Journal of the Korean
Statistical Society
Vol. 27, No. 1, 1998

# Initial Value Selection in Applying an EM Algorithm for Recursive Models of Categorical Variables

MiSook Jeong, [1] SungHo Kim [2] and KwangMo Jeong [3]

## ABSTRACT

Maximum likelihood estimates(MLEs) for recursive models of categorical variables are discussed under an EM framework. Since MLEs by EM often depend on the choice of the initial values for MLEs, we explore reasonable rules for selecting the initial values for EM. Simulation results strongly support the proposed selection rules.

**Key Words** : Conditional independence; Curved exponential family; Experts' opinion; Hyperplane of estimates; Guided selection; Model structure; Order-distortion; Probability interval.

[1] Department of Statistics, Pusan National University, Pusan, 609-735, South Korea.

[2] Basic Sciences Division, Korea Advanced Institute of Science and Technology, Daejon, 305-701, South Korea.

[3] Research Institute of Information and Communication, Department of Statistics, Pusan National University, Pusan, 609-735, South Korea.

## 1. INTRODUCTION

This paper presents an approach to the iterative computation of the MLEs for graphical models of categorical variables, some of which are latent satisfying some assumptions. For continuous variables, structural equation models(Bollen, 1989) are one of the most generic terms. If it is for finitely discrete or categorical variables, we may well consider probabilistic influence diagrams(Oliver and Smith, 1990), Bayesian networks(Pearl, 1988), graphical log-linear models(Fienberg, 1980 and Whittaker, 1990), and each family of models being suitable to use under certain circumstances of the relation.

Methods of fitting the structural equation models are well developed, although not complete. Jöreskog and Sörbom's(1986) LISREL and Bentler's (1985) EQS are most popular software packages for such models. For categorical variable models, the Iterative Proportional Fitting(IPF) algorithm and the Newton-Raphson algorithm are well known for fitting hierarchical log-linear models. Model fitting methods for hierarchical log-linear models are well established(Bishop, Fienberg, and Holland, 1975 and Agresti, 1990). An IPF algorithm for fitting the probabilistic influence diagrams(IDs) of categorical variables is considered in Kim(1997). While undirected graphs are used for graphical log-linear models, we use directed acyclic graphs for recursive models(Lauritzen and Wermuth,1983). Maximum likelihood(ML) estimation for recursive models is a simple matter if the data are complete, i.e., all the variables involved are observed. However, if a recursive model contains latent variables, the ML estimation may not necessarily be as simple as with complete data. When data are incomplete for recursive models, we will apply an EM algorithm for estimation.

The ideas underlying an EM algorithm have been presented in special cases by many authors. Dempster, Laird and Rubin(DLR)(1977) introduced the EM algorithm for computing MLEs with incomplete data. The EM technique and theory are applied for finding the estimates in the ML framework and the mode of the posterior distribution in a Bayesian framework. Each iteration of the algorithm consists of an expectation step followed by a maximization step. In many cases the M-step can be performed with a standard statistical package, thus saving us the programming time. But, because the EM algorithm performs E-step and M-step after generating initial values for the estimates, the initial values may affect the whole EM process. This phenomenon has been reported in literature(Wu, 1983).

In this paper we will investigate how much the initial values affect the final estimates in an EM process and propose useful rules for selecting appropriate initial values.

We now describe recursive models and variables that will be used in this paper. Let us turn our attention to educational testing, where tests are given to students in a paper-and-pencil format and each item is scored 0 or 1, 0 for an incorrect response and 1 for a correct response. In the field of educational testing, it is well known that task abilities are causally related to task performance(Greeno and Simon, 1988). When it comes to paper-and-pencil tests, the task performance is given in the form of item response. Thus we can say that item scores are influenced by states of the item-relevant abilities. In this paper, we will also use binary variables to represent the possession states of a given ability, 1 for the possession state and 0 for the other state. Note that in most cases the ability states are not observable and so that the abilities will be treated in the form of latent variables in this paper.

In the recursive model, each node represents a variable, and we will use the terms *node* and *variable* interchangeably within a model. The structure of a recursive model is the relationship that can easily be represented by a directed acyclic graph.

