ACT Research Report Series 96-6

Estimation of Item Response Models Using the EM Algorithm for Finite Mixtures

David J. Woodruff

Bradley A. Hanson

ACT

September 1996

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243-0168

Estimation for Item Response Models using the EM Algorithm for Finite Mixtures

David Woodruff
Bradley A. Hanson

Abstract

This paper presents a detailed description of maximum likelihood parameter estimation for item response models using the general EM algorithm. In this paper the models are specified using a univariate discrete latent ability variable. When the latent ability variable is discrete the distribution of the observed item responses is a finite mixture, and the EM algorithm for finite mixtures can be used. Maximum likelihood estimates of the item parameters and of the discrete probabilities of the latent ability distribution are given using the EM algorithm for finite mixtures. Results are presented in general for both dichotomous and polytomous item response models. The relation between the EM estimates and Bock-Aitken marginal maximum likelihood estimates is discussed.

Estimation for Item Response Models using the EM Algorithm for Finite Mixtures

The purpose of this paper is to present a fairly simple and unified treatment of how the general EM algorithm can be used to obtain maximum likelihood estimates (MLEs) of both the item parameters and the probability distribution of the latent ability variable for item response models. The approach taken in this paper is to assume the latent ability variable being measured by the items is discrete. When the latent ability variable is discrete the distribution of the observed data is a finite mixture (Titterington, Smith, and Makov, 1985). With a discrete latent ability variable the EM algorithm for finding maximum likelihood estimates for finite mixtures can be used (Dempster, Laird, and Rubin, 1977; Titterington, Smith, and Makov, 1985).

This paper clarifies previously established results using a finite mixture approach. A complete, self-contained description of maximum likelihood parameter estimates of item response models for dichotomous and polytomous items using the EM algorithm for finite mixtures is presented. The use of the finite mixture model allows a variety of previously disparate results to be consolidated using a single relatively simple approach that allows a straight-forward presentation with pedagogic value.

Versions of the results in this paper have been presented by others for a variety of specific item response models. Maximum likelihood estimates of item parameters using the EM algorithm have been presented for a variety of item response models for dichotomous items (Bock and Aitken, 1981; Thissen, 1982; Rigdon and Tsutakawa, 1983; Tsutakawa, 1984; Bartholomew, 1987; Harwell, Baker, and Zwarts, 1988; Baker, 1992) and polytomous items (Thissen and Steinberg, 1984; Bartholomew, 1987; Muraki, 1992; Wilson and Adams, 1993). The EM algorithm for finite mixtures has been applied in estimating parameters for the Rasch model by De Leeuw & Verhelst (1986) and Follmann (1988). The maximum likelihood estimates of the probabilities of the discrete latent ability distribution presented here were given by Bock and Aitken (1981), Mislevy (1984), and Titterington, Smith, and Makov (1985).

The data to be modeled are the responses of $i = 1, \ldots, N$ examinees, randomly sampled from a population of examinees, to a fixed non-random set of $j = 1, \ldots, n$ items. The responses of the N examinees to the n items are contained in a $n \times N$ matrix $\mathbf{Y}$ made up of $n \times 1$ column vectors $\mathbf{y}_1, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_N$ that contain the responses of the ith randomly sampled examinee to the n fixed items. The matrix $\mathbf{Y}$ is given by

$$\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_N]. \tag{1}$$

The jth element of $\mathbf{y}_i$ (the response of the ith randomly sampled examinee to item j) is denoted y_{ij}. It is assumed that the set of responses to each item is finite. If the responses are dichotomous then the possible values of y_{ij} are 0 and 1. If the responses are polytomous then the possible values of y_{ij} are taken to be the integers $0, 1, \ldots, L_j - 1$ (item j has L_j response categories). In practical applications values of the polytomous items need not be integers or even ordered. Note that different items may have different numbers of response categories.

Associated with item j is a set of ν_j item parameters denoted by the $\nu_j \times 1$ column vector, $\boldsymbol{\delta}_j$. The parameters for all n items are represented by $\boldsymbol{\Delta}$, the collection of all $\boldsymbol{\delta}_j$ column vectors, that is $\boldsymbol{\Delta} = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_j, \ldots, \boldsymbol{\delta}_n]$. When the number of item response categories is the same for every item (e.g., dichotomous items) then the number of parameters will typically be the same for every item so that $\nu_j = \nu$ for all j.

In addition to the observed item responses, there is a realization of a latent ability random variable Θ for each randomly sampled examinee. Unlike the realization of the item responses, the realization of Θ for the ith randomly sampled examinee (denoted θ_i) is not observed. The value θ_i is sometimes referred to as the "ability" of the ith randomly sampled examinee. In this paper the term "latent variable" will typically be used in place of "ability."

The latent random variable Θ is usually considered to be continuous. In this paper the latent variable is taken to be discrete, and estimation procedures are derived based on the discrete latent variable. This is opposed to deriving estimation procedures based on a continuous latent variable and then implementing approximations of those procedures with a discrete version of the continuous latent variable (e.g., Bock and Aitken, 1981; Muraki, 1992). In this paper the approximation of the continuous latent variable with a discrete latent variable is done in the model specification. This allows straightforward application of the EM algorithm for finite mixtures.

The latent random variable Θ can take on m known discrete values $\theta_k, k = 1, \ldots, m$, with associated unknown probabilities $\pi_k, k = 1, \ldots, m$ (a short discussion of choosing the value of m is given in the Discussion section at the end of the paper). The values of θ_k are chosen at the initiation of the estimation process and determine the scale of the latent variable. Typically the scale of the latent variable can only be known up to a linear transformation, so the model is invariant to a linear transformation of the θ_k (along with an associated transformation of the item parameters). The $m \times 1$ column vector of latent probabilities is given by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)^t$. The random variable Θ has a probability distribution defined over the population of examinees $[\Pr(\Theta = \theta_k \mid \boldsymbol{\pi})]$ that can be denoted variously as

$$\begin{aligned} \Pr(\Theta = \theta_k \mid \boldsymbol{\pi}) &= \Pr(\Theta = \theta_k \mid \pi_k) \\ &= p(\theta_k \mid \boldsymbol{\pi}) = p(\theta_k \mid \pi_k) = \pi_k . \end{aligned} \tag{2}$$

In this paper the latent random variable Θ is taken to be univariate. It is possible to generalize the formulas presented to the case of a multivariate Θ. A multivariate Θ would greatly increase the computational effort required to compute estimates.

