Stochastic Approximation EM for Exploratory Item Factor Analysis

The estimation of item and person parameters in the framework of item response theory [IRT] is a common practice in programs of educational and psychological assessment. More recently, there has been great interest and methodological development in exploratory factor analysis of categorical variables. Areas of application include understanding the structure of new forms of achievement test items, quality of life assessments, anxiety disorders, alcohol problems, and physical functioning. While some assessments are developed according to a theoretical blueprint, the empirical item structure may be informative in earlier stages of test development. Generally, factor analysis as a method of multidimensional item response theory [MIRT] is an important tool for improving or critiquing behavioral instruments.

For classical exploratory factor analysis with continuous data, a research procedure is carried out that typically consists of three steps: (1) choosing the number of factors, (2) estimating factor loadings with a minimal number of constraints for identification, and (3) rotating factor loadings to an interpretable structure. The main issue addressed in this paper concerns step (2) for ordered categorical item responses. An algorithm is proposed in which estimation is conceptualized and programmed in terms of missing or augmented data (Tanner & Wong, 1987). As shall be seen, once those data are generated the procedure is similar to that of factor analysis of continuous response variables. The work described in this paper extends that of Meng and Schilling (1996) but adds a key feature that greatly improves the efficiency of estimation. The resulting algorithm is fast and can be easily coded in statistical processing languages. No derivatives or matrix inversion is required. Below, this algorithm is investigated in two simulations and one empirical study.

Estimation Approaches

Bock and Aitken (1981) in a seminal paper had proposed an expectation maximization [EM] algorithm based on the work of Dempster, Laird, and Rubin (1977), and extended this procedure to multidimensional item-response data. As they noted, this EM algorithm has an alternative formulation according to the missing information principle. However, Bock and Aitken also realized that numerical quadrature becomes impractical for marginalizing across missing variables as the number of dimensions increases (the curse of dimensionality). Until recently, estimation of factor coefficients, including loadings and thresholds, has been a major obstacle with categorical item responses, especially with high-dimensional data or a large number of variables. A major trend in addressing computational challenges is the implementation of stochastic approaches to EM (e.g. Albert, 1992; Meng & Schilling, 1996; Cai, 2010) for marginalization. In this paper, the goal is to contribute to this trajectory with the implementation of an alternative EM algorithm for factor analysis based on stochastic approximation EM [SAEM] (Delyon, Lavielle, and Moulines, 1999).

The concept of missing data as introduced by Tanner and Wong (1987) provided an important link between Monte Carlo methods and the EM algorithm. Using this bridge, Béguin and Glas (2001) extended the Gibbs sampling approach of Albert (1992) to MIRT factor models, while others have advanced stochastic versions of the Newton-Raphson algorithm (Gu & Kong, 1998; Cai, 2010) in which Metropolis-Hastings [MH] sampling is combined with the Robbins-Monro procedure [MH-RM] for establishing convergence (Robbins & Monro, 1951). Below, the

EM algorithm is briefly described, followed by an outline of the SAEM procedure. An improved algorithm is then presented for the estimation of coefficients in MIRT models relying on the foundational work of Meng and Schilling (1996) who proposed a Gibbs sampling approach to the factor analysis of dichotomous variables based on sufficient statistics. Rubin and Thayer (1982) had described a highly similar approach for the factor analysis of observed continuous variables as this algorithm's treatment of augmented data. In the current paper, an innovative strategy is added to the general procedure of Meng and Schilling for carrying out the M-step of the EM algorithm with latent sufficient statistics.

The EM and SAEM Algorithms

The EM algorithm as described by Dempster et al. (1977) can be described as a missing data model. Given observed data *y*, let

$$l\left(\xi \mid y\right) = \log f\left(y \mid \xi\right) \tag{1}$$

be the observed data log likelihood, where ξ is a set of fixed parameters to be estimated. In the EM algorithm, the likelihood estimate of ξ on iteration *t* is obtained from the marginal log likelihood

$$Q_t = Q\left(\xi \mid \xi^{(t-1)}\right) = \int \log f\left(\psi \mid \xi\right) f\left(\psi \mid \xi^{(t-1)}\right) d\psi\,, \tag{2}$$

where ψ is a set of missing data and $\log f\left(\psi \mid \xi\right)$ the complete data likelihood.

In many situations, the maximization of Q_t is simpler than the maximization of $l\left(\xi \mid y\right)$ (Dempster et al., 1977). The EM algorithm proceeds iteratively in two steps at iteration *t*:

1. E-step: Take the expectation Q_t

2. M-step: Estimate $\xi^{(t)} = \arg\max_{\xi} Q_t$

On iteration t, Q_t is evaluated in the E-step and maximized in the M-step. Among models of the exponential family sufficient statistics exist for model parameters. In turn, if a sufficient statistic exists for a parameter then the MLE estimate must be a function of it (see Hogg, McKean, & Craig, 2014). For exponential family distributions Dempster et al. (1977) showed the E-step has the particularly simple form of updating the sufficient statistic. Despite the advantage of monotone convergence, the EM algorithm can be very slow with numerical integration over the missing data. Further, efficiency diminishes rapidly as the number of dimensions increases.

