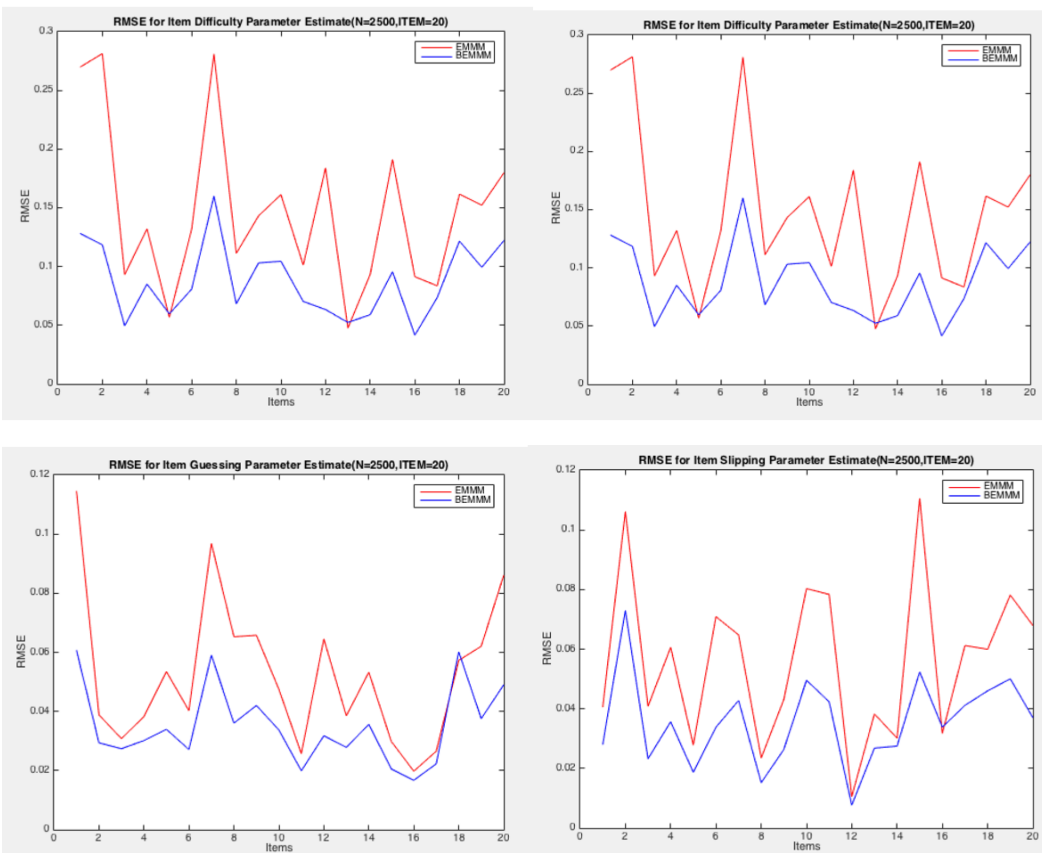

Figure 3.1 RMSEs for 2500 Examinees and 20 Items in Simulation Study 1

First, the values of bias and RMSEs for EMMM are acceptable, these results support the contention that EMMM is a feasible model. Next, it is obvious that the results for BEMMM are generally smaller than EMMM, especially for "blow-up" items, thus the conclusion is that the EMMM can be considerably improved by adding appropriate priors for item parameters.

This does not mean that EMMM is useless, because the accuracy of results is profoundly influenced by priors when using the Bayesian method. If researchers are unable to obtain appropriate priors, EMMM may be better than BEMMM.

Table 3.1 Bias for 2500 Examinees and 20 Items in Simulation Study 1

α		β		γ		ς	
EMMM	BEMMM	EMMM	BEMMM	EMMM	BEMMM	EMMM	BEMMM
-0.19	-0.10	-0.05	-0.05	-0.06	-0.02	-0.02	0.00
-0.07	0.21	0.10	-0.01	-0.01	0.00	-0.04	0.02
-0.28	-0.20	-0.07	0.00	-0.02	-0.01	-0.01	0.00
-0.24	-0.06	-0.03	0.00	-0.02	-0.01	-0.02	0.00
0.15	0.00	0.01	-0.01	-0.02	-0.01	0.00	0.00
-0.33	-0.17	0.00	0.03	-0.02	0.00	-0.03	-0.01
0.06	0.12	0.02	0.03	-0.01	0.01	-0.02	0.01
0.27	0.08	0.05	0.00	-0.01	0.00	0.00	0.01
0.00	-0.37	0.01	-0.03	-0.02	-0.02	-0.01	-0.01
-0.24	0.06	0.05	0.02	-0.02	0.00	-0.04	0.00
-0.52	-0.06	0.00	0.02	-0.02	0.00	-0.05	0.01
-0.01	-0.18	0.15	-0.01	-0.01	0.00	0.00	0.00
0.22	0.03	0.03	0.01	0.00	0.00	0.00	0.00
-0.07	0.10	0.03	0.00	-0.02	0.01	0.00	0.01
-0.46	-0.04	0.10	0.04	-0.02	0.00	-0.08	0.00
-0.48	0.00	-0.08	0.01	-0.01	0.00	-0.02	0.02
-0.52	-0.06	-0.03	0.03	-0.02	0.00	-0.05	0.01
-0.03	0.21	0.04	0.05	0.00	0.05	-0.02	0.02
0.02	0.18	0.01	0.01	-0.02	0.01	-0.02	0.01
0.10	0.07	-0.03	-0.02	-0.03	0.00	-0.01	0.01