The EM Algorithm for Finite Mixtures

Let $f(\mathbf{y} \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ be the probability distribution for the observed item responses ($\mathbf{y}$ is a vector of realizations of the item response random variables). When the latent variable is discrete $f(\mathbf{y} \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ is given by

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\Delta}, \boldsymbol{\pi}) &= \sum_{k=1}^{m} f(\mathbf{y}, \theta_k \mid \boldsymbol{\Delta}, \boldsymbol{\pi}) \\ &= \sum_{k=1}^{m} f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta}, \boldsymbol{\pi}) p(\theta_k \mid \boldsymbol{\Delta}, \boldsymbol{\pi}) \\ &= \sum_{k=1}^{m} f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta}) p(\theta_k \mid \boldsymbol{\pi}) \\ &= \sum_{k=1}^{m} f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta}) \pi_k , \end{aligned} \tag{3}$$

where $f(\mathbf{y}, \theta \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ is the joint probability distribution of the item responses and the latent variable, and $f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta})$ is the conditional probability distribution of the item responses for examinees with a fixed value of the latent variable θ_k. The third and fourth lines of Equation 3 are obtained by using the equalities

$$f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta}, \boldsymbol{\pi}) = f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta}), \tag{4}$$

and

$$p(\theta_k \mid \boldsymbol{\Delta}, \boldsymbol{\pi}) = p(\theta_k \mid \boldsymbol{\pi}) = \pi_k . \tag{5}$$

Equation 4 follows from the assumption made in item response models that conditioned on the latent variable Θ the probability of an examinee's responses to n items does not depend on the probability distribution of the latent variable in the population of examinees. Equation 5 follows from the fact that the probability distribution of Θ in the population of examinees does not depend on the item parameters for the n items.

The expression for $f(\mathbf{y} \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ in Equation 3 is a finite mixture (Titterington, Smith, and Makov, 1985). That is, from the last line in Equation 3 it can be seen that $f(\mathbf{y} \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ is a sum of component densities $f(\mathbf{y} \mid \theta_k, \boldsymbol{\Delta})$ with associated mixing weights π_k.

The EM algorithm for finding maximum likelihood estimates of the parameters of a finite mixture is described by Dempster, Laird, and Rubin (1977, section 4.3), and Titterington, Smith, and Makov (1985, section 4.3.2). The presentation of the EM algorithm for finite mixtures in this paper uses somewhat different notation than that used in those presentations. Dempster, Laird, and Rubin (1977) and Titterington, Smith, and Makov (1985) use an indicator vector $\mathbf{z}_i$ in place of θ_i, where $\mathbf{z}_i$ is of length m with a one in the position indicating the category of the latent variable for examinee i and zeros elsewhere. The present notation is more consistent with notation used in the psychometric literature.

The observed data are $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$, the missing data are $(\theta_1, \ldots, \theta_N)$, and the complete data is $[(\mathbf{y}_1, \theta_1), \ldots, (\mathbf{y}_N, \theta_N)]$. The complete data likelihood for the sample is

$$\prod_{i=1}^{N} f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi}), \tag{6}$$

where $f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ is the complete data likelihood for examinee i. The observed data likelihood for the sample is

$$\prod_{i=1}^{N} f(\mathbf{y}_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi}) = \prod_{i=1}^{N} \sum_{k=1}^{m} f(\mathbf{y}_i, \theta_k \mid \boldsymbol{\Delta}, \boldsymbol{\pi}), \tag{7}$$

where $f(\mathbf{y}_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi})$ is the observed data likelihood for examinee i. The EM algorithm uses the complete data likelihood to find values of the parameters $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ which maximize the observed data likelihood (Dempster, Laird, and Rubin, 1977).

The general EM algorithm generates a sequence of estimates $(\boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}), s = 1, 2, \ldots$, given starting values $(\boldsymbol{\Delta}^{(0)}, \boldsymbol{\pi}^{(0)})$. There are two steps in each iteration: the E step and the M step. In the E step the conditional expectation of the complete data log-likelihood is taken, where the conditional expectation is with respect to the conditional distribution of the missing data given the observed data and some fixed known values of the parameters. Let Θ_i be the random variable representing the latent variable for examinee i ($\Theta_i, i = 1, \ldots, N$ are independent and identically distributed), and let $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_N)$. The conditional expectation evaluated at the E step for iteration $s, s = 0, 1, \ldots$, is

$$Q[(\boldsymbol{\Delta}, \boldsymbol{\pi}) \mid (\boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})] = E_{\boldsymbol{\Theta}}\left\{ \log \left[\prod_{i=1}^{N} f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi}) \right] \mid \mathbf{Y}, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)} \right\}, \tag{8}$$

where the expected value is over the conditional distribution of the missing data given the observed data and fixed known values of the parameters $(\boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$. Equation 8 is the expression used in the E step of the general EM algorithm, as it is the expectation of the complete data log likelihood. Complete data sufficient statistics (other than the observations themselves) are not used.

The M step finds values of $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ that maximize the conditional expectation of the complete data log-likelihood. The M step at iteration $s, s = 0, 1, \ldots$, finds $(\boldsymbol{\Delta}, \boldsymbol{\pi}) = (\boldsymbol{\Delta}^{(s+1)}, \boldsymbol{\pi}^{(s+1)})$ to maximize $Q[(\boldsymbol{\Delta}, \boldsymbol{\pi}) \mid (\boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})]$. The new estimates $(\boldsymbol{\Delta}^{(s+1)}, \boldsymbol{\pi}^{(s+1)})$ produced in the M step at iteration s are used in the E step at iteration $s + 1$. Iterations continue until convergence is obtained.

The expectation in Equation 8 can be written as

$$
\begin{aligned}
&E_{\Theta}\left\{\log\left[\prod_{i=1}^{N} f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi})\right] \mid \mathbf{Y}, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}\right\} \\
&\qquad = E_{\Theta}\left\{\sum_{i=1}^{N} \log[f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi})] \mid \mathbf{Y}, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}\right\} \\
&\qquad = \sum_{i=1}^{N} E_{\Theta_i}\{\log[f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi})] \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}\} \\
&\qquad = \sum_{i=1}^{N}\sum_{k=1}^{m}\{\log[f(\mathbf{y}_i, \theta_k \mid \boldsymbol{\Delta}, \boldsymbol{\pi})]\, p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \\
&\qquad = \sum_{k=1}^{m}\sum_{i=1}^{N}\{\log[f(\mathbf{y}_i, \theta_k \mid \boldsymbol{\Delta}, \boldsymbol{\pi})]\, p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\}, \qquad (9)
\end{aligned}
$$

where $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$ is the conditional probability that $\Theta_i = \theta_k$ given fixed known values $\mathbf{y}_i$, $\boldsymbol{\Delta}^{(s)}$, and $\boldsymbol{\pi}^{(s)}$. Note that $\log[f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi})]$ is simply treated as a function of the discrete random variable Θ_i with respect to which its expectation is being taken. That is why in the first three lines of Equation 9 the realization of the latent random variable Θ_i for the ith randomly sampled examinee is denoted θ_i, whereas in the last two lines, where the conditional expectation has been made explicit, the θ_i (the unknown realization of the latent variable for examinee i) are changed to θ_k (the known values of the latent variable that could be realized for an examinee) in accordance with the expectation over the discrete distribution of the latent random variable. It should be noted that $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ are free unknown quantities for which estimates are found in the M step, whereas $\boldsymbol{\Delta}^{(s)}$ and $\boldsymbol{\pi}^{(s)}$ are fixed known quantities that have been computed at step $s-1$.

