UCLA
Department of Statistics Papers

Title
A Modified MCEM Approach to Full Information Binary Factor Analysis

Permalink
https://escholarship.org/uc/item/2235w0fq

Authors
An, Xinming
Bentler, Peter

Publication Date
2008-12-30

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A Modified MCEM Approach to Full Information Binary Factor Analysis

Xinming An
Department of Psychology
University of California,
Los Angeles

Peter M Bentler
Department of Psychology and Statistics
University of California,
Los Angeles

March 2008

Abstract

Since binary data are ubiquitous in educational, psychological, and social research, methods for effectively exploring the underlying factor structure of such data are still undergoing development (Schilling and Bock 2005; Maydeu-Olivares and Joe 2005; and Song and Lee 2005). Two distinct types of methods have been developed, those relying on limited information from low-order marginal and joint frequency responses and those relying on the frequencies of all distinct item response vectors. The latter approach, a full information approach to binary factor analysis, has good optimality properties but is computationally demanding. Meng and Schilling (1996) developed a Monte Carlo EM (MCEM) fitting method for this model. Compared with the Gauss-Hermite method proposed by Bock and Aitkin (1981), the MCEM method is more stable and computationally easier. Under the MCEM framework, the E step is completed by approximating the conditional expectations through observations that are simulated by Markov Chain Monte Carlo methods, while the M step is completed by conditional maximization. In the E step proposed by Meng and Schilling (1996), respondents with the same response pattern are given the same latent factor scores. When there are many very difficult or easy items, this may dramatically decrease the number of response patterns and potentially cause the method to become less stable and accurate. This paper proposes a new E step for the MCEM method to avoid this problem. Simulation studies verify that the new method yields improved results. Possible follow up research also is discussed.

1 Introduction

Questionnaires and tests that aim to measure attitude, personality, and abilities are widely used in educational, psychological, and social research. Examples of such variables include agree or disagree items and right or wrong test items. Respondents' responses to such items are often dichotomous and are hypothesized to be determined by a small number of underlying latent factors. There are well established methods, such as exploratory and confirmatory factor analysis, that can be used to explore the latent factor structure of continuous variables. However, methods for exploring the underlying factor structure of binary variables are not equally well developed. Existing approaches fall into two main categories: those based on limited information obtained from first and second-order marginal and joint frequencies (e.g., Christoffersson 1975) and those which fit latent probit or logit models that use the frequencies of all distinct response vectors (e.g., Bock and Aitkin 1981). The latter approach, which may be called the full information binary factor model, is the focus of the current paper.

Under the probit latent variable model, the probability of a correct response for a single item is a normal-ogive function of a respondent's scores on the hypothetical latent factors. The model is thus a multidimensional normal-ogive item response model or, alternatively, a multivariate latent probit regression model,with the latent factors serving as regressors. This method can also be generalized into more complicated situations like polytomous data or mixed data (e.g., Shi and Lee 1997, Shi and Lee 2000, Lee and Song 2004). Meng and Schilling (1996) developed a Monte Carlo EM (MCEM) fitting method for this model. Compared with the Gauss-Hermite method proposed by Bock and Aitkin (1981), the MCEM method is more stable and computationally easier. Under the MCEM framework, the E step is completed by approximating the conditional expectations through observations that are simulated by Markov Chain Monte Carlo methods, while the M step is completed by conditional maximization. In the E step proposed by Meng and Schilling (1996), respondents with the same response pattern are given the same latent factor scores. It can be hypothesized that when there are many difficult or easy items, this can dramatically decrease the number of model-based response patterns, causing the method to be less stable and accurate. To alleviate this problem, in this paper we propose a new E step for the MCEM method, and report on simulation studies that evaluate the proposed new method.

2 Model Description

In the probit latent variable model, the probabilities of correct response for items are determined by respondents' latent factors through normal-ogive functions of a latent response variable y that is, in turn, generated by Thurstone's common factor model with d factors

$$y_{ij} = \alpha_{1j} Z_{i1} + \cdots + \alpha_{dj} Z_{id} + \varepsilon_{ij} \tag{1}$$

and a residual uniqueness or error term. For each of $J(J > d)$ latent response variables, y_j, respondent i obtains a correct manifest response to item j when his latent response y_{ij} equals or exceeds a threshold parameter γ_j. Only the binary response is observed, and there are $J(> d)$ binary variables in total. Assuming that ε_{ij} follows a $N(0, \sigma_j^2)$ distribution, the probability that respondent i with latent trait $z_i = (z_{i1}, \cdots, z_{id})^T$ correctly responds to item j is given by

$$P\{u_{ij} = 1 | z_i, \alpha_j, \sigma_j\} = P\{y_{ij} \geq \gamma_j | z_i, \alpha_j, \sigma_j\} = \Phi(z_i^T a_j + b_j) \tag{2}$$

where Φ is the CDF of $N(0,1)$, $a_j = (a_{1j}, \cdots, a_{dj})^T$ with $a_{mj} = \alpha_{mj}/\sigma_j, (m = 1, \cdots, d)$, and $b_j = -\gamma_j/\sigma_j$. In the context of test items, b_j is usually called the item difficulty, a_{mj} is called the item slope for factor m, and α_{mj} is called the item factor loading for factor m. These parameters are collected into the vector $b = (b_1, \cdots, b_J)$ and the matrix A, which is a $d \times J$ matrix whose jth column is a_j. Then $\theta = \{A, b\}$ is the set of parameters.

Given the linking function in (2), in this model it is also assumed that conditional on a respondent's latent trait or factor score, and the model parameters, responses to different items are independent of each other. When the number of respondents, n, is large relative to the number of items, Bock and Aitkin (1981) also proposed to sort the response patterns and count the number of occurrences, s_i, of distinct patterns, $u_i, i = 1, \cdots, n_0$. In this setup, i now indexes distinct response patterns rather than respondents. Thus in the probit latent variable model, we take U as the observed response patterns and $\theta = \{A, b\}$ as the set of parameters to be estimated. Maximum likelihood is the standard method used to estimate the parameters.