The Stochastic Approximation EM algorithm (Delyon et al., 1999) which bears a similarity to the method of (Celeux & Diebolt, 1992), is a combination of stochastic EM and the Robbins–Monro method. In this approach, the E-step can be replaced by a stochastic approximation based on multiple draws of the missing data. A single draw can be used if the E-step is computationally intensive, which is the strategy taken below. Let y be the observed data, then at iteration *t* of SAEM:

1. *Missing data*. Obtain one draw of ψ from the distribution $f\left(\psi \mid \xi^{(t)}, y\right)$
2. *S-step (stochastic approximation)*. Update $Q_t = Q_{t-1} + \gamma_t \left\{\log f\left(\psi \mid \xi\right) - Q_{t-1}\right\}$, where γ_t is a positive sequence such that $\gamma_t = 1$, $\sum_{t=1}^{\infty} \gamma_t \to \infty$, and $\sum_{t=1}^{\infty} \gamma_t^2$ is finite.
3. *M-step*. Update $\xi^{(t)} = \arg\max_{\xi} Q_t$.

Implementation of SAEM (as well as EM) is highly simplified when the complete likelihood belongs to a curved exponential family. In Step 2 of the SAEM algorithm, the vector of sufficient statistics $s\left(y, \psi\right)$ is computed at iteration *t* and then updated according to

$$s^{(t)} = s^{(t-1)} + \gamma_t \left\{ S\left(y, \psi^{(t)}\right) - s^{(t-1)} \right\}. \tag{3}$$

$S\left(y, \psi^{(t)}\right)$ is the sufficient statistic calculated on iteration *t*. In the M-step, $\xi^{(t)}$ is updated as a function of s, the sufficient statistics. The SAEM algorithm has been shown to converge to the MLE under general conditions (Delyon et al., 1999).

MIRT Factor Model

Assume a set of *M* test items and *N* examinees with indices labeled $j = 1 \ldots M$ and $i = 1 \ldots N$. For ease of presentation item responses in this section are assumed to be dichotomous with correct and incorrect responses scored $Y_{ij} = 1$ and $Y_{ij} = 0$, respectively. In the next section, a more general approach is presented that is applicable to polytomous items with dichotomous items as a special case.

Assume there is a vector of *K* latent variables that account for an examinee's observed item responses. For mathematical convenience, the cumulative normal function (also termed the normal ogive) is used to model item responses rather than the logistic function, though both give nearly isomorphic results in most applications. The normal ogive models provided a flexible tool for the stochastic method.

Using the cumulative normal distribution function Φ, a multidimensional normal ogive model for a correct response on item *j* presented to examinee *i* is given by

$$\begin{aligned} P_{ij}\left(Y_{ij} = 1\right) &= \Phi\left(Y_{ij} = 1 \mid \eta_{ij}\right) \\ \eta_{ij} &= \mathbf{A}_j \boldsymbol{\theta}_i - b_j, \end{aligned} \tag{4}$$

where $\boldsymbol{\theta}_i$ is a K×1 vector of latent factor scores or abilities for examinee i and b_j is an intercept parameter. Discrimination for item j is generalized to $\mathbf{A}_j$, a 1×K vector of slopes or factor loadings which signify that an examinee's response may be affected by weighted combinations of K different skills or abilities.

In the latent space the response model is represented as a multivariate regression of missing item responses (also known as propensities) z on $\boldsymbol{\theta}$:

$$\begin{aligned} z_{ij} &= \mathbf{A}_j\boldsymbol{\theta}_i - b_j + \varepsilon_{ij} \\ &= \eta_{ij} + \varepsilon_{ij}. \end{aligned} \tag{5}$$

For examinee i, $\mathbf{z}_i$ is an $M \times 1$ vector of missing item responses. For a single item and examinee, let $\mathbf{A}_j$ denote row j of the $M \times K$ matrix $\mathbf{A}$ of slopes (or factor loadings), and a $K \times 1$ vector of latent factor scores (also a missing variable) for an examinee be specified as $\boldsymbol{\theta}_i \sim N(\mathbf{0},\mathbf{I})$. The $M \times 1$ vector $\mathbf{b}$ holds the item's intercept b_j, and $\varepsilon_{ij} \sim N(0,1)$ is a random measurement error. In exploratory factor models additional identification restrictions are required on the factor loading matrix **A**. Fox (2010) provides more details on the Bayesian approach to item response modeling.

MIRT Factor Analysis as a Missing Data Problem

Let the set of item parameters be symbolically represented as ξ, the observed data as y, and the unobserved or missing variables as ψ. Recall that persons are indexed by i and items by j. The goal is to maximize the observed data log likelihood $l(\xi \mid y) = f(y \mid \xi)$, but it is often easier to work with the Q function used in EM estimation shown in Equation (2) . Given the current estimate of the fixed item parameters $\xi^{(t)}$, the posterior distribution $f(\psi \mid \xi', y)$ of the unobserved latent variable ψ is generated and the parameter estimate is updated by maximizing $Q\left(\xi \mid \xi^{(t)}\right)$. In the case of the MIRT factor model Equation (2) can be modified as

$$Q(\xi \mid \xi') = \int \log f\left(z, \theta \mid \xi\right) f\left(\theta \mid z, \xi'\right) f\left(z \mid \xi', y\right) d\theta dz, \tag{6}$$

where ξ' represents a fixed value of the item parameters. The complete data likelihood is

$$\log f\left(z, \theta \mid \xi\right) = \exp\left[-\frac{1}{2}\sum_i \left\{\left[\mathbf{z}_i - \left(\mathbf{A}\boldsymbol{\theta}_i - \mathbf{b}\right)\right]^T \left[\mathbf{z}_i - \left(\mathbf{A}\boldsymbol{\theta}_i - \mathbf{b}\right)\right] + \boldsymbol{\theta}_i^T \boldsymbol{\theta}_i\right\}\right], \tag{7}$$