3.2 Simulation study 2: BEMMM VS BME VS MCMC

This study seeks to compare the accuracy of Bayesian EMMM, BME and MCMC under the same conditions.

Item parameter generation: The starting values of the item parameters are generated as $\alpha_j \sim N(2,\ 0.5)\mathrm{I}(\alpha_j > 0)$, $\beta_j \sim N(0,0.5)$, γ_j ~Beta(2,8) and $\varsigma_j \mid \gamma_j \sim$ $Beta(2,8)\mathrm{I}(\varsigma_j < 1-\gamma_j)$, which are also used by Culpepper (2016).

Simulation design: Sample sizes are 2500 and 5000. $\theta_j \sim N(0,1)$, and the number of items are 20. The priors are $\alpha_j \sim N(0,2)I(a \geq 0),\ \beta_j \sim N(0,2)$. Overall, 50 replications were generated for each condition in the fully crossed 3 (BEMMM vs MCMC vs BME) × 2 (2500 vs 5000) design.

Evaluation criteria: Bias and root mean squared error (RMSE) for item parameter recovery.

Item parameter recovery: The bias and RMSEs results for cognitive testing with the sample size 2500 are presented in Figure 3.2 and Table 3.2. The results for the 5000 examinees condition shows a similar pattern to the 2500 examinees condition, and are summarized in Appendix B.

Figure 3.2 shows that the RMSE values for BEMMM and MCMC are close, which supports that the BEMMM is as accurate as MCMC.