The likelihood function is needed to generate a methodology for obtaining the parameter estimates. Under the assumption of conditional independence of responses to different items, the likelihood function based on response patterns, u_i, is

$$L_i(\theta|u_i, z_i) = \prod_{j=1}^{J} [\Phi(z_i^T a_j + b_j)]^{u_{ij}} \times [1 - \Phi(z_i^T a_j + b_j)]^{1-u_{ij}} \tag{3}$$

Then the likelihood function of model parameters given all the responses patterns and latent factors from all n respondents is

$$L(\theta|U, Z) = \prod_{i=1}^{n_0} L_i = \prod_{i=1}^{n_0} \{\prod_{j=1}^{J} [\Phi(z_i^T a_j + b_j)]^{u_{ij}} \times [1 - \Phi(z_i^T a_j + b_j)]^{1-u_{ij}}\}^{s_i} \tag{4}$$

where U is the $n_0 \times J$ response matrix, and Z is the $n_0 \times d$ latent factor score matrix.

The fundamental estimation problem is that the above likelihood function involves latent variables Z which can not be observed, and hence the likelihood function can not be maximized directly. In order to make progress, as in maximum likelihood factor analysis for continuous variables, in the probit latent variable model it is typically assumed that data have been obtained from a sample of respondents drawn from a population with a certain multivariate distribution of the latent traits Z. In the exploratory factor analysis context, it is typically assumed that $z_1, \cdots, z_{n_0}$ are independently and identically distributed according to the standard d-variate normal distribution $N(0, I_d)$. Then the EM method is used to maximize the likelihood function using the standard E and M steps.

E step: In this step, the expectation of (3) is taken with respect to Z, eliminating Z from the likelihood function:

$$L_i(\theta|U) = E_Z\{\prod_{j=1}^{J}[\Phi(Zz_i^T a_j + b_j)]^{u_{ij}} \times [1 - \Phi(z_i^T a_j + b_j)]^{1-u_{ij}}\} \qquad (5)$$

M step: In this step the unconditional likelihood function is maximized:

$$L(\theta|U) = \prod_{i=1}^{n_0}\{E_Z\{\prod_{j=1}^{J}[\Phi(z_i^T a_j + b_j)]^{u_{ij}} \times [1 - \Phi(z_i^T a_j + b_j)]^{1-u_{ij}}\}\}^{s_i} \qquad (6)$$

These two steps are iteratively repeated until a specified convergence criterion is met. At this point, the maximum likelihood estimates (MLE) of all the parameters are available.

The probit latent variable model involves two indeterminacy problems. The first one is an indeterminacy of scale and location. The second one is the standard rotational indeterminacy of the multiple factor model. To fix the first indeterminacy problem, the usual choice is to set the mean vector of the latent distribution to 0 and the correlation matrix of the factors to the identity matrix I (**?**, Schilling and Bock 2005). To solve rotational indeterminacy, additional conditions have to be imposed on the model. In exploratory factor analysis, these conditions are usually imposed in an arbitary but suitable way so that the estimation method can be implemented without problems. A useful method is to restrict the factor loadings so that $\alpha_{jk} = 0, k > j$ (Schilling and Bock 2005). The resulting initial factor loading matrix is often not interpretable, so that it may subsequently be transformed according to typical rotation criteria, for example, the varimax criterion. In the restricted case of confirmatory factor analysis, the restrictions are given by the nature of the application (Christoffersson 1975).

We now turn to the methodology of interest, the MCEM approach to the binary factor analysis model.

3 MCEM Approach

For the MCEM method, the log function of equation (4) is taken, to obtain the log-likelihood function

$$logL(\theta|\theta^{(t)}, U) = \sum_{j=1}^{J}\{\sum_{i=1}^{n_0} s_i\{u_{ij}log\Phi(z_i^T a_j + b_j) + (1-u_{ij})log(1-\Phi(z_i^T a_j + b_j))\}\} \quad (7)$$

The E-step involves computing the expectation of $logL(\theta|\theta^{(t)}, U)$ over Z given U and the parameter estimate, $\theta^{(t)}$, from the previous M-step. That is

$$E(logL) = \sum_{j=1}^{J}\{\sum_{i=1}^{n_0} s_i\{u_{ij}E[log\Phi(z_i^T a_j + b_j)] + (1-u_{ij})E[log(1-\Phi(z_i^T a_j + b_j))]\}\} \quad (8)$$

where the expectations are with respect to $f(z_i|u_i, \theta^{(t)}), i = 1, \ldots, n_0$, which are the posterior distributions of z_i with response pattern $u_i, i = 1, \ldots, n_0$. To use the Monte Carlo method to calculate this expectation, draws from the above posterior distribution are taken.

As pointed out before, the prior distribution for z_i is d-variate standard normal distribution, $N(0, I_d)$. The likelihood function for z_i based on response pattern u_i is

$$L(z_i|u_i, \theta^{(t)}) = \prod_{j=1}^{J}[\Phi(z_i^T a_j + b_j)]^{u_{ij}}[1 - \Phi(z_i^T a_j + b_j)]^{1-u_{ij}} \quad (9)$$

Then the kernel of the posterior distribution for z_i is

$$f(z_i|u_i, \theta^{(t)}) \propto L(z_i|u_i, \theta^{(t)})N(0, I_d) \quad (10)$$

Although the posterior distribution is complicated, a method like Metropolis can be used to make draws from it. Once K draws are available, $\{z_{k,i}, k = 1, \ldots, K\}$, from $f(z_i|u_i, \theta^{(t)}), i = 1, \ldots, n_0$, these can be used to form the Monte Carlo estimates for the expected log-likelihood function. Specifically, (8) can be replaced by

$$E(logL) = \sum_{j=1}^{J}\{\sum_{i=1}^{n_0} s_i\{\frac{u_{ij}}{K}\sum_{k=1}^{K} log\Phi(z_{k,i}^T a_j + b_j) + \frac{1-u_{ij}}{K}\sum_{k=1}^{K} log(1 - \Phi(z_{k,i}^T a_j + b_j))\}\} \quad (11)$$

The M step then maximizes this estimated expected log-likelihood function by the Newton-Raphson method to obtain parameter estimates for the next iterate.