where the multivariate variance term is $\boldsymbol{\Sigma} = \mathbf{I}$ which constitutes an identification restriction. Note the last term in (7) represents the normal prior $\boldsymbol{\theta} \sim N\left(\mathbf{0}, \mathbf{I}\right)$ which has the effect of fixing the latent scale for ability. Let $\xi = \left(\mathbf{A}, \mathbf{b}\right)^{(t)}$ and $\mathbf{z}, \boldsymbol{\theta}$ be a set of missing latent variables. As shown by Béguin and Glass (2001), the values in $\mathbf{z}$ can be sampled from a truncated normal distribution and $\boldsymbol{\theta}$ can be sampled as

$$\begin{aligned} \hat{\boldsymbol{\theta}} &\sim \left(\mathbf{I} + \mathbf{A}^T\mathbf{A}\right)^{-1} \mathbf{A}^T \left(\mathbf{z}_i + \mathbf{b}\right) \\ \boldsymbol{\theta} &\sim N\left(\hat{\boldsymbol{\theta}}, \left(\mathbf{I} + \mathbf{A}^T\mathbf{A}\right)^{-1}\right), \end{aligned} \tag{8}$$

Maximizing $Q\left(\xi \mid \xi^{(t)}\right)$ with respect to **A** and **b** then results in

$$\mathbf{b} = n^{-1}\left(\mathbf{A}\sum_i \boldsymbol{\theta}_i - \sum_i \mathbf{z}_i\right) \tag{9}$$

$$\mathbf{A}_j = \left[\sum_i \left(\mathbf{z}_j + \mathbf{b}_j\right)\boldsymbol{\theta}_i^T\right]\left(\sum_i \boldsymbol{\theta}_i \boldsymbol{\theta}_i^T\right)^{-1}. \tag{10}$$

The SAEM algorithm's computational efficiency results from the use of sufficient statistics of latent samples within each Gibbs cycle to estimate ξ:

$$\sum_i \boldsymbol{\theta}_i, \sum_i \mathbf{z}_i, \sum_i \mathbf{z}_i \mathbf{z}_i^T, \sum_i \boldsymbol{\theta}_i \boldsymbol{\theta}_i^T, \ \sum_i \mathbf{z}_i \boldsymbol{\theta}_i^T. \tag{11}$$

As noted by Meng and Schilling (1996), it is preferable to condition on sufficient statistics to take advantage of Rao Blackwellization. They showed (p. 1258) the conditional statistics depend on two sufficient statistics of the missing data:

$$\begin{aligned} \mathbf{S}_1 &= \sum\nolimits_i \mathbf{z}_i \\ \mathbf{S}_2 &= \sum\nolimits_i \mathbf{z}_i \mathbf{z}_i^T. \end{aligned} \tag{12}$$

This suggests a simple strategy based wholly on (12):

$$\mathbf{b} = n^{-1}\left(-\mathbf{S}_1\right) = -\overline{\mathbf{z}}, \text{ given } \mathrm{E}\left[\sum\nolimits_i \boldsymbol{\theta}_i\right] = 0 \tag{13}$$

$$\begin{aligned} \mathbf{S}_2 &= \mathbf{VDV}^T + \mathbf{U} \\ \mathbf{A} &= \mathbf{VD}^{1/2}, \end{aligned} \tag{14}$$

where $\mathbf{U}$ is the identity matrix containing the variances of the latent measurement errors ε. While the solution in (9) and (10) requires further identification restrictions, e.g. the lower triangular restriction of Anderson and Rubin (1956), the eigenanalysis of $\mathbf{S}_2 - \mathbf{U}$ appears to be adequate for the identification of $\mathbf{A}$ as shown in the simulations below. For observed variables note that there are sufficient statistics serving the equivalent purpose of $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{U}$ for maximum likelihood factor analysis.

SAEM Algorithm for Exploratory MIRT Factor Analysis

Building on previous work (Bock & Aitken, 1981; Takane & de Leeuw, 1986; Albert, 1992; Meng & Schilling, 1996; Fox, 2003; Cai, 2010) the estimation procedure in this paper applies the SAEM method (Delyon et al., 1999; Kuhn & Lavielle, 2005). The algorithm below is given for polytomous items of which the procedure for dichotomous items is a special case. Assume κ ordinal categories of an observed item response. The latent regression model is given by

$$\begin{aligned} x_{ijk} &= A_j\theta_i - b_j - \tau_{jk} + \varepsilon_{ijk} \\ &= \eta_{ij} - \tau_{jk} + \varepsilon_{ijk}. \end{aligned} \tag{15}$$

where $k = 1, 2, ..., \kappa - 1$ indexes item thresholds.

1. *Draw missing values x for estimating item thresholds*.

Ordinal item option propensities x are drawn as

$$x_{ijk} \mid \eta_{ij}, \tau_{jk} \sim \begin{cases} N_{(-\infty,0)}\ \left(\eta_{ij} - \tau_{jk}, 1\right) & \text{if } y_{ij} \le k \\ N_{(0,+\infty)}\ \left(\eta_{ij} - \tau_{jk}, 1\right) & \text{if } y_{ij} > k. \end{cases} \tag{16}$$

Random values of x_{ijk} are independently generated for each individual for each item from the truncated normal distributions in (16). Item difficulties b_j and decentered thresholds τ_{jk} are obtained from sufficient statistics based on the *x* variables

$$\begin{aligned} b_j &= -\sum\nolimits_i \bar{x}_{ij\bullet} / n \\ \tau_{jk} &= -\sum\nolimits_i \left(x_{ijk} - b_j\right) / n. \end{aligned} \tag{17}$$

2. *Draw missing values z for estimating ability*.