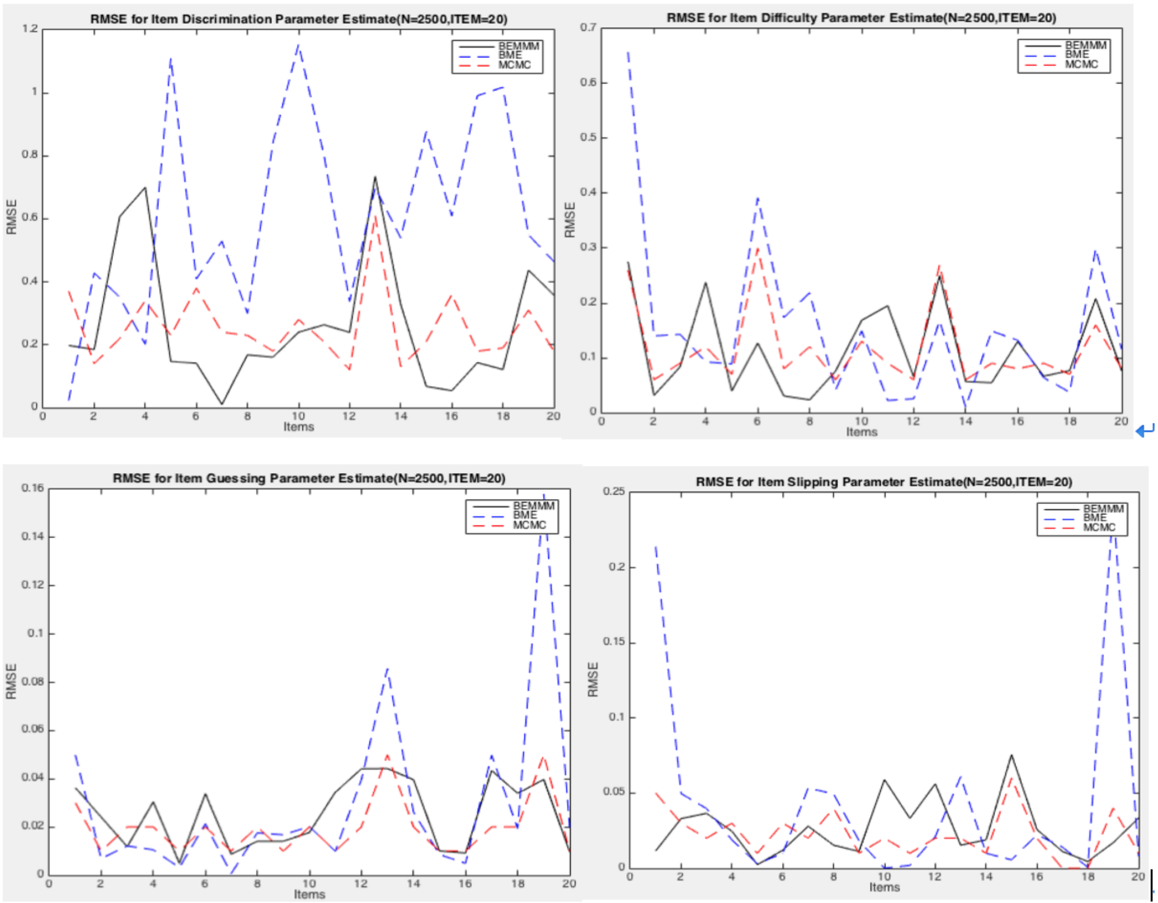

Figure 3.2 RMSEs for 2500 Examinees and 20 Items in Simulation Study 2

Table 3.1 presented the biases for Bayesian EMMM, Bayesian Modal Estimation, and MCMC. The values of BEMMM and MCMC are similar and small, while the BME has considerably larger biases. In these models, the estimates of lower and upper asymptotes parameters perform better than that of discrimination and difficulty parameters. When comparing the RMSE and bias of the 2500 condition with the results of 5000 examinees (in Appendix B), the 5000 condition shows smaller RMSEs and biases due to the increase in the sample size.

To sum up the results of RMSE and bias, it is sufficient to say that BEMMM is as accurate as MCMC, and BME is less accurate than the other.

Table 3.2 Bias for 2500 Examinees and 20 Items in Simulation Study 2

α			β			γ			ς		
MCMC	BEMMM	BME	MCMC	BEMMM	BME	MCMC	BEMMM	BME	MCMC	BEMMM	BME
0.19	0.02	-0.02	0.05	0.02	0.66	-0.01	-0.01	-0.05	0.00	0.01	0.21
0.07	0.04	0.43	0.01	-0.03	0.14	0.01	0.02	-0.01	0.01	0.03	0.05
-0.02	0.34	0.35	0.01	0.04	0.14	0.00	0.00	-0.01	0.00	0.04	0.04
-0.07	0.48	0.20	0.04	0.07	0.09	0.00	0.02	-0.01	0.01	0.02	0.02
-0.02	-0.15	1.11	0.02	-0.02	0.09	0.00	0.00	0.00	0.00	0.00	0.00
0.21	-0.14	0.41	0.01	0.07	-0.39	0.01	0.02	-0.02	-0.01	-0.01	-0.01
-0.06	-0.01	0.53	0.02	0.01	0.17	0.00	0.00	0.00	-0.01	0.03	0.05
0.16	-0.17	0.30	0.03	0.00	0.22	0.01	0.00	-0.02	0.03	0.01	0.05
0.04	-0.14	0.84	0.02	-0.01	0.04	0.00	0.01	0.02	0.00	0.01	-0.02
0.19	0.04	1.15	0.05	0.02	0.15	0.00	0.01	0.02	0.02	0.06	0.00
0.05	0.07	0.80	0.01	-0.17	0.02	0.01	0.03	-0.01	0.00	0.02	0.00
0.05	0.23	0.34	0.00	-0.03	-0.03	0.00	0.04	-0.04	0.01	0.05	0.02
-0.45	-0.58	-0.70	0.09	0.25	-0.17	-0.04	-0.02	-0.09	-0.01	-0.02	0.06
0.06	0.18	0.54	-0.01	-0.06	0.01	0.01	0.03	-0.03	0.01	0.02	0.01
0.15	0.07	0.88	0.01	-0.03	0.15	0.00	0.01	0.01	0.05	0.08	-0.01
-0.15	0.01	0.61	0.01	0.01	0.13	0.00	0.01	0.00	0.00	0.02	0.02
0.03	-0.11	0.99	0.00	-0.03	-0.06	0.01	0.04	-0.05	0.00	0.01	-0.01
0.07	-0.12	1.02	-0.01	0.08	-0.04	0.01	0.03	-0.02	0.00	0.00	0.00
-0.02	0.41	-0.55	-0.02	0.00	0.30	-0.03	0.03	-0.16	-0.01	0.01	0.24
0.00	0.27	0.46	0.02	0.05	0.12	0.00	0.01	-0.02	0.00	0.03	0.01

CHAPTER 4: EXAMPLE USING BULLY DATA

This section reports results of an application of the 4PLM to bullying items collected as part of the 2005-2006 Health Behavior in School-Aged Children (Iannotti, 2005) study, which has been used before (i e., Culpepper, 2016). This study seeks to confirm the practicability of BEMMM and compare the accuracy and speed of BEMMM with MCMC.

The response matrix consists of 7491 adolescents, ages 11, 13 and 15, and ten bullying items. The original polytomous items were dichotomized as:

$$W_{ij} = \begin{cases} 1 & \text{Student i bullied another student as asked in item j} \\ 0 & \text{Student i did not bully another student as asked in item j} \end{cases}. \qquad (4.1)$$

Some students who did not bully may not report bullying behaviors so the γ_j might be close to 0. The priors for both α_j and β_j are $N(0,2)$, γ_j ~Beta(2,8), and $\varsigma_j \mid \gamma_j \sim Beta(2,8)\mathrm{I}(\varsigma_j < 1-\gamma_j)$.