Although it is possible to use the Metropolis method make draws from any distribution, including the posterior distribution $f(z_i|u_i, \theta)$ used here, as long as a function proportional to the density can be calculated, its efficiency can be improved by taking into account the properties associated with different distributions. To improve this point for current model, Meng and Schilling (1996) proposed to further augment $\{U, Z\}$ to $\{U, Z, X\}$, where the new data $X = (x_{ij})$ are assumed to have independent conditional normal distributions given Z:

$$x_{ij}|z_i, a_j, b_j \sim N(z_i^T a_j + b_j, 1) \tag{12}$$

Then equation (2) can be expressed as

$$Pr(u_{ij} = 1|z_i, a_j, b_j) = Pr\{x_{ij} \geq 0|z_i, a_j, b_j\} = \Phi(z_i^T a_j + b_j) \tag{13}$$

In other words, x_{ij} is defined so that $u_{ij} = 1_{(x_{ij} \geq 0)}$, where 1_A is the indicator function of a set A. If x_i was observed, the new posterior distribution for z_i will be

$$f(z_i|x_i, \theta) \propto N(0, I_d) \prod_{j=1}^{J} N(x_{ij}; z_i^T a_j + b_j, 1). \tag{14}$$

Rearranging the kernel of this posterior distribution, it follows that

$$z_i|x_i, \theta \sim ind.N_d(VA(x_i - b), V), i = 1, \ldots, n_0 \tag{15}$$

where $V = (I + AA^T)^{-1}$. Because given Z, U, and θ, x_{ij}'s are independent $N(z_i^T a_j + b_j, 1)$ truncated at the left by zero if $u_{ij} = 1$ and at the right by zero if $u_{ij} = 0$, drawing from $f(X|Z, U, \theta)$ is the same as drawing from $n_0 \times J$ univariate truncated normal distributions. These facts suggest the use of the Gibbs sampler to iterate between draws from $f(\theta|X, U, \theta)$,which is given by (15), and from $f(X|\theta, U, \theta)$ until the equilibrium distribution $f(X, Z|U, \theta)$ is reached.

If both X and Z were observed, then the new log-likelihood function for θ will be

$$logL(\theta|X,Z) = \sum_{j=1}^{J}\sum_{i=1}^{n_0} s_i log(N(x_{ij}; z_i^T a_j + b_j, 1)) \propto \sum_{j=1}^{J}\sum_{i=1}^{n_0} s_i \frac{-1}{2}(x_{ij} - z_i^T a_j - b_j)^2. \tag{16}$$

From this log-likelihood function, it can be seen that the estimation of θ, conditional on X, Z, can based on the sufficient statistics,$\{Z^TZ, Z^TX, 1^TX, 1^TZ\}$. The M step here is equivalent to the MLE estimation of J separate multiple regressions, by considering the $x_j = (x_{1j}, \ldots, x_{nj})^T$ as dependent variables, and the Z as predictors. Thus this E step only requires computing the conditional expectations of those sufficient statistics. The M step is also very easy to implement.

Following Meng and Schilling (1996), let x_i^T be the ith row of $X, i = 1, \ldots, n_0$. Then

$$x_i|\theta \sim iidN_J(b, I + A^TA). \tag{17}$$

Further, let $e(u_i|\theta) = E[(x_i - b)|u_i, b, A]$ and $D(u_i|\theta) = E[(x_i - b)(x_i - b)^T|u_i, b, A]$. Then it can be shown that

$$E[1^TX|U,\theta] = \sum_{i=1}^{n_0} s_i[e^T(u_i|\theta) + b^T] \tag{18}$$

$$E[1^TZ|U,\theta] = [\sum_{i=1}^{n_0} s_i e^T(u_i|\theta)]A^TV^T \tag{19}$$

$$E[Z^TX|U,\theta] = VA\sum_{i=1}^{n_0} s_i[D(u_i|\theta) + e(u_i|\theta)b^T] \tag{20}$$

$$E[Z^TZ|U,\theta] = n * V + VA[\sum_{i=1}^{n_0} s_i D(u_i|\theta)]A^TV^T \tag{21}$$

Suppose that K draws $\{X_k, Z_k, k = 1, \ldots, K\}$ from $f(X, Z|U, \theta)$ have been obtained. Meng and Schilling (1996) suggested that there are two ways to use these draws in the E step. The first way is simply to replace the sufficient statistics $\{Z^TZ, Z^TX, 1^TX, 1^TZ\}$ by their corresponding sample averages. The second way is to use only $\{X_k, k = 1, \ldots, K\}$ to compute

$$\hat{e}(u_i|\theta) = \frac{1}{K}\sum_{k=1}^{K}[x_{i,k} - b] \tag{22}$$

and

$$\hat{D}(u_i|\theta) = \frac{1}{K}\sum_{k=1}^{K}[x_{i,k} - b][x_{i,k} - b]^T \tag{23}$$

and substitute them for $e(u_i|\theta)$ and $D(u_i|\theta)$ in (18)-(21).

To summarize, the MCEM method developed by Meng and Schilling (1996) works as follows

E Step: Make K draws from each posterior distribution $f(x_i, z_i|u_i, \theta)$. Estimate $\hat{e}(u_i|\theta)$ and $\hat{D}(u_i|\theta)$ by equations (22) and (23). Then substitute them for $e(u_i|\theta)$ and $D(u_i|\theta)$ in (18)-(21) to obtain the expectation of sufficient statistics.

M Step: Estimate the parameters by equations below

$$A = (Z^T Z - (1^T Z)^T (1^T Z)/n)^{-1}(Z^T X - (1^T Z)^T (1^T X)/n) \tag{24}$$

$$b = (1^T X - 1^T ZA)/n; \tag{25}$$

They found this to be a successful methodology, although it has been suggested more recently that adaptive quadrature would be more effective (Schilling and Bock 2005). Staying within the general framework of Meng and Schilling (1996), we now develop an alternative potential improvement.