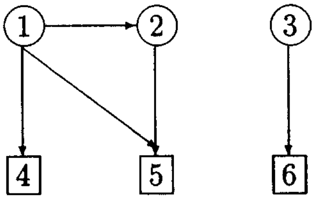


**Figure 1.1** A recursive model.

**Example 1.1** A simple example of the recursive models to be dealt with in this paper is given in Figure 1.1. The nodes 1, 2, and 3 in circles are latent variables denoting ability states of abilities 1, 2, and 3, respectively, and the nodes 4, 5 and 6 in squares are item score variables of the items 1, 2, and 3, respectively. The states of abilities 1, 2, and 3 are represented by $X_1, X_2$, and $X_3$, respectively, and the scores of items 1, 2, and 3 by $X_4, X_5$, and $X_6$, respectively.

The arrow between abilities represents a prerequisite relation, and the arrow between ability and item score represents a cause-effect relation. When there is no arrow between a pair of abilities, those in the pair are marginally independent of each other. According to the graph in Figure 1.1, ability 1 is prerequisite to ability 2, ability 1 affects item score 1, abilities 1 and 2 affect item score 2, and also ability 3 affects item score 3. But, Abilities 1 and 2 are marginally independent of ability 3.

The joint probability of the six categorical variables, $X_1, \ldots, X_6$, can be expressed by

$$P(x_1, \ldots, x_6) = P(x_1, x_2)P(x_3)P(x_4|x_1)P(x_5|x_1, x_2)P(x_6|x_3). \qquad (1.1)$$

Here $P(X_4 = 1|X_1 = 1)$ denotes the conditional probability that a randomly chosen examinee with ability 1 answers item 1 correctly. Since $X_1, X_2$, and $X_3$ are unobservable, we resort to EM to estimate the probabilities as appearing in the right-hand side of (1.1).

This paper consists of 4 sections. In section 2, we descrive an EM algorithm for finding MLEs of recursive models. And we briefly display the sensitivity of the estimates by an EM algorithm to the selection of initial values. In section 3, we propose some selection rules of initial values, and the asymptotic distributions of the parameter estimates for recursive models are presented. Merits of the proposal selection rules are described through a geometric investigation of the initial values and the final estimates by EM. Finally, section 4 sums up the results of this paper and some further concluding remarks follow in the section.

## 2. PARAMETER ESTIMATES FOR RECURSIVE MODELS

### 2.1 Recursive models

Consider a node which has at least one parent node. If the node does not have any child node, we will call it a *terminal* node, otherwise a *non-terminal* node. If a node is not connected to any other node in a graph, we will call it an *isolated* node. Since estimation for an isolated variable is equivalent to estimation for a single multinomial variable, we will consider only the recursive models without isolated nodes in this paper.

Let a recursive model, say $\mathcal{R}$, involve the categorical variables, $X_1, \ldots,$ $X_K$, where for $1 \leq L < K$, the variables, $X_{L+1}, \ldots,$ $X_K$, are terminal nodes and are conditionally independent given the first $L$ non-terminal nodes. In this paper, all the terminal nodes are observable variables and the non-terminal nodes are latent or unobservable variables. Denote by $\omega$ the index set of $X_1, \ldots, X_K$, by $\phi$ the index set of $X_{L+1}, \ldots, X_K$. We denote by $\varphi$ and $\theta$

any nonempty subsets of $\omega$; $\varphi$ for latent variables only, i.e, $\varphi = \{1, 2, \ldots, L\}$. $X_\theta$ denotes the row vector of $X's$ indexed in $\theta$. Let $n$ be the sample size. For notational convenience, we will write $P(X_\theta = x_\theta) = P(x_\theta)$.

As for $\mathcal{R}$, we denote by $\theta_i$ the index set of variable $X_{L+i}$ and its parent variable(s) and let $\varphi_i = \varphi \cap \theta_i$. Then if we let $T = K - L$, the probability model of $R$ is given by

$$P(x_\omega) = P(x_\varphi) \prod_{i=1}^{T} P(x_{\theta_i} | x_{\varphi_i}), \qquad (2.1)$$

where $P(x_\varphi)$ are expressed in various formulae according to the probability dependence structure of $X_1, \ldots, X_L$, which are latent variables. For instance, as for the model in Figure 1.1,

$$P(x_\varphi) = P(x_{\{1,2\}})P(x_3).$$

Let

$$m_\omega = m(x_\omega) = nP(x_\omega)$$

denote the cell means of $X_1$, $X_2, \ldots, X_K$ at the cell-entry $x_\omega$ and $m_\phi$ denote the cell means of the $(K - L)$-dimensional contingency table of observables, $X_{L+1}, \ldots, X_K$.