Estimation of $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ can be simplified by separating Equation 9 into two additive terms with the first depending only on $\boldsymbol{\Delta}$ and the second depending only on $\boldsymbol{\pi}$. In this way the derivative of the first term can be taken with respect to $\boldsymbol{\Delta}$, the result set to zero and solved for $\boldsymbol{\Delta}$. Similarly, the derivative of the second term can be taken with respect to $\boldsymbol{\pi}$, the result set to zero and solved for $\boldsymbol{\pi}$. Consequently, M step estimates can be calculated separately for $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$. The estimates of $\boldsymbol{\pi}$ are easy to compute (a closed-form solution exists). The computation of estimates of $\boldsymbol{\Delta}$ will typically require iterative numerical methods.

To separate Equation 9 into one part depending only on $\mathbf{\Delta}$ and one part depending only on $\boldsymbol{\pi}$ note that using Equations 4 and 5 (substituting $\mathbf{y}_i$ for $\mathbf{y}$)

$$\begin{aligned} f(\mathbf{y}_i, \theta_k \mid \mathbf{\Delta}, \boldsymbol{\pi}) &= f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}, \boldsymbol{\pi}) p(\theta_k \mid \mathbf{\Delta}, \boldsymbol{\pi}) \\ &= f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}) \pi_k \,. \end{aligned} \tag{10}$$

Substituting Equation 10 into Equation 9 gives

$$\begin{aligned} &\sum_{k=1}^{m} \sum_{i=1}^{N} \{\log[f(\mathbf{y}_i, \theta_k \mid \mathbf{\Delta}, \boldsymbol{\pi})] p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \\ &\qquad = \sum_{k=1}^{m} \sum_{i=1}^{N} \{\log[f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}) \pi_k] p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \,. \end{aligned} \tag{11}$$

The right side of Equation 11 can be written as

$$\begin{aligned} &\sum_{k=1}^{m} \sum_{i=1}^{N} \{\log[f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta})] p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \\ &\qquad + \sum_{k=1}^{m} \sum_{i=1}^{N} \{\log(\pi_k) p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \,. \end{aligned} \tag{12}$$

The first term in Equation 12 involves only $\mathbf{\Delta}$, and the second term in Equation 12 involves only $\boldsymbol{\pi}$ (all other terms, including θ_k, are constants). The first term in Equation 12 will be denoted by

$$\phi(\mathbf{\Delta}) = \sum_{k=1}^{m} \sum_{i=1}^{N} \{\log[f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta})] p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \,, \tag{13}$$

and the second term in Equation 12 will be denoted

$$\psi(\boldsymbol{\pi}) = \sum_{k=1}^{m} \sum_{i=1}^{N} \{\log(\pi_k) p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \,. \tag{14}$$

In the E step at iteration $s, s = 0, 1, \ldots$, the mN (m categories of the latent variable for each of N examinees) conditional probabilities $p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$ are computed using the values of $\mathbf{\Delta}^{(s)}$ and $\boldsymbol{\pi}^{(s)}$ computed in the M step at iteration $s - 1$, or in the case of $s = 0$ the starting values. Note that for the first iteration (s=0) of the EM algorithm, the values of $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(0)}, \boldsymbol{\pi}^{(0)})$ can be specified directly instead of specifying a $\boldsymbol{\Delta}^{(0)}$ and $\boldsymbol{\pi}^{(0)}$ and computing $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(0)}, \boldsymbol{\pi}^{(0)})$. The values of $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$ are then substituted into Equations 13 and 14 at iteration s, and the M step at iteration s consists of two parts: (1) finding values of $\boldsymbol{\Delta}$ that maximizes $\phi(\boldsymbol{\Delta})$ [these will be $\boldsymbol{\Delta}^{(s+1)}$], and (2) finding values of $\boldsymbol{\pi}$ that maximize $\psi(\boldsymbol{\pi})$ [these will be $\boldsymbol{\pi}^{(s+1)}$].

Using the definition of conditional probability, the probabilities $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$ that are computed in the E step can be expressed as

$$\begin{aligned} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) &= \frac{f(\mathbf{y}_i, \theta_k \mid \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})}{f(\mathbf{y}_i \mid \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})} \\ &= \frac{f(\mathbf{y}_i, \theta_k \mid \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})}{\sum_{k'=1}^{m} f(\mathbf{y}_i, \theta_{k'} \mid \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})} \\ &= \frac{f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}^{(s)}, \pi_k^{(s)}) p(\theta_k \mid \boldsymbol{\Delta}^{(s)}, \pi_k^{(s)})}{\sum_{k'=1}^{m} f(\mathbf{y}_i \mid \theta_{k'}, \boldsymbol{\Delta}^{(s)}, \pi_{k'}^{(s)}) p(\theta_{k'} \mid \boldsymbol{\Delta}^{(s)}, \pi_{k'}^{(s)})} \\ &= \frac{f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}^{(s)}) \pi_k^{(s)}}{\sum_{k'=1}^{m} f(\mathbf{y}_i \mid \theta_{k'}, \boldsymbol{\Delta}^{(s)}) \pi_{k'}^{(s)}}, \end{aligned} \tag{15}$$

where $\pi_k^{(s)}$ is the kth element of $\boldsymbol{\pi}^{(s)}$. The subscript k' on the right side of Equation 15 is used in sums over all possible values of the latent variable, whereas the subscript k denotes a specific value of the latent variable. The final result in Equation 15 for $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$ is an application of Bayes Theorem. Equation 15 is used for the E step computation. Note that Equation 15 applies for any item response model.

Details of the EM algorithm for calculating estimates of the item parameters and latent variable distribution for item response models with dichotomous and polytomous items are presented in the following sections.

Computing Item Parameter Estimates for Dichotomous Items

This section presents details of the EM algorithm for computing item parameter estimates for dichotomous item response models (where $L_j = 2$ for all j). The two possible responses to each item are scored 0 (incorrect) and 1 (correct).

For dichotomous items, Equation 13 can be written in terms of item response functions for each item. The item response function for item j is a function of a value of the latent variable and the item parameters associated with item j. The item response function gives the probability of an examinee with latent variable value θ answering item j correctly, and will be denoted $P(\theta, \boldsymbol{\delta}_j)$. The probability of an examinee with a latent variable value of θ answering item j incorrectly is $Q(\theta, \boldsymbol{\delta}_j) = 1 - P(\theta, \boldsymbol{\delta}_j)$. Examples of item response functions are normal and logistic ogives (Lord, 1980).