4 Modified MCEM

As noted previously, when the number of respondents, n, is large relative to the number of items, Bock and Aitkin (1981) proposed to sort the response patterns and count the number of occurrences of distinct patterns. Meng and Schilling (1996) also adopted this strategy in their MCEM approach, which has the effect that every respondent who has the same pattern of observed responses also is assumed to have identical latent trait scores. Although this simplifying assumption does make the likelihood equation and the M step easier to deal with, a potential disadvantage of doing so is that some information is lost. This information might be obtainable from the E step if it were possible to allow respondents to obtain different latent trait scores drawn from the same posterior distribution, especially when items are very difficult or very easy and hence many respondents will share the same response patterns.

We thus propose a modified MCEM procedure to alleviate this drawback. We treat respondents that have the same response pattern, u_i, as having the same posterior distribution, $f(z_i, x_i|u_i, \theta)$, but we do not force such respondents to have exactly the same latent trait scores. This is achieved as follows. In the E step, for subjects with the same response pattern u_i, we make K draws for each respondent from the same posterior distribution $f(z_i, x_i|u_i, \theta)$. We repeat this procedure for all the response patterns. Then in the M step, we treat all the $\sum_{i=1}^{n_0} s_i \times K$ draws as a whole and use them to estimate all the item parameters $\theta = \{A, b\}$. The potential benefits of doing so are that it may be possible to obtain better parameter estimates, and potentially the MCEM computations may be more stable when there are many very difficult or very easy items or binary variables.

The modified MCEM (MMCEM)method works as follows

E Step: Respondents having the same response pattern are treated as having the same posterior distribution, $f(x_i, z_i|u_i, \theta)$, but not as having exactly the same value for latent variables. For each respondent with a given response pattern, u_i, K draws are taken. Totally there are $K \times s_i$ draws from the posterior distribution $f(x_i, z_i|u_i, \theta)$ and $\sum_{i=1}^{n_0} s_i \times K = K \times n$ draws for the E step.

M Step: The parameters are estimated by treating X as dependent variables, and Z as predictors using linear multiple regression.

As usual, the E and M steps are iterated until a convergence criterion is met.

5 Simulation Study

In the last section, we introduced a modified procedure for the E step of MCEM approach. We expect that this MMCEM method will work better when there are some very difficult or very easy items, but this conjecture needs empirical evaluation. In this section, we describe four simulation studies aimed at evaluating the performance of the proposed MMCEM method as compared to the original MCEM method. We focus on the convergence rates of these alternative procedures as well as accuracy of the parameter estimates. Convergence is monitored by plotting the largest parameter change against the EM iteration, and accuracy was assessed by computing the mean square error (MSE) between the estimated and generated loadings for each factor.

Before discussing these simulation studies, we need to evaluate the Gibbs sampler to decide how to set up the E step. Figures 1 show that the Gibbs sampler mixes very fast in our setup, with autocorrelations typically decaying to near zero after five iterations. To be more conservative, in the simulations below we discard the first 15 iterations, and then choose every tenth iteration afterward until the required number of draws is reached.

5.1 Two Factor Simulations

To evaluate the performance of both methods in a low dimensional situation, we introduce two simulations of 32 item binary response tests here.

Simulation 1: 32-item test with factor loadings as shown in table 1. Item difficulties are set to zero. 1500 responses were simulated, yielding 1500 response patterns.

Simulation 2: 32-item test with factor loadings and difficulties are shown in table 1. 1500 responses was simulated, yielding 1130 response patterns.

To assess the estimation accuracy and convergence rate, the original MCEM and modified MMCEM methods are applied with different specifications to the data generated from simulations 1 and 2. These specifications balance the total number of draws in the E step for the two methods.

MCEM: Use 8 draws from the posterior distribution of each response pattern for simulation 1 and 11 draws for simulation 2, yielding 1500*8=12000 draws for simulation 1 and 11*1130=12430 draws for simulation 2.

MMCEM: Use 8 draws for both simulation 1 and 2, yielding 8*1500=12000 draws for both simulations.

In simulation 1, every subject has their own response pattern. In such a situation, the two MCEM methods are equivalent. Figure 2, which plots the largest parameter change against the number of iterations, shows that both methods converged in around 20 iterations. Their mean square errors (MSEs) are also roughly the same, 0.0025 for MCEM, and 0.0023 for MMCEM.

In simulation 2, many subjects share the same response pattern. Thus we expect that the MCEM method will show more fluctuation and converge slower than the MMCEM method. Comparing the left panels of Figures 3 and 4, with the fixed approximate 12000 random numbers in E step, the MMCEM method converges in around 40 iterations, but the MCEM method shows substantial fluctuation, suggesting nonconvergence. When the number of MCEM draws is increased from 11 to 33 and 55, as shown in the middle and right panels of Figure 3, its fluctuation decreases. With 55 MCEM draws, its largest parameter change gradually decreases as the number of iterations increase; it can be considered as converged at roughly 50 iterations. However, the MMCEM method converges similarly with roughly 1/5 the number of draws. At convergence, the MSEs of the two methods also are very similar at 0.0047 for MMCEM with 40 draws and 0.0055 for MCEM with 55 draws.

5.2 Five Factor Simulations

To evaluate the performance of the two methods in a higher dimensional situation, we utilize two five factor models whose specifications are given in Table 2 and as follows.

Simulation 3: 32-item test with factor loadings as shown in Table 2, and item difficulties set to zero. 1500 responses were simulated, yielding 1500 response patterns.

Simulation 4: 32-item test with factor loadings and difficulties as shown in Table 2. 1500 responses were simulated, yielding 1270 response patterns.

The MCEM and MMCEM methods are applied to the five factor simulation data. Again slightly different specifications are used in order to equalize the total number of draws.

MCEM: Use 8*1500=12000 draws for simulation 3 and 10*1270 =12700 draws for simulation 4.

MMCEM: Use 8*1500=12000 draws for both simulations 3 and 4.

For simulation 3, as is expected, and is shown in Figure 5, the convergence rates of the two methods are basically the same. Furthermore, their MSEs are also very close, 0.0025 for MCEM, 0.0028 for MMCEM. Comparing Figures 2 and 5, it can be seen that the convergences of both methods with 5 dimensions are roughly the same as they were in the 2 dimension situation. Their accuracies are also quite similar under 2 or 5 dimensions. These results show that Monte Carlo methods are very suitable for solving the high dimensional integration associated with the binary factor model.