For the recursive model, the marginal and conditional probabilities on the right-hand side of equation (2.1) are parameters where the conditioned variables only are observable. The probability model of the recursive model of categorical variables pertains to the exponential family, and the MLEs for the recursive model are obtained without difficulty. The likelihood function of a recursive model is expressed as the likelihood function of a multinominal distribution model if the parameters are regarded as the cell means. But if we take the parameter space as consisting of marginal or conditional probabilities, the likelihood function for complete data is given in Lauritzen(1995). As for a recursive model of $K$ categorical variables, the log-likelihood function is given by

$$\begin{aligned} \lambda(P) &= \sum_{x_\omega} n(x_\omega)(\sum_{i\in\omega} logP(x_i|x_{\theta_i})) \\ &= \sum_{i\in\omega} \sum_{x_{\theta_i}} n(x_{\theta_i}) logP(x_i|x_{\theta_i}), \end{aligned}$$

where $n(x_\theta)$ is the number of the cases that fall into the category $x_\theta$.

The MLEs for the model (2.1) are given by

$$\hat{P}(x_\varphi) = \frac{\hat{m}_\varphi}{n}$$

and

$$\hat{P}(x_{\theta_i}|x_{\varphi_i}) = \frac{\hat{m}_{\theta_i}}{\hat{m}_{\varphi_i}},$$

where $\hat{m}_\varphi$, $\hat{m}_{\theta_i}$, and $\hat{m}_{\varphi_i}$ denote the MLEs of the cell means of the frequency tables of $X_\varphi$, $X_{\theta_i}$, and $X_{\varphi_i}$, respectively. The degree of freedom for the Pearson $\chi^2$ statistic is

$$2^{K-L} - 1 - \sum_{i=1}^{K} 2^{t_i},$$

where $t_i$ is the number of the parent nodes of variable $X_i$.

## 2.2 EM algorithm for recursive models

An EM process for a recursive model of categorical variables goes as follows.

**STEP 1** : Generation of the initial values.
The initial values $\hat{m}_\omega^{(0)}$ for the cell means of the frequency table of $X_\omega$ are generated under the structure of a given recursive model.

**STEP 2** : Likelihood maximization using the initial values.
The initial values $\hat{m}_\omega^{(0)}$ are used for obtaining new estimates $\hat{m}_\omega^{(1)}$ via likelihood maximization. $\hat{m}_\omega^{(1)}$ are obtained via

$$\hat{m}_\omega^{(1)} = \hat{m}_\varphi^{(0)} \prod_{i=1}^{T} \frac{\hat{m}_{\theta_i}^{(0)}}{\hat{m}_{\theta_i \cap \varphi}^{(0)}}.$$

**STEP 3** : E-step.
The missing cells, which are due to latent variables, are filled in. For this filling-in, the observed cell frequencies $n_\phi$ and the current estimates $\hat{m}_\omega^{(r)}$ that are obtained at the preceding M-step are used. The E-step yields

$$\hat{m}_\omega^{(r+1)} = n_\phi \frac{\hat{m}_\omega^{(r)}}{\hat{m}_\phi^{(r)}}. \qquad (2.2)$$

**STEP 4** : M-step.
The new estimates $\hat{m}_\omega^{(r+2)}$ are obtained by likelihood maximization. The M-step yiels

$$\hat{m}_\omega^{(r+2)} = \hat{m}_\varphi^{(r+1)} \prod_{i=1}^{T} \frac{\hat{m}_{\theta_i}^{(r+1)}}{\hat{m}_{\theta_i \cap \varphi}^{(r+1)}} . \tag{2.3}$$

Steps 3 and 4 make one cycle of the iteration.