For item j with item parameters $\boldsymbol{\delta}_j$ and randomly sampled examinee i with ability value θ_k the probability of item response y_{ij} is given by

$$f(y_{ij} \mid \theta_k, \boldsymbol{\delta}_j) = P(\theta_k, \boldsymbol{\delta}_j)^{y_{ij}} Q(\theta_k, \boldsymbol{\delta}_j)^{1-y_{ij}} . \tag{16}$$

It is assumed that conditioned on the value of the latent variable for the examinee, the examinee's responses to the items are mutually independent (this is the assumption of local independence). Under local independence $f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})$ can be written as

$$f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}) = \prod_{j=1}^{n} f(y_{ij} \mid \theta_k, \boldsymbol{\delta}_j) = \prod_{j=1}^{n} P(\theta_k, \boldsymbol{\delta}_j)^{y_{ij}} Q(\theta_k, \boldsymbol{\delta}_j)^{1-y_{ij}} . \tag{17}$$

Equation 13 can be written using Equation 17 as

$$\begin{aligned}
\phi(\boldsymbol{\Delta}) &= \sum_{k=1}^{m} \sum_{i=1}^{N} \left\{ \log \left[\prod_{j=1}^{n} f(y_{ij} \mid \theta_k, \boldsymbol{\delta}_j) \right] p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right\} \\
&= \sum_{k=1}^{m} \sum_{i=1}^{N} \sum_{j=1}^{n} \{ \log [P(\theta_k, \boldsymbol{\delta}_j)^{y_{ij}} Q(\theta_k, \boldsymbol{\delta}_j)^{1-y_{ij}}] p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \} \\
&= \sum_{k=1}^{m} \sum_{j=1}^{n} \left\{ \log [P(\theta_k, \boldsymbol{\delta}_j)] \left[\sum_{i=1}^{N} y_{ij} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right] \right\} \\
&\quad + \sum_{k=1}^{m} \sum_{j=1}^{n} \left\{ \log [Q(\theta_k, \boldsymbol{\delta}_j)] \left[\sum_{i=1}^{N} (1 - y_{ij}) p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right] \right\} \\
&= \sum_{k=1}^{m} \sum_{j=1}^{n} \left\{ \log [P(\theta_k, \boldsymbol{\delta}_j)] \left[\sum_{i=1}^{N} y_{ij} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right] \right\} \\
&\quad + \sum_{k=1}^{m} \sum_{j=1}^{n} \left\{ \log [Q(\theta_k, \boldsymbol{\delta}_j)] \left[\sum_{i=1}^{N} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right] \right\} \\
&\quad - \sum_{k=1}^{m} \sum_{j=1}^{n} \left\{ \log [Q(\theta_k, \boldsymbol{\delta}_j)] \left[\sum_{i=1}^{N} y_{ij} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right] \right\} .
\end{aligned} \tag{18}$$

A simpler computational formula for Equation 18 can be obtained by using Equation 15 to compute

$$n_k^{(s)} = \sum_{i=1}^{N} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) = \sum_{i=1}^{N} \frac{f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}^{(s)})\pi_k^{(s)}}{\sum_{k'=1}^{m} f(\mathbf{y}_i \mid \theta_{k'}, \boldsymbol{\Delta}^{(s)})\pi_{k'}^{(s)}}, \tag{19}$$

and

$$r_{jk}^{(s)} = \sum_{i=1}^{N} y_{ij} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) = \sum_{i=1}^{N} \frac{y_{ij} f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}^{(s)})\pi_k^{(s)}}{\sum_{k'=1}^{m} f(\mathbf{y}_i \mid \theta_{k'}, \boldsymbol{\Delta}^{(s)})\pi_{k'}^{(s)}}. \tag{20}$$

Substituting Equations 19 and 20 into Equation 18 gives

$$\phi(\boldsymbol{\Delta}) = \sum_{k=1}^{m} \sum_{j=1}^{n} \{\log[P(\theta_k, \boldsymbol{\delta}_j)] r_{jk}^{(s)} + \log[Q(\theta_k, \boldsymbol{\delta}_j)](n_k^{(s)} - r_{jk}^{(s)})\}. \tag{21}$$

The quantity $n_k^{(s)}$ can be thought of as a provisional estimate of the number of examinees in the sample with ability value θ_k. The quantity $r_{jk}^{(s)}$ can be thought of as a provisional estimate of the number of examinees in the sample with ability value θ_k who answer item j correctly. Note the notational distinction that though n denotes the number of items on the test, the $n_k^{(s)}$ represent estimates of the number of examinees with specified ability value θ_k at iteration s.

The E step at iteration s consists of computing the values $n_k^{(s)}$ and $r_{jk}^{(s)}$ using values of $\boldsymbol{\Delta}^{(s)}$ and $\pi_k^{(s)}$ computed in iteration $s-1$ ($\boldsymbol{\Delta}^{(0)}$ and $\pi_k^{(0)}$ are starting values used in iteration 0). In the M step at iteration s the values of $n_k^{(s)}$ and $r_{jk}^{(s)}$ computed in the E step are substituted into Equation 21 and the value of $\boldsymbol{\Delta}$, namely $\boldsymbol{\Delta}^{(s+1)}$, that maximizes $\phi(\boldsymbol{\Delta})$ is found. Maximization methods such as Newton-Raphson (Dennis and Schnabel, 1983) involve computing the first and second partial derivatives of $\phi(\boldsymbol{\Delta})$, which in turn involves computing the first and second derivatives of $\log[P(\theta_k, \boldsymbol{\delta}_j)]$ and $\log[Q(\theta_k, \boldsymbol{\delta}_j)]$ with respect to $\boldsymbol{\Delta}$. These derivatives can be quite complex depending on the form of $P(\theta_k, \boldsymbol{\delta}_j)$. Baker (1992) gives detailed derivations of these partial derivatives for various forms of $P(\theta_k, \boldsymbol{\delta}_j)$. The details of computing the maximum in the M step will not be presented in this paper.

Computing an Estimate of the Latent Variable Distribution

This section present details of the EM algorithm for computing an estimate of $\boldsymbol{\pi}$. The procedures presented in this section apply to any item response model for either dichotomous or polytomous items.

Equation 14 can be written as

$$\begin{aligned}\psi(\boldsymbol{\pi}) &= \sum_{k=1}^{m}\sum_{i=1}^{N}\{\log(\pi_k)p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\} \\ &= \sum_{k=1}^{m}\log(\pi_k)\sum_{i=1}^{N}p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})\,. \end{aligned} \tag{22}$$

The E step substitutes the values of $n_k^{(s)}$ from Equation 19 *into equation* 22 to obtain

$$\psi(\boldsymbol{\pi}) = \sum_{k=1}^{m}\log(\pi_k)n_k^{(s)}\,. \tag{23}$$

The m π_k must sum to one because the $\pi_k, k = 1, \ldots, m$, represent the probabilities for the discrete random variable Θ. A Lagrange multiplier is used to maximize Equation 22 subject to the constraint that the π_k sum to one. The function to maximize is

$$\begin{aligned}\Psi(\boldsymbol{\pi}, \lambda) &= \sum_{k=1}^{m}\log(\pi_k)n_k^{(s)} + \lambda\left(\sum_{k=1}^{m}\pi_k - 1\right) \\ &= \sum_{k=1}^{m}\log(\pi_k)n_k^{(s)} + \lambda\sum_{k=1}^{m}\pi_k - \lambda\,. \end{aligned} \tag{24}$$