The convergence rates for simulation 4 are shown in Figures 6 and 7. Again the MMCEM method converges faster than the MCEM method. As expected, once converged, their MSEs are also basically the same, 0.004 for MCEM, and 0.0038 for MMCEM.

6 Discussion

MCEM methods have a number of theoretical strong points, including that they are very suitable for high dimensional models, that the M steps often allow closed form estimation, and that they can be easily adapted to various complicated models such as the mixed data latent variable model (Shi and Lee 2000) and the nonlinear structural model for hierarchically mixed data (Lee and Song 2004). Of course, these methods also have disadvantages, including (1) in the E step, random numbers need to be drawn from a truncated normal distribution, which may be difficult; (2) when a model involves a lot of missing data, the resulting EM algorithm may have a very slow convergence rate; and (3) variation introduced by Monte Carlo simulation make it impossible to obtain any definite values for the likelihood-ratio criterion which is needed for model testing (Schilling and Bock (2005)). Concerning the last point, Lee and Song (2004) proposed an approach based on the idea of path sampling (Meng and Wong 1996), which can be used to compute the observed-data likelihoods associated with specific models of interest, or their ratios, so that the likelihood ratio or the Bayesian information criteria (BIC) can be evaluated for model comparison or hypothesis testing. This interesting proposal may be adaptable to the current context, but so far not many theoretical or simulation results are known on the performance of this sampling based model comparison approach.

In conclusion, simulation studies in Meng and Schilling (1996) and Schilling and Bock (2005) indicate that the MCEM method is more accurate than non-adaptive quadrature based methods. However, the simulation studies reported in the last section indicate that the MCEM method proposed by Meng and Schilling (1996) can be improved when items are very difficult or very easy and hence there may be fewer response patterns than number of respondents. The Meng-Schilling methodology requires that respondents having the same response pattern also have the same latent trait scores. As we have shown, this assumption can be relaxed, yielding a more stable and accurate MMCEM method that can outperform the original MCEM method.

References

Bock, R. D. and Aitkin, M. (1981). Marginal maximum-likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46:443–459.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40:5–32.

Lee, S. and Song, X. (2004). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics*, 60:624–636.

Maydeu-Olivares, A. and Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2(n) contingency tables: a unified framework. *Journal of the American Statistical Association*, 100:1009–1020.

Meng, X. and Wong, W. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6:831–860.

Meng, X. L. and Schilling, S. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, 91:1254–1267.

Schilling, S. and Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70:533–555.

Shi, J. Q. and Lee, S. Y. (1997). Estimation of factor scores with polytomous data by the EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 50:215–226.

Shi, J.-Q. and Lee, S.-Y. (2000). Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society: Series B*, 62:77–87.

Song, X. Y. and Lee, S. Y. (2005). Maximum likelihood analysis of nonlinear structural equation models with dichotomous variables. *Multivariate Behavioral Research*, 40:151–177.

Appendix: Tables and Figures

Figure 1: Autocorrelation Plot of Gibbers sampler

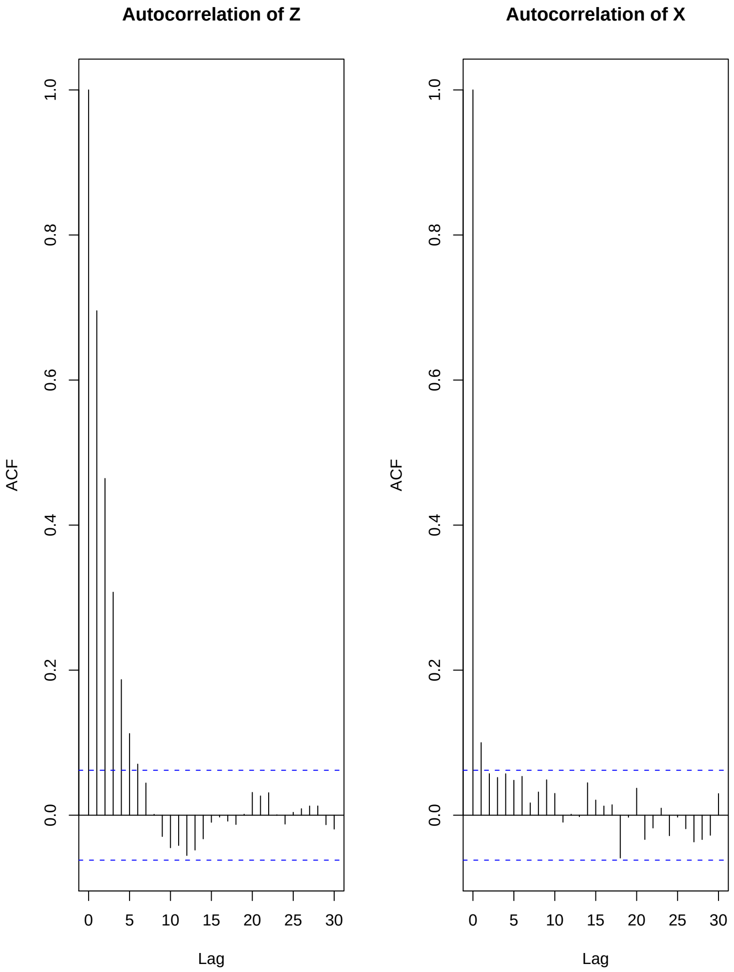

Table 1: Factor Loadings (Difficulties) for Simulation 1 and 2. Other Loadings are 0

Items					
1-6	6-12	12-18	18-25	25-28	29-32
F1	F2	F1	F2	F1	F2
0.72 (1.5)	0.41	0.37	0.70	0.29	0.75(1.5)
0.69 (1.6)	0.63(1.6)	0.77(1.3)	0.82(1.2)	0.67(1.3)	0.71(1.0)
0.80 (1.5)	0.75(1.3)	0.76(1.4)	0.67(1.3)	0.81(1.4)	0.69(1.2)
0.73 (1.2)	0.68(1.5)	0.75(1.6)	0.77(1.1)	0.78(1.5)	0.68(1.3)
0.67 (1.2)	0.72(1.7)	0.61(1.3)	0.75(1.5)	*	*
0.71 (1.6)	0.67(1.3)	0.83(1.1)	0.73(1.6)	*	*
*	0.78(1.4)	0.70(1.1)	0.81(1.1)	*	*
*	*	*	0.62(1.3)	*	*