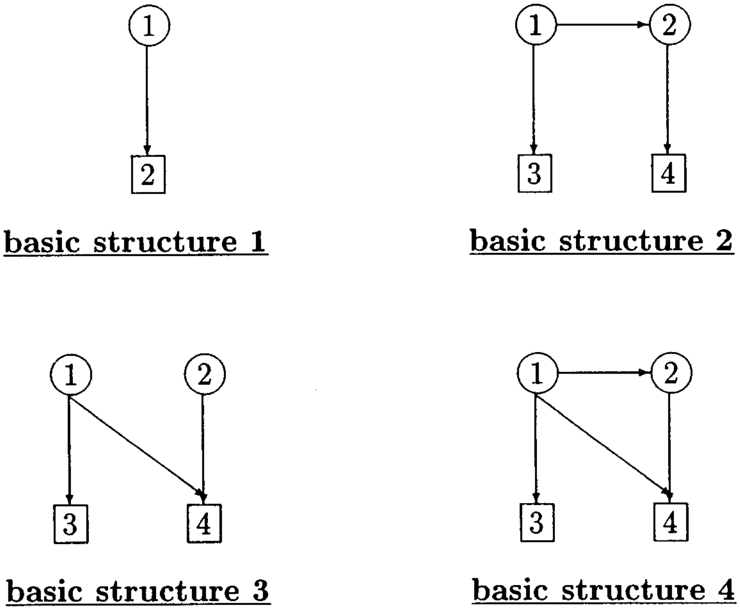
**STEP 5** : Convergence.
Steps 3 and 4 are repeated until convergence takes place. The stopping rule is to stipulate when cycle-to-cycle changes are small enough. For some small number $\varepsilon$, we stop the iteration when $|\hat{m}_\omega^{(r+2)} - \hat{m}_\omega^{(r)}| < \varepsilon$. In this paper we took 0.001 for $\varepsilon$. The MLEs for the marginal and conditional probabilities are computed from the final estimates $\hat{m}_\omega$ of the cell means.

## 2.3 Sensitivity of the estimates by the EM algorithm

As mentioned by Wu(1983), estimates from EM are often subject to the initial values used. In this subsection, we will investigate, confined to the four basic structures as in Figure 2.1, possible influences of the initial values upon the final estimates. It may sound senseless to mention basic structures for recursive models, since there are infinitely many different structures of the model. But to get an insight into the influence of the initial values, we will focus on the simple structures such as those in Figure 2.1 and call them "basic" structures. These structures involve at most 4 variables, where the numbers of latent and observable variables are half and half, respectively.

In basic structure 1, $X_1$ and $X_2$ denote the ability and the item score variables, respectively. On the other hand, in basic structure 2 through basic structure 4, the first two variables $X_1$ and $X_2$ denote the ability states and the other two variables $X_3$ and $X_4$ item scores.

In the simulated example below, the sample size $(n)$ is 500,000 and the stopping criterion $(\varepsilon)$ of EM is 0.001 .



**basic structure 1**

**basic structure 2**

**basic structure 3**

**basic structure 4**

**Figure 2.1.** Four basic structures of recursive models. Circles denote ability state variables and boxes item score variables.

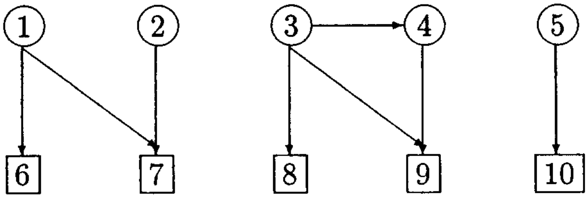


**Figure 2.2.** Model 1. Basic structures 1, 3, and 4 are nested therein.

**Example 2.1** We assume that $X_1$ through $X_5$ are latents and $X_6$ through $X_{10}$ are item score variables. The 10 variables are all binary. We will call the model in Figure 2.2 Model 1. Basic structures 1, 3 and 4 are nested in the model. As for Model 1, there are 20 parameters to estimate, which are $P(X_1 = 1)$, $P(X_2 = 1)$, $P(X_3 = 1)$, $P(X_4 = 1|X_3 = x_3), x_3 = 0, 1$, $P(X_5 = 1)$, $P(X_6 = 1|X_1 = x_1), x_1 = 0, 1$, $P(X_7 = 1|X_1 = x_1, X_2 = x_2), x_1 = 0, 1, x_2 = 0, 1$, $P(X_8 = 1|X_3 = x_3), x_3 = 0, 1$, $P(X_9 = 1|X_3 = x_3, X_4 = x_4), x_3 = 0, 1, x_4 = 0, 1$, and $P(X_{10} = 1|X_5 = x_5), x_5 = 0, 1$. Since only 5 variables are observable, the degree of freedom for the Pearson $\chi^2$ statistic is 11.