For dichotomous items $\kappa = 2,$ and so $z_{ij} = x_{ij}$ because there is just one item propensity. For $\kappa > 2,$ draw *z* from the truncated normal distribution

`

$$z_{ij} \mid \eta_{ijk} \sim \begin{cases} TN_{\left(d_{jL},\infty\right)}\left(\eta'_{ij}, 1\right) & \text{if } y_{ij} = L = \kappa - 1 \\ \quad ... & \\ TN_{\left(d_{j1},d_{j2}\right)}\left(\eta'_{ij}, 1\right) & \text{if } y_{ij} = 2 \\ TN_{\left(-\infty,d_{j1}\right)}\left(\eta'_{ij}, 1\right) & \text{if } y_{ij} = 1 \end{cases} \tag{18}$$

where $d_{jk} = b_j + \tau_{jk}$ and $\eta'_{ij} = \mathbf{A}_j \boldsymbol{\theta}_i.$

3. *Update sufficient statistics.*

$$\boldsymbol{\mu}_{x,jk} = \boldsymbol{\tau}_{jk}^{(t-1)} + \gamma_t \left\{ \overline{\mathbf{x}}_{jk} - \boldsymbol{\mu}_{x,jk}^{(t-1)} \right\}$$

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{\Sigma}_{zz}^{(t-1)} + \gamma_t \left(\mathbf{S}_{zz} - \boldsymbol{\Sigma}_{zz}^{(t-1)} \right)$$

where γ_t is the current value of the Robbins-Monro gain coefficient. The second equation specifically refers to the update to the augmented data covariance $\boldsymbol{\Sigma}_{zz}^{(t)} = \mathbf{S}_{zz} = \mathbf{z}\mathbf{z}^T$.

4. *Obtain* $\mathbf{A} = \mathbf{V}\mathbf{D}^{1/2}$, $\mathbf{b}$, *and item thresholds from* $\boldsymbol{\Sigma}_{zz}$ *and* $\boldsymbol{\mu}_{x,jk}$.
5. *Sample* $\boldsymbol{\theta}$.
6. *Repeat Steps 1-5 until convergence*.

In Step 2, note that convergence is defined relative to the sufficient statistics rather than values of fixed item parameters. Due to rotational indeterminacy individual fixed parameters such as factor loadings should not be monitored (except in simulations) during iterations.

To monitor convergence of parameters, a window size (say *W* = 3) is selected along with a tolerance constant ε. Iterations are terminated when the maximum covariance change for the trace of $\boldsymbol{\Sigma}_{zz}$ is less than $\varepsilon = .001$ for *W* iterations. See Houts and Cai (2017) for an example of this convergence strategy with respect to individual parameters during MH-RM iterations. Robbins-Monro iterations require a defined step size; let $\gamma_t = (1/t)^{\alpha}$, $t \geq 0$, where $1/2 < \alpha \leq 1$ ensures convergence (Delyon et al., 1999). A larger step size (say $\alpha = 1$) may accelerate the rate of convergence, but may increase Monte Carlo error or result in a local maximum. A smaller step size $\alpha < 1$ may result in less Monte Carlo error but also slows convergence. Jank (2006) provided a set of recommendations to monitor convergence to a global solution. Moreover, examination of the eigenvalues of $\boldsymbol{\Sigma}_{zz}$ during iterations can help to identify the dimensionality *K* of the factor solution. After convergence, the solution can be rotated to obtain a simple structure. While rotation for interpretation can be a challenging task, this a different problem from estimation accuracy; the target rotation for the simulations will be the simulated loadings.

In the following three sections two simulation studies and an empirical example are given for demonstrating the new algorithm. In the two simulations, parameter estimates and CPU times are compared to those obtained with the MH-RM algorithm as implemented in the software package flexMIRT (Houts & Cai, 2017). As noted below the comparison of two algorithms must consider that a number of parameters beyond the algorithm per se may affect the speed and quality of estimation. This paper addresses the topic of computational architecture in the discussion section.

For both simulations abilities of the examinees were sampled from the multivariate normal prior of $\boldsymbol{\theta}$ of 5 and 10 dimensions (*K* = 5 and *K* = 10). In the 5D simulations the loadings from the first dimension are sampled from a beta B(2.5,3,0.2,1.7) while the items with a bifactor structure load on a second latent factor sampled from a more constrained B(2.5,3,0.1,0.9); this bifactor structure is devised to assess parameter estimation of smaller absolute values of loadings in the orthogonal dimension. In the 10D simulations the loadings from each dimension are sampled from B(2.5,3,0.2,1.7). All item difficulties for both simulations were sampled from $N(0,1)$.

Simulation Study with Bifactor Structure

The first simulation was run for *K* = 5 dimensional data following a bifactor structure. Simulated 5-category item responses were generated for 100 items with the cumulative normal response model for a sample size of *n* = 10,000. The first 20 items loaded only in the general factor, and each ensuing set of 20 items loaded on the general factor and one specific orthogonal factor. Both the SAEM and MH-RM algorithms were applied to this data set. Default options were used for the MH-RM algorithm, except that CPU time for standard error computation was minimized by specifying 10 Stage 4 iterations (settings specific to flexMIRT). Both analyses

(SAEM and MH-RM) were run on a desktop computer with 6GB of memory using an Intel i7 processor with four cores running at 3.33 GHz. For both analyses, 10 processors were used.[1] However, CPU time does not decrease linearly with the number of cores, and coding efficiencies were primarily attained through vectorization.[2] The SAEM algorithm was written entirely in R (R Core Team, 2017).