Table 4.1 contains the point estimates and standard deviations for BEMMM and MCMC. The estimates of item parameters are very similar between the two methods, and the standard deviations for the item thresholds and intercepts for BEMMM are smaller than for MCMC. The values of γ_j are close to 0, which confirmed the meaning of γ_j as mentioned above. The point estimates of the upper asymptote for item 1 and 2 are quite larger than 0, which means students bullied another student but answered no to these items. The results support that 4PLM is useful when treating socially desirable responding.

Compared with the time Culpepper (2016) reported "The Gibbs sampler required approximately 47 min to complete 100,000 iterations with N = 7491 using a 2.4GHz processor and 6GB of RAM," the time for the BEMMM is only about 60 seconds. BEMMM might be useful in practice as a time-saving method.

Table 4.1 Estimated Item Parameters for bullying data from the HBSC study

Item	α				β			
	MCMC(SE)		BEMMM(SE)		MCMC(SE)		BEMMM(SE)	
1	4.44	(0.44)	3.93	(0.41)	0.67	(0.10)	0.63	(0.03)
2	3.54	(0.33)	3.15	(0.00)	0.74	(0.09)	0.77	0.00
3	1.21	(0.07)	1.18	(0.03)	1.10	(0.06)	1.14	(0.02)
4	1.41	(0.06)	1.40	(0.02)	1.83	(0.06)	1.94	(0.03)
5	1.63	(0.09)	1.48	(0.02)	2.32	(0.11)	2.27	(0.03)
6	1.93	(0.10)	1.79	(0.03)	2.93	(0.13)	2.94	(0.04)
7	2.44	(0.15)	2.29	(0.03)	4.05	(0.22)	4.14	(0.05)
8	1.58	(0.08)	1.45	(0.02)	2.13	(0.09)	2.09	(0.03)
9	2.23	(0.15)	2.06	(0.03)	3.75	(0.22)	3.76	(0.05)
10	2.29	(0.16)	2.09	(0.03)	3.89	(0.24)	3.87	(0.05)
Item	γ				ς			
	MCMC(SE)		BEMMM(SE)		MCMC(SE)		BEMMM(SE)	
1	0.00	(0.00)	0.00	(0.00)	0.17	(0.01)	0.16	(0.01)
2	0.00	(0.00)	0.00	(0.00)	0.18	(0.01)	0.16	0.00
3	0.01	(0.01)	0.00	(0.00)	0.03	(0.02)	0.03	(0.01)
4	0.00	(0.00)	0.00	(0.00)	0.01	(0.01)	0.01	(0.01)
5	0.01	(0.00)	0.01	(0.00)	0.01	(0.01)	0.01	(0.01)
6	0.00	(0.00)	0.00	(0.00)	0.01	(0.01)	0.01	(0.01)
7	0.00	(0.00)	0.00	(0.00)	0.01	(0.01)	0.01	(0.01)
8	0.01	(0.00)	0.00	(0.00)	0.01	(0.01)	0.01	(0.01)
9	0.01	(0.00)	0.01	(0.00)	0.01	(0.01)	0.01	(0.01)
10	0.00	(0.00)	0.00	(0.00)	0.01	(0.01)	0.01	(0.01)

CHAPTER 5: DISCUSSION

Following the renewed interest in 4PL, a reformulated 4PL algorithm was proposed. This section summarizes the study, comes to conclusions, discusses certain issues and possible directions for future research.

5.1 Conclusion

In section 2, the mathematical derivation of the new algorithm was presented. The method of adding a latent variable separates the estimation of lower and upper asymptotes parameters from the estimation of discrimination and difficulty parameters, and makes the algorithm fast and stable. The combination with the Bayesian method provides extra information through the priors, which solves the issue of implausible estimates.

Based on the results of simulation studies in Section 3 and the bully data in Section 4, the author can conclude that the Bayesian EMMM yields comparable estimates with MCMC and performs better than the BME with respect to accuracy. Furthermore, the speed of BEMMM is significantly faster than the MCMC (60 seconds vs. 47 minutes.)

Overall, BEMMM and MCMC are more accurate than BME. While the BEMMM is much faster than the MCMC, the BEMMM should be regarded as a suitable algorithm for 4PLM.

5.2 Discussion

There are two main interpretations of γ and ς parameters. One regards γ as the probability of producing a correct response by random guessing (Waller, 1974; Hambleton & Cook, 1977) and ς is the possibility of slipping (Culpepper, 2016). The other thinks of γ as the possibility of success by low-proficiency students and ς is the possibility of failure by high-proficiency students.