We let $\omega = \{1,2,\dots,10\}$, $\phi = \{6,7,\dots,10\}$, $\varphi = \{1,2,\dots,5\}$, $\theta_1 = \{1,6\}$, $\theta_2 = \{1,2,7\}$, $\theta_3 = \{3,8\}$, $\theta_4 = \{3,4,9\}$, and $\theta_5 = \{5,10\}$. The structure of Model 1 implies that $X_1$, $X_2$, $X_3$, $X_5$ are marginally independent of each other, that $X_4$ is dependent on $X_3$, that $X_6,\dots,X_{10}$ are conditionally independent given $X_1,\dots,X_5$, that $X_7$ depends on $X_1$ and $X_2$, that $X_9$ depends on $X_3$ and $X_4$, and that $X_6, X_8$ and $X_{10}$ depend on $X_1, X_3$ and $X_5$, respectively.

We assume for this example that the item score variable for item $i$ is denoted by $X_{i+5}$, for $i = 1,\dots,5$ and that the proportion correct of item $i$ by $\bar{X}_{i+5}$.

Table 2.1 shows the MLEs via EM when the proportions correct of items are $\bar{X}_6 = .730$, $\bar{X}_7 = .344$, $\bar{X}_8 = .822$, $\bar{X}_9 = .762$, and $\bar{X}_{10} = .399$.

In Table 2.1, $P$ and $P^{(0)}$ denote the actual probabilities and the initial values for their estimates, respectively. The probability estimates $\hat{P}$ are MLEs via EM. The actual probabilities are from Kim(1994). And the $MSE(\hat{P})$ is the mean squared error of $\hat{P}$ obtained through 100 replications with the sample size 100,000 for each replication. An asterisk($*$) is attached to the ineffective estimates, not belonging to the probability interval(to be explained in subsection 3.4) at the significance level 0.05. As Table 2.1 shows, the MSEs of the estimates, when the initial values are chosen randomly, are relatively large. This indicates that the estimates by EM are largely dependent on the initial values. Thus we ought to be cautious in selecting the initial values.

## 3. GUIDED SELECTION FOR INITIAL VALUES

This section presents a guideline for selecting the initial values and considers the asymptotic distribution of parameter estimates. Using the asymptotic distribution, we will explore appropriate rules for selecting the initial values.

Before setting off towards the rules, we will have a look into the geometry of EM as applied to recursive models of categorical variable. The E-step is implemented through expression (2.2). Once implemented, the new estimates $\hat{m}_\omega^{(r+1)}$ satisfy that

$$\hat{m}_\phi^{(r+1)} = n_\phi.$$

**Table 2.1.** Actual probabilities and estimated probabilities for Model 1.

| | $P$ | $P^{(0)}$ | $\hat{P}$ | $MSE(\hat{P})$ |
|---|---|---|---|---|
| $P(X_1 = 1)$ | .896 | .500 | .554* | .1182 |
| $P(X_6 = 1 \mid X_1 = 0)$ | .100 | .300 | .497* | .1621 |
| $P(X_6 = 1 \mid X_1 = 1)$ | .801 | .950 | .917* | .0127 |
| $P(X_2 = 1)$ | .400 | .500 | .472 | .0052 |
| $P(X_7 = 1 \mid X_1 = 0, X_2 = 0)$ | .000 | .001 | .004* | .0000 |
| $P(X_7 = 1 \mid X_1 = 0, X_2 = 1)$ | .098 | .200 | .470* | .1406 |
| $P(X_7 = 1 \mid X_1 = 1, X_2 = 0)$ | .101 | .200 | .108 | .0000 |
| $P(X_7 = 1 \mid X_1 = 1, X_2 = 1)$ | .797 | .900 | .814 | .0003 |
| | | | | |
| $P(X_3 = 1)$ | .900 | .950 | .855 | .0021 |
| $P(X_8 = 1 \mid X_3 = 0)$ | .106 | .200 | .096 | .0001 |
| $P(X_8 = 1 \mid X_3 = 1)$ | .901 | .990 | .945 | .0020 |
| $P(X_4 = 1 \mid X_3 = 0)$ | .851 | .500 | .555* | .0865 |
| $P(X_4 = 1 \mid X_3 = 1)$ | .898 | .900 | .864 | .0012 |
| $P(X_9 = 1 \mid X_3 = 0, X_4 = 0)$ | .000 | .100 | .173* | .0297 |
| $P(X_9 = 1 \mid X_3 = 0, X_4 = 1)$ | .206 | .400 | .550* | .1205 |
| $P(X_9 = 1 \mid X_3 = 1, X_4 = 0)$ | .200 | .500 | .333* | .0177 |
| $P(X_9 = 1 \mid X_3 = 1, X_4 = 1)$ | .899 | .950 | .904 | .0000 |
| | | | | |
| $P(X_5 = 1)$ | .398 | .800 | .574* | .0302 |
| $P(X_{10} = 1 \mid X_5 = 0)$ | .100 | .100 | .026 | .0056 |
| $P(X_{10} = 1 \mid X_5 = 1)$ | .850 | .900 | .676* | .0297 |