Commented [A1]: This is impossible on a 4-core i7. You can only use 8 threads.

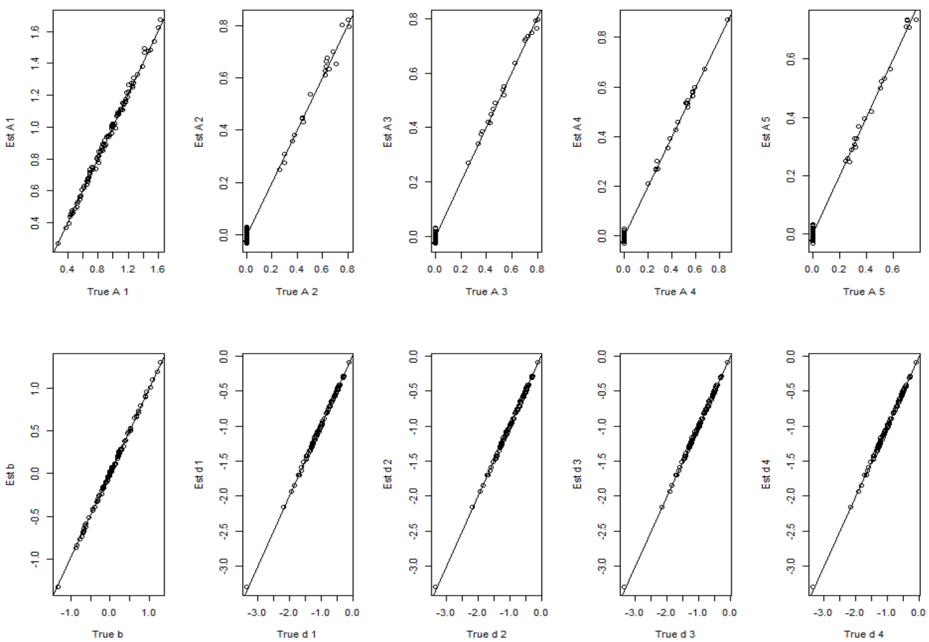

Figure 1. SAEM plots for bifactor loadings and thresholds.

The SAEM algorithm required 14.3 minutes of CPU time (226 iterations, and 10 burn in). Estimated loadings are plotted against true values in Figure 1, where unbiased coefficients fall on the 45-degree diagonal in the plots. Estimated factor loadings were obtained by rotating the **A** matrix to the matrix of true loadings using the TargetT function of the R package GPArotation (Bernaards & Jennrich, 2005). This rotation can be applied in each cycle to monitor convergence;

Commented [A2]: This seems like an incredibly small number of burn-in iterations… ?

however, this monitoring had no effect on estimation. In non-simulated data a rotation to simple structure obtained after convergence would be interpreted, recognizing that rotation may involve both substantive and statistical reasoning. Within the SAEM algorithm the monitoring functions (rotations) were included in CPU time resulting in a conservative benchmark. In Figure 1 it can be seen that the factor structure is accurately estimated. A moderate degree of bias is evident for some factor loadings, but all estimates of thresholds deviated negligibly from the 45 degree line.

The MH-RM algorithm as implemented by Cai, du Toit, and Thissen (2011) was also used to estimate the factor structure using standard options and applying a multidimensional graded response model.[3] This analysis required 28.6 minutes of CPU time (200, 100, and 2011 cycles for Stages 1, 2, and 3, respectively, with 10 Stage 4 cycles). After coefficients were estimated and scaled by dividing by the constant $D = 1.749$,[4] the factor loading matrix was rotated to the true coefficient matrix (unnormalized) target.[5] In Figure 2 the rotated factor loadings are plotted against true values.

In Figure 2 it can be seen that the loadings generally fall on or near the 45-degree line, but several deviations are observable for the first two factors. This may be a function of the starting values (obtained in Stages 1 and 2) for the MH-RM procedure (Stage 3), but it would take additional CPU time to obtain a better start. Of course, it is possible that the MH-RM algorithm could be tailored to reduce CPU time. This is also true for the SAEM algorithm for which starting values were generated randomly.

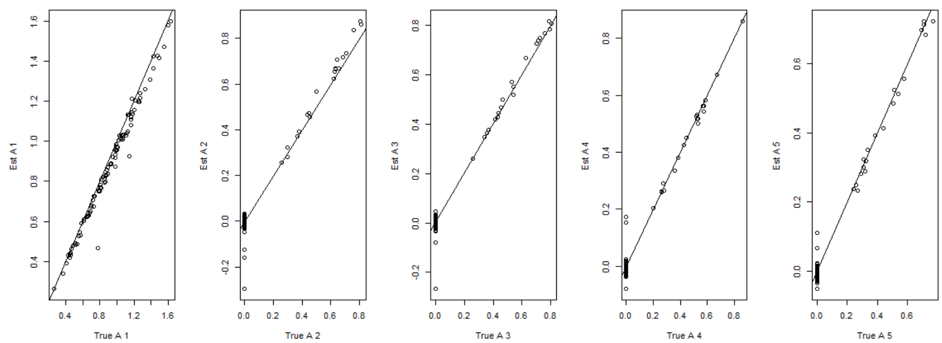

Figure 2. MH-RM plots for bifactor loadings only.