In section 2, the formulation of estimated γ and ς parameters was expressed using artificial data:

$$\varsigma_j = \frac{\sum_{k=1}^{q}\left(\overline{f_{jk}}^{(W)} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\varsigma} - 1}{\sum_{k=1}^{q}\overline{f_{jk}}^{(W)} + \alpha_j^{\varsigma} - 1 + \beta_j^{\varsigma} - 1}, \quad \gamma_j = \frac{\sum_{k=1}^{q}\left(\overline{r_{jk}} - \overline{r_{jk}}^{(W)}\right) + \alpha_j^{\gamma} - 1}{\sum_{k=1}^{q}\left(\overline{f_{jk}} - \overline{f_{jk}}^{(W)}\right) + \alpha_j^{\gamma} - 1 + \beta_j^{\gamma} - 1}. \quad (5.1)$$

According to the definition of artificial data, there are the third meaning of γ and ς from a mixture-modeling perspective: γ is the proportion of examinees who do not know the answer but respond correctly within the group of examinees who do not know the answer, ς is the proportion of examinees who know the answer but respond incorrectly among examinees know the answer. The interpretation is intuitive and meets with the meaning of “guessing” and “slipping.” For the bully data, γ is interpreted as the proportion of students who did not bully but answered that he bullied others among all students who did not bully, ς is the proportion of students who bullied others but answered no among all students who bullied others.

The simulation studies and the real data example may be oversampled so researchers should be cautious when applying the BEMMM in practice. At the least, it can be used to get the starting values for the MCMC in order to save time, and can be used to check with the BME. In practice, the best way is to use different algorithms to check upon each other and combine the advantages of every method.

5.3 Possible Directions for Future Research

First, Zheng, et al. (2017) proposed a similar latent mixture-modeling-based algorithm for the 3PLM which they called the EMM. Researchers could compare the model-fit of EMM and EMMM when applied to cognitive data, if the results of EMMM are better, 4PLM may be a more appropriate model in practice.

Second, by carrying on the latent mixture model perspective, researchers can introduce two latent variables to represent the lower and upper asymptotes parameters separately.

Third, as regards the accuracy of predictions, the estimated standard error (SE) is different for each different model or algorithm, so it is necessary to calculate the SE for BEMMM in order to judge the stability of estimation. Cai and Lee (2009) proposed the supplemented EM, which can be applied to 4PLM in future.

Finally, due to the increasing number of applications of the 4PLM in several areas of research, the BEMMM can be applied in such areas of practical testing such as computerized adaptive testing.

REFERENCES

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of educational statistics*, *17*(3), 251-269.

Barton, M. A., & Lord, F. M. (1981). AN UPPER ASYMPTOTE FOR THE THREE-PARAMETER LOGISTIC ITEM-RESPONSE MODEL. *ETS Research Report Series*, *1981*(1).

Béguin, A. A., & Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*(4), 541-561.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46(4), 443-459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model forn dichotomously scored items. *Psychometrika*, *35*(2), 179-197.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). MMPI-2: Manual for administration and scoring.

Cai, L., & Lee, T. (2009). Covariance structure model fit testing under missing data: An application of the supplemented EM algorithm. *Multivariate Behavioral Research*, *44*(2), 281-304.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1-29.

Cheng, Y., & Liu, C. (2015). The effect of upper and lower asymptotes of IRT models on computerized adaptive testing. *Applied Psychological Measurement*, *39*(7), 551-565.

Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *psychometrika*, *81*(4), 1142-1163.

Feuerstahler, L. M., & Waller, N. G. (2014). Estimation of the 4-parameter model with marginal maximum likelihood. *Multivariate behavioral research*, *49*(3), 285-285.

Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. Springer Science & Business Media.

Guo, S. (2017). *Bayesian Expectation-Maximization-Maximization: a latent-mixture-modeling-based Bayesian algorithm for the three-parameter logistic model.* (Master's thesis, University of Illinois at Urbana-Champaign). Retrieved from http://hdl.handle.net/2142/97545

Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of educational measurement*, *14*(2), 75-96.

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

Ho, R. G., & Yen, Y. C. (2005). Design and evaluation of an XML-based platform-independent computerized adaptive testing system. *IEEE Transactions on Education*, *48*(2), 230-237.

Iannotti, R. J. (2013). Health behavior in school-aged children (HBSC), 2009-2010. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research*.

Liao, W. W., Ho, R. G., Yen, Y. C., & Cheng, H. C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. *Social Behavior and Personality: an international journal*, *40*(10), 1679-1694.

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 509-525.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Magis, D. (2013). A note on the item information function of the four-parameter logistic model. *Applied Psychological Measurement*, *37*(4), 304-315.

McDonald, R. P. (1967). Nonlinear factor analysis (Psychometric Monographs, No. 15). Richmond, VA: Psychometric Corporation.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, *51*(2), 177-195.

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items?. *Psychological Methods*, *8*(2), 164.

Rulison, K. L., & Loken, E. (2009). I've fallen and I can't get up: can high-ability students recover from early mistakes in CAT?. *Applied Psychological Measurement*, *33*(2), 83-101.

Thomas, A., O'Hara, B., Ligges, U., & Sturtz, S. (2006). Making BUGS Open. R News 6: 12–17.