**NOTE: An asterisk (*) is attached when $\hat{P}$ is not contained in the corresponding probability interval at the significance level 0.05.**

That is, when the new estimates are marginalized on $X_\phi$, the marginals are the same as $n_\phi$. This means geometrically that the points $\{\hat{m}_\omega^{(r+1)}(x_\omega)\}$ in the space of $[0,1]^{2^K-1}$ lie in the hyperplane $\mathcal{H}_1$ given by

$$\mathcal{H}_1 = \{m_\omega(x_\omega); m_\phi(x_\phi) = n_\phi(x_\phi) \text{ for all possible configurations } x_\phi\}.$$

On the other hand, the M-step is carried out through expression (2.3). $\hat{m}_\omega^{(r+2)}$ is defined in terms of as many factors as appearing in the right-hand side of (2.3). The factors can be in the form of marginal or conditional probabilities. The marginal $\hat{m}_\varphi^{(r+1)}$ can be further factorized according to the structure of $X_\varphi$. We know that a model structure that is expressed in terms of conditional independence relationship implies a constraint for $\{m_\omega\}$. A good example is given in section 2.7 of Bishop, Fienberg, and Holland(1975). As for our recursive models, the relationship among $X_\omega$ is fully representable via directed acyclic graph. For instance, as described in Example 2.1, we can read a variety of marginal or conditional independencies from Figure 2.2. These independencies restrict the set $\{m_\omega\}$ into a hyperplane where the independencies are satisfied. Interested readers are refered to Figure 2.7-3 in Bishop et al(1975, p.53) for a hyperplane of independence of two binary variables. For convenience' sake, we will denote by $\mathcal{H}_2$ the hyperplane that satisfies the independence relationship among the variables in $X_\omega$ of a given recursive model. Then the estimates $\hat{m}_\omega^{(r+1)}$ in expression (2.3) from an M-step must lie in the hyperplane $\mathcal{H}_2$.

Therefore, the final estimates from an EM algorithm should be contained in $\mathcal{H}_1 \cap \mathcal{H}_2$. Visualization of $\mathcal{H}_1 \cap \mathcal{H}_2$ is impossible when the K-dimensional contingency table contains 5 cells or more for which the cell means are to be estimated.

The EM problem is an optimization problem for the likelihood function where the domain of the likelihood function is confined to $\mathcal{H}_1 \cap \mathcal{H}_2$. Whether the final estimates $\{\hat{m}_\omega\}$ from an EM are the global maximum point of the likelihood function or a local maximum point depends on the shape of the hyperplane $\mathcal{H}_1 \cap \mathcal{H}_2$, which is hard to visualize or analyze when it is of dimension 4 or higher.

It is important to note that the recursive model of categorical variables belongs to a curved exponential family in general. The natural parameter space is of the cell means of $X_\omega$ and the cell means are obtained through a joint probability model such as in expression (2.1), where each marginal or conditional probability on the righthand side is a parameter. Fisher(1925) classified such a model as belonging to a curved multinomial family. Efron(1978) explored the relation between data point and parameter space confined to (curved) exponential families. He showed that the MLE point is located at a point in the parameter space which is closest, under some condition, to the data point. Our problem here is that we can hardly figure out the shape of $\mathcal{H}_1 \cap \mathcal{H}_2$.