Simulation Study with Subscale Structure

The second simulation was run for K = 10 dimensional item responses following a subscore structure. The first 10 items loaded on the first subscale, the second 10 items loaded on the second subscale, and so forth. Simulated 4-category responses were generated for 100 items with the cumulative normal response model for a sample size of N = 100,000. Both the SAEM and MH-RM algorithms were applied to this data set. Default options were again used for the MH-RM algorithm, as described above. Both analyses were run on a desktop computer with 64GB of memory using an Intel i7-6950X processor with ten cores running at 3.00 GHz. Ten processors were used for both analyses. The effects of using multiple processors was not investigated.

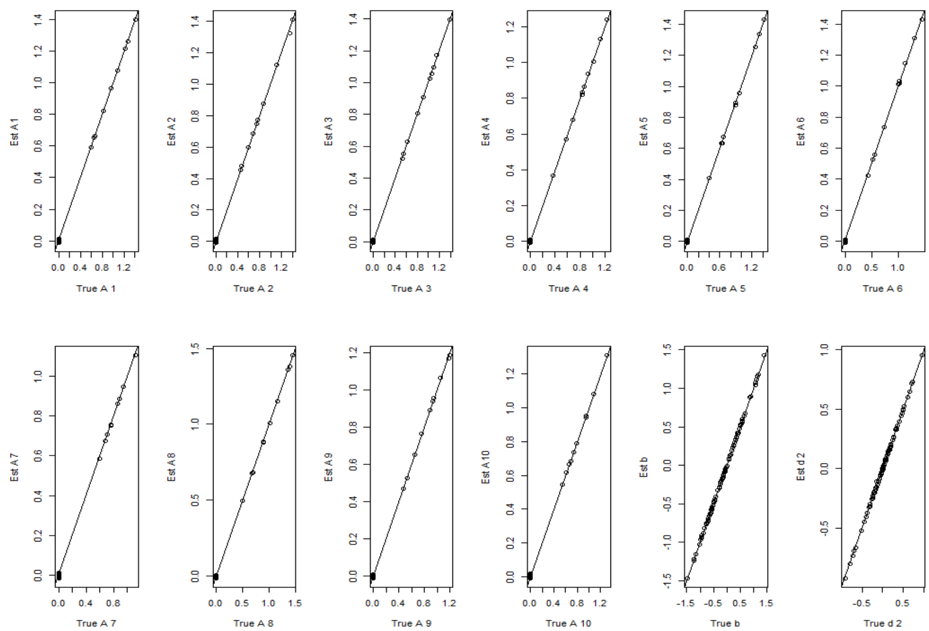

Figure 3. SAEM estimates plotted against true values (only second threshold shown).

The SAEM algorithm required 20.5 minutes of CPU time (75 iterations, and 50 burn in). Estimated loadings are plotted against true values in Figure 3. Unbiased coefficients fall on the 45-degree diagonal. Estimated factor loadings were obtained by rotating the **A** matrix to the matrix of true loadings within each cycle to monitor convergence. The factor structure is accurately estimated. A small degree of bias is evident for two factor loadings (factor 7 and 8), but all estimates touch the 45-degree line. Thresholds converged much more quickly than factor loadings.

The MH-RM procedure as described above was also used to estimate the factor structure. This required 70.6 minutes of CPU time (200, 100, and 742 cycles for Stages 1, 2, and 3, respectively, with 10 Stage 4 cycles). Upon convergence estimated loadings were rotated to the true coefficients. In Figure 4 the rotated factor loadings are plotted against true values.

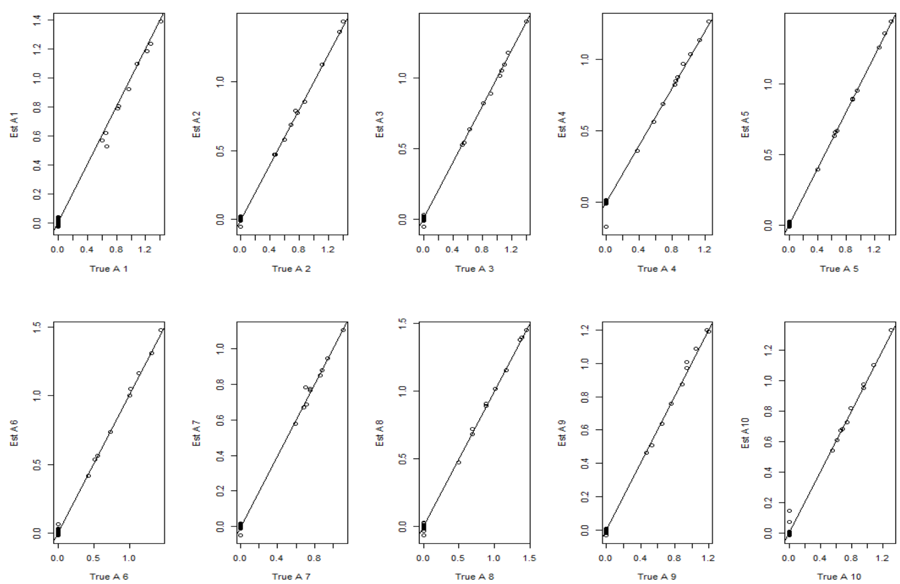

Figure 4. MH-RM estimates plotted against true values for factor loadings only.

In Figure 4 estimated coefficients generally track true values closely, but notably more deviations for the 45-degree line occur for the MH-RM solution for factors 1, 4, 7, and 9 than for the SAEM solution. While these deviations are small, it should be recognized that they are estimated with a very large sample size of N = 100,000.