The partial derivatives of $\Psi(\boldsymbol{\pi}, \lambda)$ with respect to π_k are

$$\frac{\partial\Psi(\boldsymbol{\pi}, \lambda)}{\partial\pi_k} = \frac{n_k^{(s)}}{\pi_k} + \lambda\,, \tag{25}$$

for $k = 1, \ldots, m$. The partial derivative of $\Psi(\boldsymbol{\pi}, \lambda)$ with respect to λ is

$$\frac{\partial\Psi(\boldsymbol{\pi}, \lambda)}{\partial\lambda} = \sum_{k=1}^{m}\pi_k - 1\,. \tag{26}$$

Setting Equation 25 equal to zero gives

$$n_k^{(s)} = -\lambda\pi_k\,, \tag{27}$$

for $k = 1, \ldots, m$. Summing both sides of Equation 27 over k gives

$$\sum_{k=1}^{m}n_k^{(s)} = -\lambda\sum_{k=1}^{m}\pi_k = -\lambda\,. \tag{28}$$

because setting Equation 26 equal to zero implies that $\sum_k \pi_k = 1$. Substituting the value for $-\lambda$ given by Equation 28 into Equation 27 gives

$$n_k^{(s)} = \pi_k \sum_{k=1}^{m} n_k^{(s)} . \tag{29}$$

Solving Equation 29 for π_k gives values of $\pi_k^{(s+1)}$ for iteration s of the EM algorithm. The estimates of $\pi_k^{(s+1)}$ given by Equation 29 are

$$\begin{aligned} \pi_k^{(s+1)} &= \frac{n_k^{(s)}}{\sum_{k'=1}^{m} n_{k'}^{(s)}} \\ &= \frac{\sum_{i=1}^{N} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})}{\sum_{k'=1}^{m} \sum_{i=1}^{N} p(\theta_{k'} \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})} \\ &= \frac{1}{N} \sum_{i=1}^{N} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}^{(s)}) \pi_k^{(s)}}{\sum_{k'=1}^{m} f(\mathbf{y}_i \mid \theta_{k'}, \boldsymbol{\Delta}^{(s)}) \pi_{k'}^{(s)}} . \end{aligned} \tag{30}$$

Going from the second to the third line of Equation 30 follows from the fact that

$$\sum_{k'=1}^{m} p(\theta_{k'} \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) = 1 . \tag{31}$$

The m values of $\pi_k^{(s+1)}$ are the new estimates of π_k computed at iteration s. These values are used in the E step at iteration $s+1$. The values of π_k given by Equation 30 are the same as those presented in Bock and Aitken (1981), Mislevy (1984), and Titterington, Smith, and Makov (1985).

The values of $\pi_k^{(s+1)}$ for the final iteration of the EM algorithm (when convergence is achieved) are not estimates of a posterior distribution for Θ. Rather, they are maximum likelihood estimates of the marginal distribution of the discrete random variable Θ defined over the population of examinees.

After the final EM iteration the latent variable scale for most item response models can be set by linearly transforming the values of θ_k so that the mean and variance of the latent variable distribution are equal to specified values. The item parameters would also need to be transformed to be on the same scale.

To summarize the general EM algorithm for dichotomous item response models, the E step at iteration s consists of computing the $n_k^{(s)}$ and the $r_{jk}^{(s)}$ as given in Equations 19 and 20, and the M step at iteration s consists of finding estimates $\mathbf{\Delta}^{(s+1)}$ that maximize Equation 21 and computing the $\pi_k^{(s+1)}$ as given in Equation 30.

Computing Item Parameter Estimates for Polytomous Items

This section presents details of the EM algorithm for computing item parameter estimates for polytomous items. The L_j possible responses to item j are scored $0, 1, \ldots, L_j - 1$.

Equation 13 can be written in terms of item category response functions for each item. The item category response functions for item j are functions of the latent variable and the item parameters for item j. The item category response functions give the probability that an examinee with latent variable value θ will respond in item response category l of item j. The item category response functions for item j will be denoted $P_l(\theta, \boldsymbol{\delta}_j)$, $l = 0, 1, \ldots, L_j - 1$, where $P_l(\theta, \boldsymbol{\delta}_j)$ is the item category response function for the response category corresponding to item score l.

For item j with item parameters $\boldsymbol{\delta}_j$ and randomly sampled examinee i with ability value θ_k the probability of item response y_{ij} is given by

$$f(y_{ij} \mid \theta_k, \boldsymbol{\delta}_j) = \prod_{l=0}^{L_j-1} P_l(\theta_k, \boldsymbol{\delta}_j)^{\mathrm{I}_{\{y_{ij}=l\}}} , \tag{32}$$

where $\mathrm{I}_{\{y_{ij}=l\}}, l = 0, 1, \ldots, L_j - 1$ is equal to 1 if $y_{ij} = l$ and zero otherwise.

Under local independence $f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta})$ can be written using Equation 32 as

$$f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}) = \prod_{j=1}^{n} f(y_{ij} \mid \theta_k, \boldsymbol{\delta}_j) = \prod_{j=1}^{n} \prod_{l=0}^{L_j-1} P_l(\theta_k, \boldsymbol{\delta}_j)^{\mathrm{I}_{\{y_{ij}=l\}}} . \tag{33}$$

Equation 13 can be written using Equation 33 as

$$\begin{aligned}
\phi(\mathbf{\Delta}) &= \sum_{k=1}^{m} \sum_{i=1}^{N} \left\{ \log\left[\prod_{j=1}^{n} f(y_{ij} \mid \theta_k, \boldsymbol{\delta}_j) \right] p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right\} \\
&= \sum_{k=1}^{m} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{l=0}^{L_j-1} \{ \log[P_l(\theta_k, \boldsymbol{\delta}_j)^{\mathrm{I}_{\{y_{ij}=l\}}}] p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \} \\
&= \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{l=0}^{L_j-1} \left\{ \log[P_l(\theta_k, \boldsymbol{\delta}_j)] \left[\sum_{i=1}^{N} \mathrm{I}_{\{y_{ij}=l\}} p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \right] \right\}
\end{aligned} \tag{34}$$

Equation 34 can be written as

$$\phi(\mathbf{\Delta}) = \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{l=0}^{L_j-1} \log[P_l(\theta_k, \boldsymbol{\delta}_j)] r_{jkl}^{(s)}, \tag{35}$$

where

$$r_{jkl}^{(s)} = \sum_{i=1}^{N} \mathrm{I}_{\{y_{ij}=l\}} p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) = \sum_{i=1}^{N} \frac{\mathrm{I}_{\{y_{ij}=l\}} f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}^{(s)}) \pi_k^{(s)}}{\sum_{k'=1}^{m} f(\mathbf{y}_i \mid \theta_{k'}, \mathbf{\Delta}^{(s)}) \pi_{k'}^{(s)}}. \tag{36}$$

Note that the sum of $r_{jkl}^{(s)}$ over item response categories equals $n_k^{(s)}$:

$$\begin{aligned} \sum_{l=0}^{L_j-1} r_{jkl}^{(s)} &= \sum_{l=0}^{L_j-1} \sum_{i=1}^{N} \mathrm{I}_{\{y_{ij}=l\}} p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \\ &= \sum_{i=1}^{N} \left[p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) \sum_{l=0}^{L_j-1} \mathrm{I}_{\{y_{ij}=l\}} \right] \\ &= \sum_{i=1}^{N} p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}) = n_k^{(s)}. \end{aligned} \tag{37}$$

The $r_{jkl}^{(s)}$ can be thought of as provisional estimates of the number of examinees in the sample with ability value θ_k who responded in category l of item j. The n_k, as before, may be considered provisional estimates of the number of examinees with ability value $\theta_k, k = 1, \ldots, m$.