Waller, M. I. (1974). Removing the effects of random guessing from latent trait ability estimates. *ETS Research Report Series*, *1974*(1).

Waller, N. G., & Reise, S. P. (2010). Measuring psychopathology with non-standard IRT models: Fitting the four-parameter model to the MMPI. In *Measuring psychological constructs with model-based approaches*. American Psychological Association.

Waller, N. G., & Feuerstahler, L. (2017). Bayesian modal estimation of the four-parameter item response model in real, realistic, and idealized data sets. *Multivariate Behavioral Research*, *52*(3), 350-370.

Yen, Y. C., Ho, R. G., Laio, W. W., Chen, L. J., & Kuo, C. C. (2012). An empirical evaluation of the slip correction in the four parameter logistic models with computerized adaptive testing. *Applied Psychological Measurement*, *36*(2), 75-87.

Zhang, J. (2005). Estimating multidimensional item response models with mixed structure. ETS Research Report 05-22. Princeton, NJ: Educational Testing Service.

Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement*, *36*(5), 375-398.

Zheng, C., Meng, X., Guo, S., & Liu, Z. (2017). Expectation-Maximization-Maximization: A Feasible MLE Algorithm for the Three-Parameter Logistic Model Based on a Mixture Modeling Reformulation. *Frontiers in psychology*, *8*, 2302.

Zimowski, M. F., Muraki, E., Mislevy, R., & Bock, R. D. (1996). BILOG-MG. *Multiple group IRT analysis and test maintenance for binary items*.

APPENDIX A: DEFINITION OF SYMBOLS IN EMMM AND BEMMM

Y -- Responses

W -- Newly defined discrete augmented variable

θ -- Ability parameter

a -- Discrimination parameter

b -- Difficulty parameter

γ -- Lower asymptote parameter

ς -- Upper asymptote parameter

P^* -- Two-parameter logistic model

ψ -- Item parameter vector which contains four item parameters

τ -- Hyper-parameter, contains the parameters of examinee population ability distribution

η -- Mean and variance of the prior distributions for item parameters

$\alpha,\ \beta$ -- Parameters for Beta distribution

X_k -- Nodes on the ability scale

$\overline{f_{jk}}$ -- Expected number of examinees with ability X_k

$\overline{r_{jk}}$ -- Expected number of examinees with ability X_k who response correctly

$\overline{f_{jk}}^{(W)}$ -- Expected number of examinees with ability X_k who know the answer

$\overline{r_{jk}}^{(W)}$ -- Expected number of examinees with ability X_k who know the answer and response correctly