Table 2: Factor Loadings (Difficulties) for Simulation 3 and 4. Other Loadings are 0

Items				
1-6 F1	6-12 F2	12-18 F3	18-25 F4	25-32 F5
0.72 (1.5)	0.41	0.37	0.70	0.29
0.69 (1.6)	0.63(1.6)	0.77(1.3)	0.82(1.2)	0.67(1.3)
0.80 (1.5)	0.75(1.3)	0.76(1.4)	0.67(1.3)	0.81(1.4)
0.73 (1.2)	0.68(1.5)	0.75(1.6)	0.77(1.1)	0.78(1.5)
0.67 (1.2)	0.72(1.7)	0.61(1.3)	0.75(1.5)	0.75(1.5)
0.71 (1.6)	0.67(1.3)	0.83(1.1)	0.73(1.6)	0.71(1.0)
*	0.78(1.4)	0.70(1.1)	0.81(1.1)	0.69(1.2)
*	*	*	0.62(1.3)	0.68(1.3)

Figure 2: MCEM and MMCEM for Simulation 1

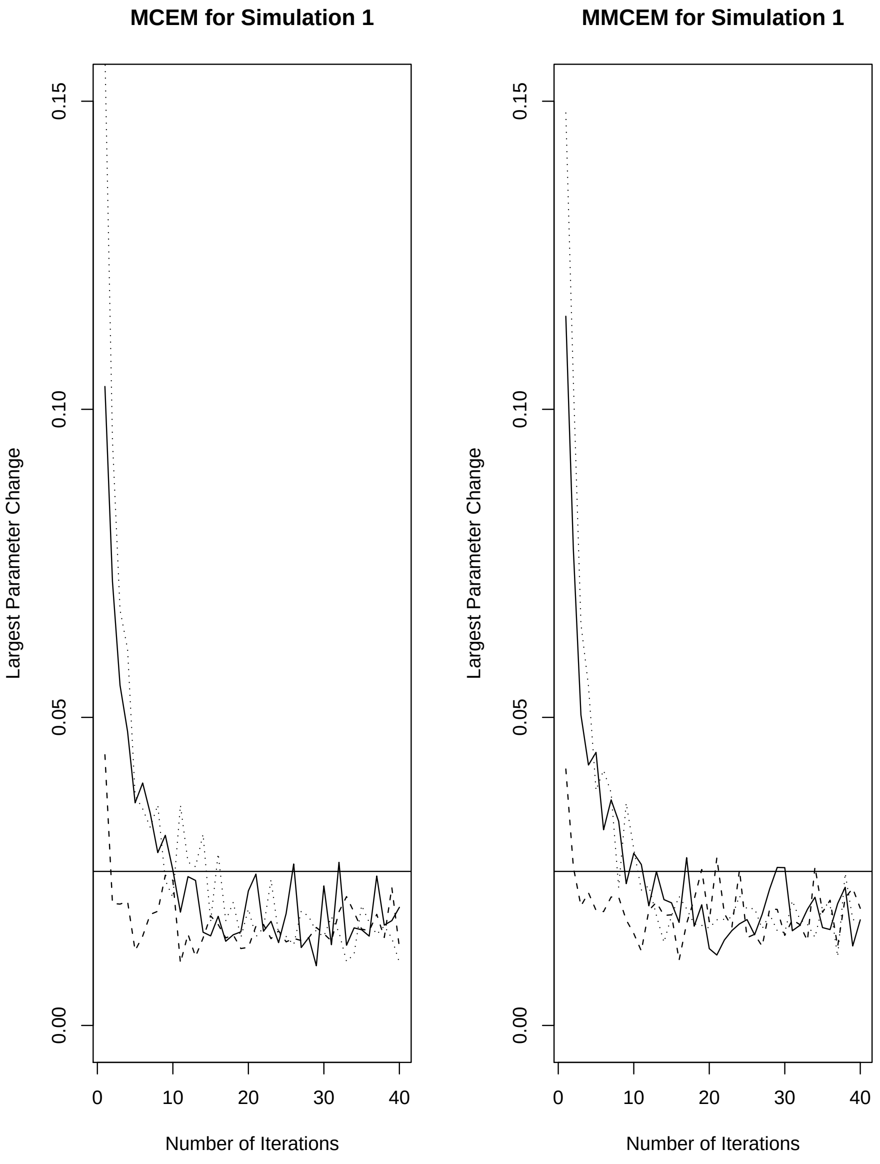

Figure 3: MCEM for Simulation 2