The E step at iteration s consists of computing the values of $r_{jkl}^{(s)}$ and $n_k^{(s)}$ using values of $\mathbf{\Delta}^{(s)}$ and $\pi_k^{(s)}$ computed in iteration $s - 1$ ($\mathbf{\Delta}^{(0)}$ and $\pi_k^{(0)}$ are starting values used in iteration 0). In the M step at iteration s the values of $n_k^{(s)}$ computed in the E step are substituted into Equation 30 to obtain the estimates of $\pi_k^{(s)}$, and the values of $r_{jkl}^{(s)}$ computed in the E step are substituted into Equation 35 and the value of $\mathbf{\Delta}$, namely $\mathbf{\Delta}^{(s+1)}$, that maximizes $\phi(\mathbf{\Delta})$ is found. As was the case for dichotomous items, this can involve computing the first and second partial derivatives of $\phi(\mathbf{\Delta})$, which in turn involves computing the first and second derivatives of $\log[P_l(\theta_k, \boldsymbol{\delta}_j)]$ with respect to $\mathbf{\Delta}$. For an example, see Muraki (1992) where details of the M step computation for the generalized partial credit model are presented.

For the case of two response categories for every item ($L_j = 2$ for all j) Equation 35 is equivalent to Equation 21 with $r_{jk1}^{(s)} = r_{jk}^{(s)}$ and $r_{jk0}^{(s)} = n_k^{(s)} - r_{jk}^{(s)}$. Furthermore, $n_k^{(s)} = r_{jk0}^{(s)} + r_{jk1}^{(s)}$.

Computing Bayes Modal Estimates using the EM Algorithm

The EM algorithm can be modified to produce the posterior mode of $(\boldsymbol{\Delta}, \boldsymbol{\pi})$ (Dempster, Laird, and Rubin, 1977, pg. 6). If the log of the prior distribution of $(\boldsymbol{\Delta}, \boldsymbol{\pi})$ is $G[(\boldsymbol{\Delta}, \boldsymbol{\pi})]$ then to produce Bayesian modal estimates of $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ (instead of maximum likelihood estimates) $Q[(\boldsymbol{\Delta}, \boldsymbol{\pi}) \mid (\boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})] + G[(\boldsymbol{\Delta}, \boldsymbol{\pi})]$ is maximized in the M step. Note that the E step calculation does not change. The M step calculation for $\boldsymbol{\pi}$ does not change if a uniform prior is used for $\boldsymbol{\pi}$. Prior distributions for $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ as contained in $G[(\boldsymbol{\Delta}, \boldsymbol{\pi})]$ are not the same as the values $\boldsymbol{\Delta}^{(0)}$ and $\boldsymbol{\pi}^{(0)}$ used to start the EM iterations. The values of $\boldsymbol{\Delta}^{(0)}$ and $\boldsymbol{\pi}^{(0)}$ are starting values, not prior distributions. Mislevy (1986) and Tsutakawa and Lin (1986) discusses Bayes modal estimation using the EM algorithm for the 3-parameter and 2-parameter logistic IRT models (see also Harwell and Baker, 1991).

Marginal Maximum Likelihood Using the Bock-Aitken Algorithm

This section discusses the relationship between the EM algorithm given above and the algorithm for marginal maximum likelihood given by Bock and Aitken (1981). Bock and Aitken (1981) start with a continuous latent variable and use a discrete version of the latent variable for computational purposes (numerical quadrature). Here, a discrete latent variable is specified in the model.

A typical implementation of marginal maximum likelihood uses the marginal distribution of the observed variables as calculated using a *specified* distribution of the latent variable (although it is possible to estimate the distribution along with the item parameters). Marginal maximum likelihood estimates of the item parameters are those that maximize the marginal likelihood of the observed variables. If $f(\mathbf{y}_i, \theta_i \mid \boldsymbol{\Delta}, \boldsymbol{\pi}^*)$ is the joint likelihood of the observed and missing data for examinee i, and $p(\theta_k \mid \pi_k^*)$ specifies a marginal discrete distribution for the latent variable (with known probabilities π_k^*), then the marginal likelihood for examinee i is

$$f(\mathbf{y}_i \mid \mathbf{\Delta}) = \sum_{k=1}^{m} f(\mathbf{y}_i, \theta_k \mid \mathbf{\Delta}, \boldsymbol{\pi}^*)$$
$$= \sum_{k=1}^{m} f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}) p(\theta_k \mid \pi_k^*)$$
$$= \sum_{k=1}^{m} f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}) \pi_k^* . \tag{38}$$

The log of the marginal likelihood for the whole sample is

$$\log \prod_{i=1}^{N} f(\mathbf{y}_i \mid \mathbf{\Delta}) = \sum_{i=1}^{N} \log[f(\mathbf{y}_i \mid \mathbf{\Delta})]$$
$$= \sum_{i=1}^{N} \log \left\{ \sum_{k=1}^{m} f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta}) \pi_k^* \right\} . \tag{39}$$

The log-likelihood in Equation 39 is the same as the observed data log-likelihood (the log of Equation 7) that is maximized by the EM algorithm with the exception that the π_k^* are treated as known values in Equation 39 whereas the π_k are parameters to be estimated in the EM algorithm presented previously. Thus, the EM algorithm could be used to maximize the log-likelihood in Equation 39 using initial values $\pi_k^{(0)} = \pi_k^*$ and setting $\pi_k^{(s+1)} = \pi_k^{(0)} = \pi_k^*$ for all iterations. The M step for the parameters of the latent variable distribution would not be performed; rather the same values of $\pi_k^{(s)}$ would be used for every iteration.