w -- $P^* \times \left(1 - P^*\right)$

APPENDIX B: FIGURES AND TABLES IN SIMULATION STUDIES

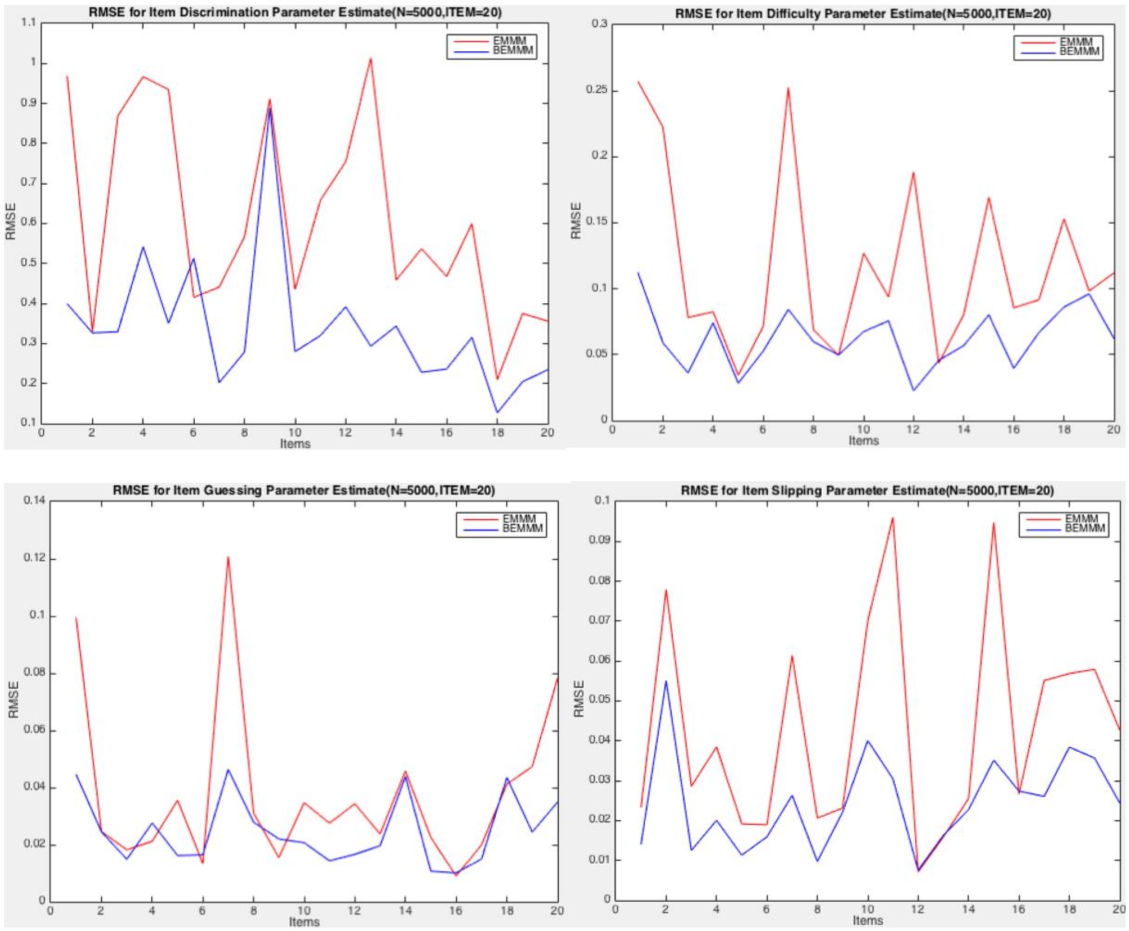

Figure B.1 RMSEs for 5000 Examinees and 20 Items in Simulation Study 1

Table B.1 Bias for 5000 Examinees and 20 Items in Simulation Study 1

α		β		γ		ς	
EMMM	BEMMM	EMMM	BEMMM	EMMM	BEMMM	EMMM	BEMMM
0.02	-0.11	0.00	-0.02	-0.04	-0.01	-0.01	0.00
-0.26	0.15	0.17	-0.01	-0.01	0.01	-0.06	0.04
-0.63	-0.18	-0.06	0.03	-0.01	0.01	-0.01	0.01
-0.11	0.01	-0.03	0.04	0.00	0.00	-0.01	0.01
0.01	-0.06	0.01	0.00	-0.01	0.00	0.00	0.01
-0.28	0.15	-0.06	0.03	-0.01	0.00	-0.01	0.00
-0.22	-0.06	-0.14	-0.01	-0.08	-0.01	-0.03	0.01
-0.51	-0.02	0.06	0.04	-0.02	0.02	-0.01	0.00
0.10	-0.79	0.05	-0.01	0.00	-0.02	-0.01	-0.02
-0.27	-0.09	0.02	0.02	-0.02	0.00	-0.03	0.00
-0.64	-0.19	0.07	0.06	-0.02	0.00	-0.09	-0.01
-0.04	-0.03	0.18	0.00	0.00	0.00	0.00	0.00
0.47	0.06	0.04	0.02	0.01	0.01	0.00	0.00
-0.09	0.13	0.06	0.02	0.00	0.03	0.00	0.01
-0.49	-0.11	0.07	0.03	-0.02	0.00	-0.07	0.01
-0.45	0.03	-0.08	0.02	-0.01	0.01	-0.02	0.02
-0.55	-0.18	-0.01	0.04	-0.01	0.00	-0.04	0.02
-0.18	0.07	0.00	0.04	-0.03	0.04	-0.03	0.02
-0.25	-0.06	0.06	0.05	-0.01	0.01	-0.03	0.00
-0.23	0.05	-0.04	-0.01	-0.04	0.00	-0.02	0.01