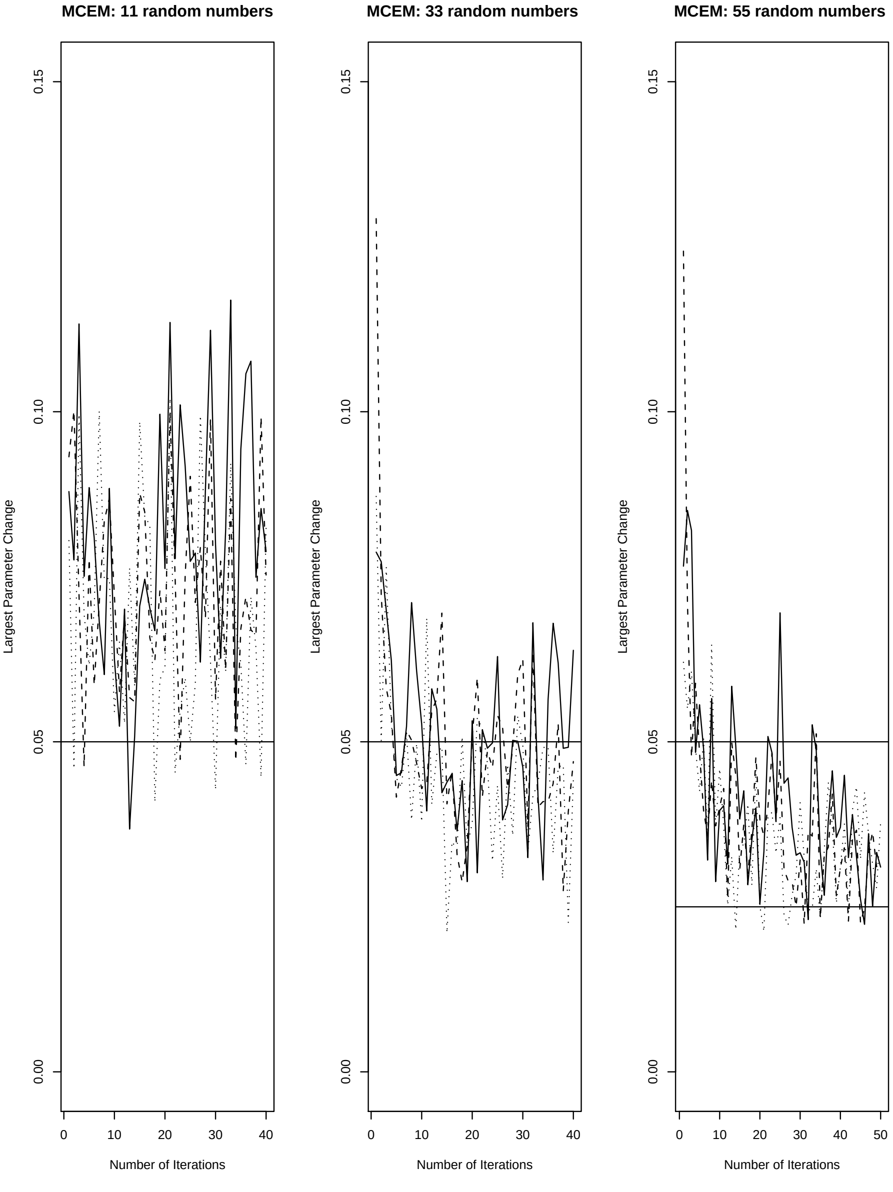

Figure 4: MMCEM for Simulation 2

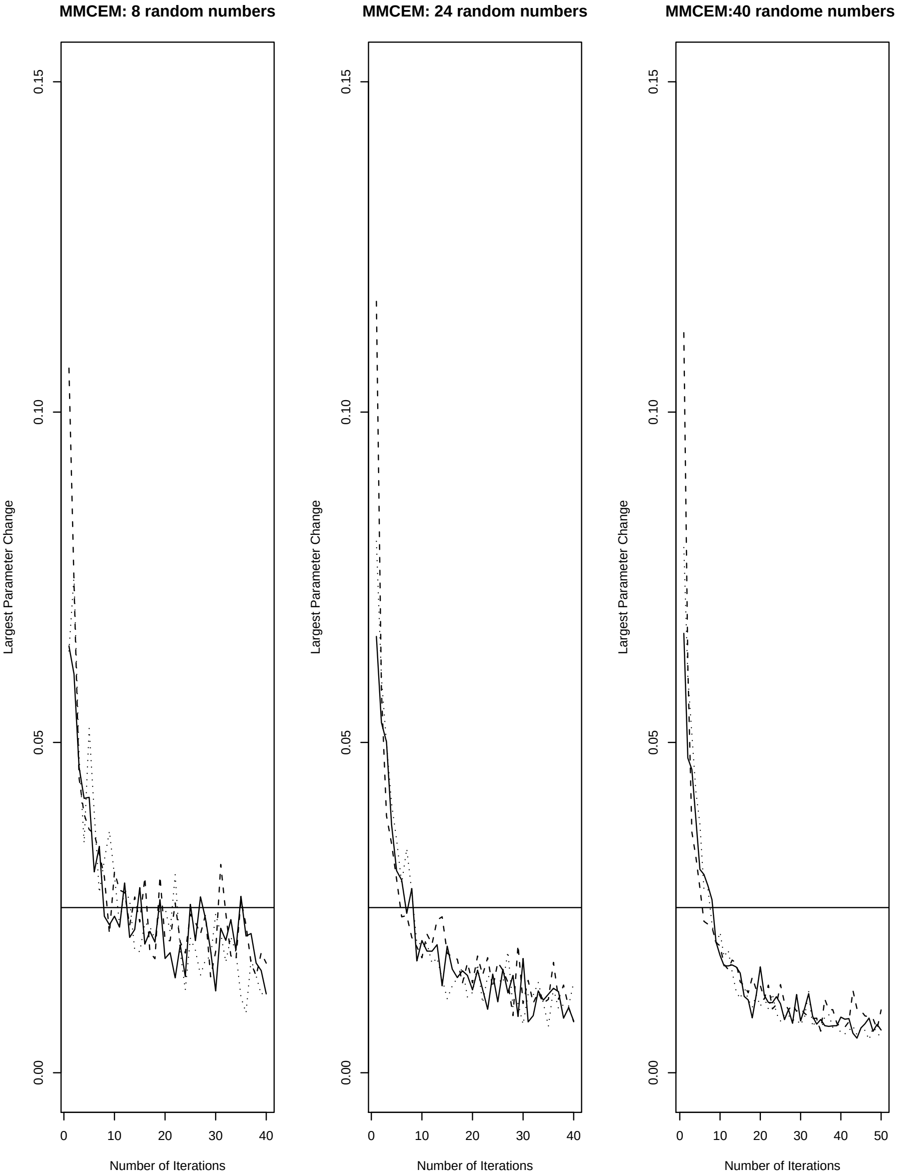

Figure 5: MCEM and MMCEM for Simulation 3

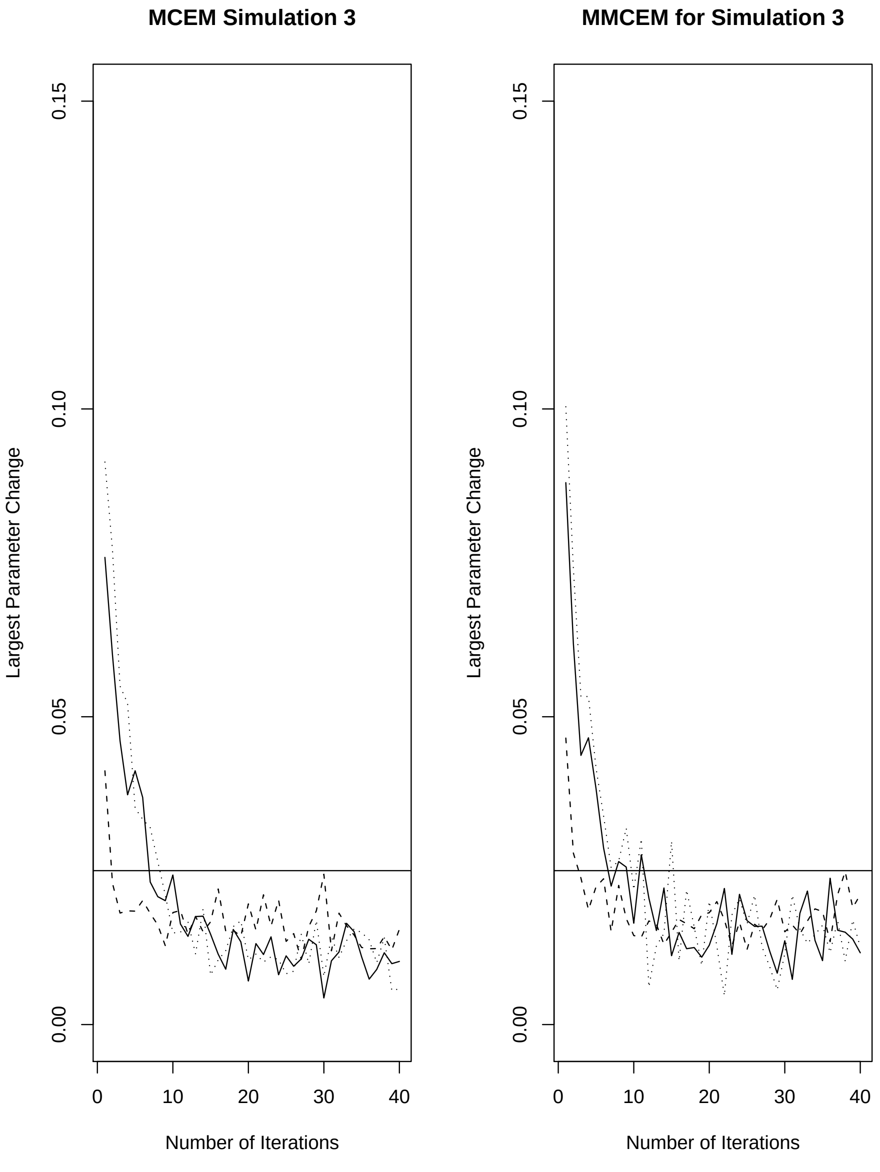

Figure 6: MCEM for Simulation 4

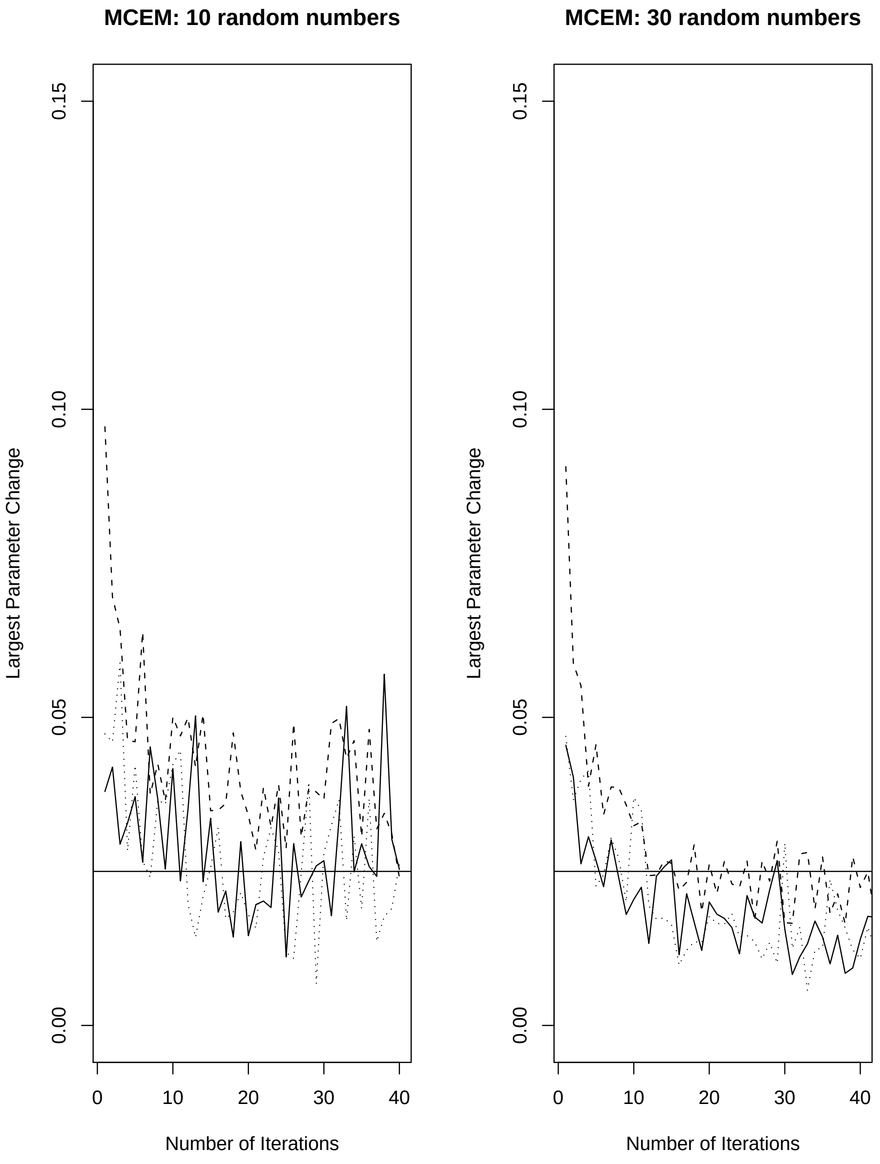

Figure 7: MMCEM for Simulation 4