Empirical Example: Quality of Life Data

In this section, the SAEM method is applied to data for a tryout form of 24 polytomous items (five categories) assessing the social quality of life for the Pediatric Quality of Life scale. These data for 753 children were previous analyzed by DeWalt, Thissen, Stucky et al. (2013) and

Cai (2010). Cai obtained solutions for 1-5 dimensions and compared two methods (EM with adaptive quadrature, MH-RM) in terms of factor structure and computation time. While the methods provided similar factor structure for a $K = 5$ factor solution, the EM algorithm required over 50 times more CPU time than MH-RM which required 1.58 minutes. For each of the 24 items four thresholds were estimated along with factor coefficients including 24*4 = 96 intercept parameters were estimated along with 24*K factor loadings.

The SAEM algorithm with 303 iterations was applied to the same data set, using the trace convergence criterion. The 3D solution (1.12 minutes CPU time) was found to be highly interpretable. In Table 1 loadings are given for the unstandardized loadings rotated to simple structure by the oblimin transformation. Note all items are coded with the same polarity. All items loading on Factor 1 involve negative affect whether this is regarding other kids, friends, or adults. This factor is primarily marked by two items: "Other kids were mean to me," and "Other kids made fun of me." In contrast, the highest loadings on Factor 2 involve positive affect. Finally, Factor 3 primarily involved three items regarding family. Factors 1 and 2 were correlated (after reversing polarity of negative affect items) about $r = .5$, but nearly uncorrelated with Factor 3. Item 16 is of some interest in itself: "My teachers understood me." Its loadings are relatively small and split across Factors 2 and 3. This perhaps demonstrates the dual role of teachers as caregivers for their students, but it also suggests that another dimension may be required for assessing this function.

Distinct factors for items having positive and negative valence is a common finding, especially with Likert-type items (e.g., Tomas & Oliver, 1999; Alexandrov, 2010). As noted by Alexandrov (2010) items with this dual structure can change the dimensionality of the construct and are often considered as a method effect; the exclusive use of positively worded Likert items with a fairly high level of intensity is thus recommended. In sum, this analysis illustrates the role of a purely exploratory approach. It is somewhat different from the example of Cai (2010) who presented a factor solution rotated to a partially specified structure.

Item	Factor 1	Factor 2	Factor3	Item Description
1	.51 (.05)	.37 (.05)	-.12 (.07)	I wished I had more friends.
2	-.22 (.06)	.17 (.04)	.48 (.12)	I want to spend more time with my family.
3	.01 (.04)	.45 (.05)	.13 (.06)	I was good at talking with adults.
4	.22 (.06)	.88 (.09)	.09 (.04)	I was good at making friends.
5	.61 (.05)	.22 (.04)	.01 (.06)	I have trouble getting along with other kids my age.
6	.81 (.09)	-.22 (.05)	.79 (.15)	I had trouble getting along with my family.
7	1.49 (.16)	.19 (.07)	-.29 (.13)	Other kids were mean to me.
8	.64 (.04)	-.09 (.03)	.22 (.06)	I got into a yelling fight with other kids.
9	.14 (.05)	1.03 (.08)	-.05 (.05)	I felt accepted by other kids my age.
10	-.13 (.06)	.73 (.10)	.73 (.17)	I felt loved by my parents or guardians.
11	.72 (.07)	.11 (.05)	-.12 (.07)	I was afraid of other kids my age.
12	1.49 (.13)	.02 (.05)	-.18 (.10)	Other kids made fun of me.
13	1.00 (.08)	.16 (.04)	.00 (.05)	I felt different from other kids my age.
14	.48 (.05)	.10 (.04)	-.06 (.06)	I got along better with adults than other kids my age.
15	.16 (.05)	1.02 (.07)	-.03 (.06)	I felt comfortable with other kids my age.
16	.15 (.05)	.39 (.05)	.31 (.09)	My teachers understood me.
17	.63 (.10)	-.23 (.05)	.93 (.20)	I had problems getting along with my parents or guardians.
18	.27 (.06)	.98 (.09)	.27 (.09)	I felt good about how I got along with my classmates.
19	.65 (.06)	.12 (.04)	.19 (.07)	I felt bad about how I got along with my friends.
20	.74 (.07)	.24 (.05)	-.03 (.07)	I felt nervous when I was with other kids my age.
21	.62 (.07)	.31 (.05)	.06 (.04)	I did not want to be with other kids.
22	.06 (.04)	1.10 (.07)	-.03 (.05)	I could talk with my friends.
23	.13 (.05)	.99 (.08)	-.14 (.08)	Other kids wanted to be with me.
24	.04 (.03)	.98 (.07)	-.15 (.08)	I did things with other kids my age.

Table 1. Unstandardized factor loadings for the SAEM 3D solution for Social Quality of Life prototype scale. Standard errors in parentheses.

Let **A*** represent the value of the factor loading matrix upon convergence. Standard errors for QOL factor loadings were obtained by the following procedure. A chain of length 1000 for each coefficient in **A** was generated post-convergence with $\gamma = 1$. On each cycle, each sampled value of **A** was rotated to the target **A*** and saved. Standard errors [SE] were then obtained with the R package mcmcse (Dai & Jones, 2017; Flegal, 2012) using the Tukey method, which provided slightly higher SEs than other methods. In Table 1 SEs on the third factor are larger for higher loadings; however, the absolute values of the largest loadings still exceed 4 SEs.