The description of the Bock-Aitken algorithm presented here follows that of Harwell, Baker, and Zwarts (1988). The maximum value of the log-likelihood in Equation 39 as a function of $\mathbf{\Delta}$ will occur at a value of $\mathbf{\Delta}$ for which the derivative of Equation 39 is equal to zero. The maximum likelihood estimates of the item parameters are found by solving the system of equations given by setting the derivatives of Equation 39 with respect to the item parameters equal to zero. The derivative of the log-likelihood (Equation 39) with respect to δ_{rj} (the r-th item parameter for item j) is

$$\sum_{i=1}^{N} \frac{\partial \log[f(\mathbf{y}_i \mid \boldsymbol{\Delta})]}{\partial \delta_{rj}} = \sum_{i=1}^{N} \frac{1}{f(\mathbf{y}_i \mid \boldsymbol{\Delta})} \frac{\partial f(\mathbf{y}_i \mid \boldsymbol{\Delta})}{\partial \delta_{rj}}$$
$$= \sum_{i=1}^{N} \frac{1}{f(\mathbf{y}_i \mid \boldsymbol{\Delta})} \sum_{k=1}^{m} \frac{\partial f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})\pi_k^*}{\partial \delta_{rj}}$$
$$= \sum_{i=1}^{N} \frac{\pi_k^*}{f(\mathbf{y}_i \mid \boldsymbol{\Delta})} \sum_{k=1}^{m} \frac{\partial f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})}{\partial \delta_{rj}} \tag{40}$$

Using the equality

$$\frac{\partial f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})}{\partial \delta_{rj}} = \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})]}{\partial \delta_{rj}} f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta}), \tag{41}$$

Equation 40 can be written as

$$\sum_{i=1}^{N} \frac{\partial \log[f(\mathbf{y}_i \mid \boldsymbol{\Delta})]}{\partial \delta_{rj}} = \sum_{i=1}^{N} \frac{f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})\pi_k^*}{f(\mathbf{y}_i \mid \boldsymbol{\Delta})} \sum_{k=1}^{m} \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})]}{\partial \delta_{rj}}$$
$$= \sum_{i=1}^{N} \frac{f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})\pi_k^*}{\sum_{k=1}^{m} f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})\pi_k^*} \sum_{k=1}^{m} \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})]}{\partial \delta_{rj}}$$
$$= \sum_{i=1}^{N} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}, \boldsymbol{\pi}^*) \sum_{k=1}^{m} \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})]}{\partial \delta_{rj}}$$
$$= \sum_{k=1}^{m} \sum_{i=1}^{N} \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})]}{\partial \delta_{rj}} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}, \boldsymbol{\pi}^*). \tag{42}$$

Equation 15 (with $\boldsymbol{\Delta}^{(s)}$ replaced by $\boldsymbol{\Delta}$) is used in going from the second to third line in Equation 42.

The item parameter estimates that are computed in the M step of the EM algorithm are the solutions to a system of equations that results from setting the derivatives of Equation 13, with respect to the item parameters, equal to zero. The derivative of Equation 13 with respect to δ_{rj} is

$$\sum_{k=1}^{m} \sum_{i=1}^{N} \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \boldsymbol{\Delta})]}{\partial \delta_{rj}} p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)}). \tag{43}$$

Equations 43 and 42 are identical with the exception that in Equation 43 $\boldsymbol{\Delta}^{(s)}$ is used in place of $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}^{(s)}$ is used in place of $\boldsymbol{\pi}^*$ in the function p. This distinction can have a significant effect on the computational effort needed to solve the system of equations. In Equation 43 the item parameters that are being solved for only appear in the derivative (the first term in the product inside the sums), but not in the function p (where $\mathbf{\Delta}^{(s)}$ is treated as a constant). In Equation 42 the item parameters appear in both the derivative and the function p. It is usually the case that the derivative in Equation 43 will depend only on the item parameters for item j (and consequently Equation 43 will depend only on the item parameters for item j). Consequently, in the EM algorithm the parameter estimates for each item can be solved for separately. In contrast, Equation 42 will depend on the parameters for all items — not just the parameters for item j. The computationally simpler approach of separately solving for the parameters of each individual item cannot be used with Equation 42. The difference between using Equation 42 versus Equation 43 for estimating the item parameters is illustrated by Tanner (1996, Section 4.1) for the two-parameter logistic item response model.

Marginal maximum likelihood estimates that are solutions to the system of equations given by setting Equation 42 equal to zero for all item parameters have been presented for some specific item response models. Thissen (1982) presented marginal maximum likelihood estimates of the item parameters for the one-parameter logistic model for dichotomous items. Bock and Lieberman (1970) presented marginal maximum likelihood estimates for a two-parameter normal ogive model for dichotomous items, but their solution is only computationally practical for a small number of items.

The computational complexity of estimates based on using Equation 42 led Bock and Aitken (1981) to suggest a two-step algorithm for computing marginal maximum likelihood estimates of the item parameters. Iteration s ($s = 0, 1, \ldots$) of the Bock-Aitken algorithm consists of two steps. In the first step of the Bock-Aitken algorithm at iteration s item parameters computed at iteration $s-1$ ($\mathbf{\Delta}^{(s)}$, where $\mathbf{\Delta}^{(0)}$ are starting values) are used to calculate the values of $p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^*)$ using Equation 15. In the second step at iteration s item parameters that are the solution of

$$\sum_{i=1}^{N} \sum_{k=1}^{m} \frac{\partial \log[f(\mathbf{y}_i \mid \theta_k, \mathbf{\Delta})]}{\partial \delta_{rj}} p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^*) = 0 \tag{44}$$

are found ($\mathbf{\Delta}^{(s+1)}$). It is easier to solve for $\mathbf{\Delta}$ in Equation 44 than in Equation 42 because

Equation 42 contains $\boldsymbol{\Delta}$ as a part of the function p, whereas there is no $\boldsymbol{\Delta}$ in the function p in Equation 44 ($\boldsymbol{\Delta}^{(s)}$ *in the function* p *in* Equation 44 *is known*). The values of $\boldsymbol{\Delta}^{(s+1)}$ found in iteration s are used in iteration $s+1$ to compute $p(\theta_k \mid \mathbf{y}_i, \boldsymbol{\Delta}^{(s+1)}, \boldsymbol{\pi}^*)$. This two-step process continues until the item parameters converge.

The left side of Equation 44 is equal to Equation 43 with $\boldsymbol{\pi}^*$ substituted for $\boldsymbol{\pi}^{(s)}$, so the item parameters that are the solution to Equation 44 are the same as the item parameters that maximize Equation 13 (with $\boldsymbol{\pi}^*$ substituted for $\boldsymbol{\pi}^{(s)}$). Consequently, the Bock-Aitken algorithm is identical to the EM algorithm where the parameters of the latent variable distribution, namely $\boldsymbol{\pi}$, are specified and not estimated. The first step of the Bock-Aitken algorithm corresponds to the E step of the EM algorithm, and the second step of the Bock-Aitken algorithm corresponds to the M step of the EM algorithm with the exception that $\boldsymbol{\pi} = \boldsymbol{\pi}^*$ is fixed and need not be estimated.