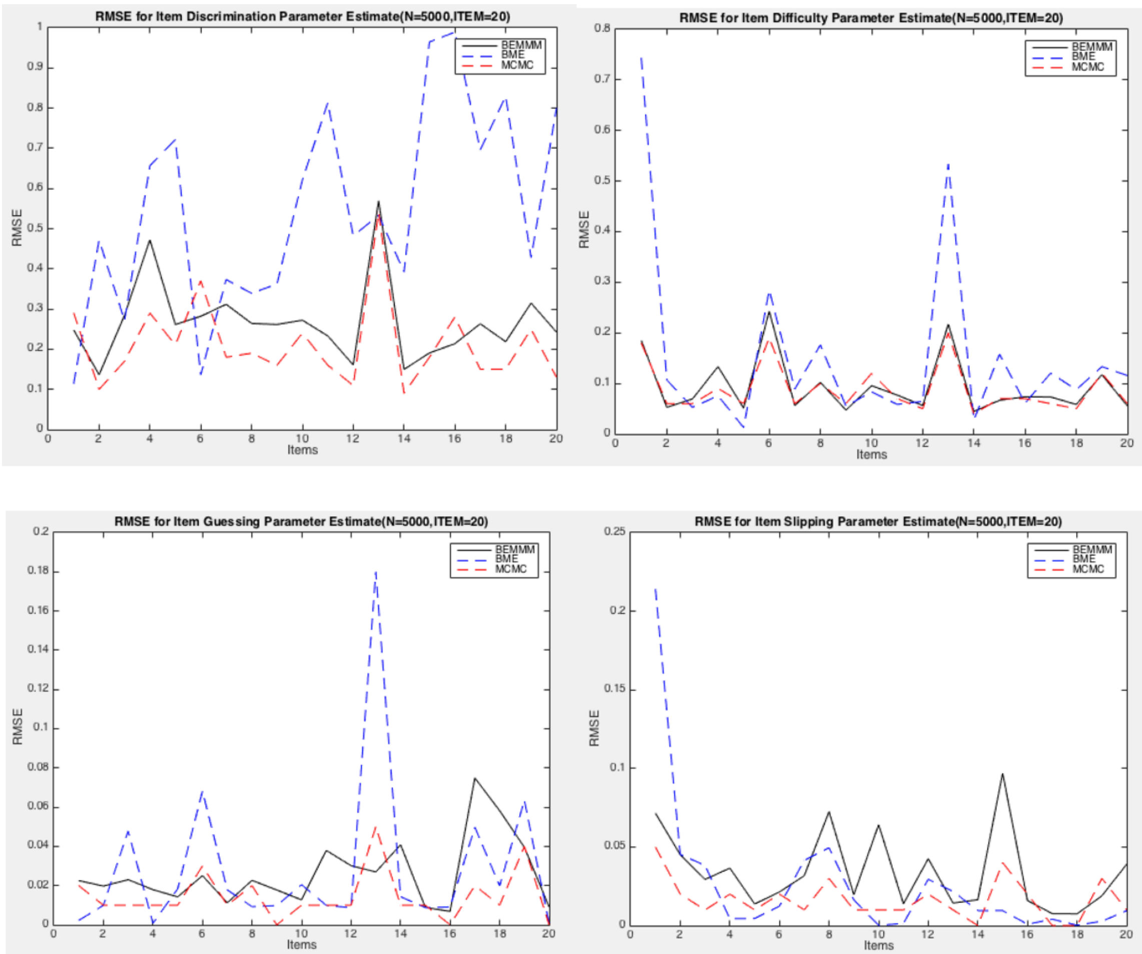

Figure B.2 RMSEs for 5000 Examinees and 20 Items in Simulation Study 2

Table B.2 Bias for 5000 Examinees and 20 Items in Simulation Study 2

α			β			γ			ς		
MCMC	BEMMM	BME	MCMC	BEMMM	BME	MCMC	BEMMM	BME	MCMC	BEMMM	BME
0.13	0.11	0.11	0.05	-0.05	0.74	-0.00	0.00	0.00	-0.01	0.05	0.21
0.03	0.05	0.47	0.04	0.01	0.11	0.00	0.02	-0.01	-0.00	0.04	0.05
-0.01	0.13	0.27	0.02	0.02	0.05	-0.00	0.02	-0.05	-0.00	0.03	0.04
-0.04	0.16	0.66	0.03	0.02	0.08	-0.00	0.01	0.00	-0.01	0.03	0.00
0.00	0.09	0.72	0.04	0.01	0.01	0.00	0.01	-0.02	0.00	0.01	0.00
0.18	-0.14	0.14	-0.00	0.14	-0.28	0.01	0.01	-0.07	-0.00	-0.01	0.01
-0.06	0.17	0.37	0.03	0.01	0.09	-0.00	0.01	-0.02	-0.00	0.03	0.04
0.11	0.16	0.34	0.04	0.02	0.18	0.01	0.02	-0.01	0.01	0.06	0.05
-0.04	0.13	0.36	0.04	0.00	0.06	0.00	0.02	-0.01	-0.00	0.02	0.02
0.18	0.16	0.62	0.08	0.02	0.08	0.00	0.01	-0.02	0.01	0.06	0.00
0.04	0.05	0.81	0.02	0.02	-0.06	0.00	0.04	-0.01	0.00	0.01	0.00
0.03	0.08	0.48	0.02	0.00	0.06	0.00	0.03	-0.01	0.00	0.04	0.03
-0.28	-0.37	-0.53	0.03	0.07	-0.53	-0.03	0.00	-0.18	-0.00	-0.01	0.02
0.01	0.05	0.39	0.01	0.01	-0.03	0.00	0.04	-0.01	0.00	0.02	0.01
0.14	0.11	0.96	0.03	-0.01	0.16	0.00	0.01	0.01	0.04	0.09	0.01
-0.09	0.02	0.99	0.04	-0.01	0.06	-0.00	0.01	-0.01	-0.00	0.01	0.00
0-.03	0.08	0.70	0.02	0.00	-0.12	-0.01	0.07	-0.05	0.00	0.01	0.00
0.04	0.08	0.83	0.02	0.01	-0.09	0.01	0.06	-0.02	0.00	0.01	0.00
-0.01	0.09	0.43	0.02	0.00	-0.13	-0.02	0.02	-0.06	-0.01	0.01	0.00
-0.02	0.11	0.80	0.04	0.01	0.12	0.00	0.01	0.00	-0.00	0.04	0.01