Discussion

In this study, the new SAEM algorithm for factor analysis was shown to provide nearly unbiased estimation of item parameters in two large-scale simulation studies. The speed of the SAEM procedure was also compared to the MH-RM procedure (Cai, du Toit, & Thissen, 2011), using the work of Cai (2010). The SAEM algorithm investigated was demonstrated to be faster for exploratory factor analysis of categorical response variables in larger data sets, but there are several limitations to generalizing these results. Depending on factors such as setting the criterion and tolerance for convergence, window size, starting values, gain sequences, and programming language, CPU time can change drastically. Moreover, to examine bias in simulation studies, factor loadings need to be rotated to a target structure. The method of rotation may also affect the results, though the same method of rotation was used for the SAEM and MH-RM solutions. Taking all of this into consideration, the new algorithm appears to have much potential and its implementation in a compiled language such as C++ may further decrease CPU time. If the SAEM algorithm is modified to bypass estimates of thresholds falling below the convergence, this will provide a substantial speed-up because most CPU time in the SAEM algorithm is spent on generating missing values for estimating thresholds. The most obvious advantage of the new algorithm is the wide access it provides to IRT factor analysis of high dimensional and/or large data sets. Missing data in the form of incomplete response vectors do not pose a serious obstacle. Under the assumption of *missing at random* (Rubin & Little, 2002), missing value imputation is naturally integrated into the algorithm.

Because of simple and short code (no matrix inversion, no derivatives), this innovation can be easily incorporated into research procedures involving factor analysis of categorical data. Necessary future work involves the computation of standard errors, which can be more expensive in terms of CPU time than the estimation of parameters in large-scale analyses. Several other extensions to the proposed algorithm would be useful for applying this new approach to estimation of a greater variety of MIRT models. First, development is required to extend the approach to IRT models with guessing, though inclusion of a *c* parameter would not bypass stability problems encountered in practice with lower asymptotes. Second, the procedure may be amenable to other methods of setting identification restrictions on **A**, suggesting its potential application to confirmatory factor analysis.

End Notes

[1] The item parameters and data set used for the simulations are available upon request from the second author.

[2] A lot has changed in statistical computing since 1996, both in terms of software capacity and hardware. Meng and Schilling reported using 6 minutes of CPU time for 20 iterations involving an analysis of 1000 response patterns for a 32-item test with dichotomous items. In the empirical example in this paper, about 300 iterations for a 24-item test of polytomous items with $n = 753$ were run in just over 1 minute.

[3] A logistic graded response model is more similar to a cumulative normal model than a generalized partial credit model, and thus a fairer basis for comparison,

[4] Savalei (2006) showed that $D = 1.749$ provides a slightly better scaling than the traditional $D = 1.7$.

[5] The same rotation procedure was used for both the SAEM and MH-RM procedures. If the estimated **A** was first rotated to an orthogonal solution using the varimax transformation, the target rotation performed more effectively.

References

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

Anderson, T.W. & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (ed.) *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (pp. 111–150), Berkeley: University of California Press.

Alexandrov, A. (2010). Characteristics of single-item measures in Likert scale format. *Electronic Journal of Business Research Methods,* 8 (1), 1-12.

Béguin, A. A. & Glas, C.A.W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66, 541-562.

Bernaards, Coen A. and Jennrich, Robert I. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*, 65, 676-696.

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

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, 75, 33-57.

Cai, L., du Toit, S.H.C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software]. Chicago: SSI International.

Celeux, G. & Diebolt, J. (1992). A stochastic approximation type EM algorithm for the mixture problem, *Stochastics and Stochastics Reports*, 41, 119-134.

Dai, N. & Jones, G.L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *Journal of Multivariate Analysis*, 159, 184-199.

Delyon, B., Lavielle, M., & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, 27, 94-128.

Dempster, A.P. Laird, N.M. Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39, 1–38.

Flegal, J.M. (2012). Applicability of subsampling bootstrap methods in Markov chain Monte Carlo. In L. Plaskota & H. Woźniakowski (Eds.) *Monte Carlo and Quasi-Monte Carlo Methods 2010* (pp. 363-372), *Springer Proceedings in Mathematics & Statistics*, 23, 363-372.

Fox, J.-P. (2003) Stochastic EM for estimating the parameters of a multilevel IRT model. *British Journal of Mathematical and Statistical Psychology*, 56, 65-81.

Fox, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. New York: Springer.

Gu, M.-G. & Kong, F.-H. (1998). A stochastic approximation algorithm with Markov chain Monte-Carlo method for incomplete data estimation problems. *Proceedings of the National Academy of Science*, 95, 7270-7274.

Hogg, R.V., McKean, J., & Craig, A.T. (2014) *Introduction to mathematical statistics*, *7th ed*. Boston, MA: Pearson Education.

Houts, C. R., & Cai, L. (2017). *flexMIRT users manual version 3.5: Flexible multilevel item factor analysis and test scoring*. Chapel Hill, NC: Vector Psychometric Group.

Jank, W. (2006). Implementing and Diagnosing the Stochastic Approximation EM Algorithm. *Journal of Computational and Graphical Statistics*, 4, 803-829.

Kuhn E. & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, 49, 1020–1038.

Meng, X.L. & 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.

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

Robbins, H. & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, 22, 400–407.

Rubin, D.B., Thayer, D.T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, 47, 69–76.

Savalei, V. (2006). Logistic approximation to the normal: The KL rationale. *Psychometrika*, 71, 763-767.

Takane, Y., and de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 393-408.

Tanner, M.A. & Wong, W.H. (1987). The Calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.

Tomas, J.M. & Oliver, A. (1999). Rosenberg’s Self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, 6, 84-98.