This section has discussed marginal maximum likelihood estimates for the case where the latent variable distribution is assumed known. It is also possible to estimate the latent variable probabilities along with the item parameters. In this case the Bock-Aitken algorithm is the same as the EM algorithm described in previous sections (where both $\boldsymbol{\pi}$ and $\boldsymbol{\Delta}$ were estimated). This last statement assumes that both methods start with the same θ_k values (nodes in numerical quadrature), and the same $\pi_k^{(0)}$ values (weights in numerical quadrature).

Summary

This paper presents detailed derivations of established results using a finite mixture approach. Estimates of parameters for item response models using the EM algorithm were derived treating the latent ability variable as discrete, in which case the distribution of observed item responses is a finite mixture. Maximum likelihood estimates of the item parameters and the latent variable distribution were obtained by a straightforward application of the general EM algorithm for finite mixtures. General results were presented for dichotomous item response models and for polytomous item response models. A closed-form solution for estimates of the latent ability distribution was given that applies to any item response model. Estimates for the item parameters will depend on the specific form of the item response functions, and will usually require iterative numerical procedures.

Finally, it was shown that the EM algorithm is the same as the Bock-Aitken algorithm for marginal maximum likelihood estimation of the item parameters.

Discussion

This paper focused on the case of a univariate real-valued latent variable. It is straightforward to extend the estimation procedures presented for dichotomous and polytomous items to other cases of interest. One example is the case of latent class models where the discrete latent variable is nominal. Everitt (1984) and Bartholomew (1987) discuss using the EM algorithm to obtain maximum likelihood estimates for latent class models. Another example is the case of a multivariate latent variable. It is straightforward to generalize the formulas presented in this paper for a discrete univariate latent variable to a discrete multivariate latent variable although the larger number of categories for a multivariate discrete latent variable could greatly increase the amount of computation required.

Besides the estimates of the item parameters and the discrete ability distribution that have been presented, an estimate of each examinee's ability may also be of interest. An empirical Bayes approach to obtaining an estimate of the ability value for the ith examinee is to use the distribution of the latent ability variable given by Equation 15 with the values of $\mathbf{\Delta}^{(s)}$ and $\boldsymbol{\pi}^{(s)}$ given by the final iteration of the EM algorithm (Tsutakawa and Soltys, 1988; Bock and Aitken, 1981; Bernardo and Smith, 1994). Then the mean or mode of the distribution $p(\theta_k \mid \mathbf{y}_i, \mathbf{\Delta}^{(s)}, \boldsymbol{\pi}^{(s)})$, $k = 1, \ldots, m$, can be used as an estimate of the ability of examinee i. These estimates are not true Bayesian estimates. A true Bayesian analysis would be based on the distribution $p(\theta_k \mid \mathbf{y}_i)$ with the item parameters marginalized out (see Equation 5 in Tsutakawa and Soltys, 1988). Tsutakawa and Soltys (1988) present an approximation to a Bayesian solution for dichotomous item response models.

In this paper the latent ability variable has been taken to be discrete in the model specification. It may be more natural to specify a continuous distribution for the latent ability variable. For the Rasch model it has been shown that as long as enough levels of the latent ability variable are used ($[n+2]/2$ if n is even or $[n+1]/2$ if n is odd, where n is the number of items) then the class of models using a discrete latent ability variable is the same as the class of models using a continuous ability latent variable, and the maximum likelihood estimates of the item parameters using the EM algorithm described in this paper are asymptotically identical to conditional maximum likelihood estimates of the item parameters (De Leeuw & Verhelst, 1986; Follmann, 1988; Lindsay, Clogg, & Grego, 1991).

If an estimate of a continuous ability is needed, then various methods (e.g., kernel estimators) may be used to fit a continuous distribution to the final estimates of the π_k (Tapia and Thompson, 1978). If the latent ability variable is assumed to be continuous and in the E step of the general EM algorithm the rectangle rule is used to compute the integral of the latent ability variable, then that method is computationally similar to the procedure presented here that assumes a discrete latent ability variable with the proviso that the θ_k values are equally spaced.

It seems likely that as long as enough levels of a discrete latent variable are used not much, if anything, will be lost by assuming a discrete rather than a continuous latent variable. When a real-valued discrete latent variable is used, one needs to decide on the number of levels to use and the values of the latent variable to use at each level (a similar decision would need to be made when assuming a continuous latent variable if numerical quadrature were employed). As noted above some theoretical results on the number of levels needed are available for the Rasch model.

Results pertaining to the estimation of histograms for observed continuous random variables may have some value in obtaining a rough estimate of the number of levels for the latent ability variable. Terrell and Scott (1985) propose $(2N)^{1/3}$, or a convenient slightly larger integer, as the optimal number of bins to use in constructing a histogram from continuous data. Alternatively, goodness of fit tests for various values of m, the number of ability levels, could be used to select a value of m that best fits the data (Titterington, Smith, and Makov, 1985, pg. 150). Experience indicates that 20 levels of the latent variable are about the minimum number needed to give reasonable results for two and three parameter logistic models for dichotomous items.

References

Baker, F. B. (1992). *Item response theory parameter estimation techniques.* New York: Marcel Dekker.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis.* New York: Oxford University Press.

Bernardo, J. M., & Smith A. F. M. (1994). *Bayesian Theory.* New York: John Wiley & Sons.

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

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

De Leeuw, J. & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics, 11*, 183-196.

Dennis, J. E. & Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations.* Englewood Cliffs, N.J.: Prentice-Hall.

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

Everitt, B. S. (1984). A note on parameter estimation for Lazarsfeld's latent class model using the EM algorithm. *Multivariate Behavioral Research, 19*, 79-89.

Follmann, D. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika, 53*, 553-562.

Harwell, M. R., Baker, F. B., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: A didactic. *Journal of Educational Statistics, 13*, 243-271.

Harwell, M. R., Baker, F. B. (1991). The use of prior distributions in marginalized Bayesian item parameter estimation: A didactic. *Applied Psychological Measurement, 15*, 375-389.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.

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

Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86*, 96-107.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm.

Applied Psychological Measurement, 16, 159-176.

Rigdon, S. E., & Tsutakawa, R. K. (1983). Parameter estimation in latent trait models. *Psychometrika, 48*, 567-574.

Tanner, M. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (3rd ed.). New York: Springer-Verlag.

Tapia, R. A., & Thompson, J. R. (1978). *Nonparametric probability density estimation.* Baltimore: Johns Hopkins University Press.

Terrell, G. R., & Scott, D. W. (1985). Oversmooth nonparametric density estimates. *Journal of the American Statistical Association, 80*, 209-214.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47*, 175-186.

Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49*, 501-519.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* New York: John Wiley & Sons.

Tsutakawa, R. K. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics, 9*, 263-276.

Tsutakawa, R. K., & Lin H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika, 51*, 251-267.

Tsutakawa, R. K., & Soltys, M. J. (1988). Approximation for Bayesian ability estimation. *Journal of Educational Statistics, 13*, 117-130.

Wilson, M., & Adams, R. J. (1993). Marginal maximum likelihood estimation for the ordered partition model. *Journal of Educational Statistics, 18*, 69-90.