Thus it is desirable to try several different initial points for EM and choose the best among the sets of the final estimates $\{\hat{m}_\omega\}$ as maximum likelihood estimates. The "best" is in the sense that the value of the Pearson chi-square statistic is the smallest among the several sets of final estimates from EM. Note that the log-likelihood function is concave and the Pearson chi-square statistic is convex in the estimates and so that the set of estimates which give the largest value to the log-likelihood function give the smallest value to the Pearson chi-square statistic. As aforementioned, it is hard to see if the "largest" value is the global maximum value, when $\mathcal{H}_1 \cap \mathcal{H}_2$ is hard to analyze. This is why we should exercise our discretion so that the initial values might fall within a reasonable range, when the model is relatively complex. The "best" estimates are, in this respect, an outcome of the mixture of the model, data, and the discretion, which we will elaborate on in the rest of the paper.

### 3.1 Useful rules for selecting the initial values

Given a recursive model of abilities and item scores, we aim to find the marginal and conditional probabilities of abilities as well as the conditional probabilities of item scores given abilities. Because ability variables are latent, we use the proportions correct of items in selecting some reasonable initial values. In the rest of this subsection, we will derive, confined to the four basic structures, useful results for selecting reasonable initial values.

**Theorem 3.1. (For basic structure 1)** Assume that $X_1$ and $X_2$ are related as in basic Structure 1 and suppose that

$$P(X_2 = 1|X_1 = 1) \geq \alpha_{11},\ P(X_2 = 1|X_1 = 0) \leq \alpha_{10} \tag{3.1}$$

for real constants $\alpha_{10}$ and $\alpha_{11}$ with $0 \leq \alpha_{10} < \alpha_{11} \leq 1$. Then we have

$$l \leq P(X_1 = 1) \leq u, \tag{3.2}$$

where

$$\begin{aligned} u &= min(\frac{P(X_2 = 1) - P(X_2 = 1|X_1 = 0)}{\alpha_{11} - \alpha_{10}}, 1) \\ l &= max(\frac{P(X_2 = 1) - \alpha_{10}}{P(X_2 = 1|X_1 = 1) - \alpha_{10}}, 0) \end{aligned}$$

**Proof.** The probability of a correct response to item 1 is

$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 1)P(X_2 = 1|X_1 = 1) \\ &\quad + \{1 - P(X_1 = 1)\}P(X_2 = 1|X_1 = 0). \end{aligned} \tag{3.3}$$

From (3.1) and (3.3) follows the desired result. □

We may choose as an initial value for $P(X_1 = 1)$ any value which satisfies (3.2). In practice, we can consult experts for $\alpha$'s in (3.1).

**Remark 3.1.** Under condition (3.1), the maximal value of $(P(X_2 = 1) - \alpha_{10})/(P(X_2 = 1|X_1 = 1) - \alpha_{10})$ is $(P(X_2 = 1) - \alpha_{10})/(\alpha_{11} - \alpha_{10})$, and the maximal value of $(P(X_2 = 1) - P(X_2 = 1|X_1 = 0))/(\alpha_{11} - \alpha_{10})$ is $P(X_2 = 1)/(\alpha_{11} -$

$\alpha_{10})$. Because $P(X_2 = 1|X_1 = 1)$ and $P(X_2 = 1|X_1 = 0)$ are unknown, we use $min\{P(X_2 = 1)/(\alpha_{11} - \alpha_{10}), 1\}$ and $max\{(P(X_2 = 1) - \alpha_{10})/(\alpha_{11} - \alpha_{10}), 0\}$ instead of $u$ and $l$ in (3.2), respectively, when we determine the initial value for $P(X_1 = 1)$.

**Theorem 3.2. (For basic structure 2)** Assume that $X_1$ through $X_4$ are related as in basic structure 2 and suppose that

$$P(X_3 = 1|X_1 = 1) \geq \alpha_{11},\ P(X_3 = 1|X_1 = 0) \leq \alpha_{10} \tag{3.4}$$
$$P(X_4 = 1|X_2 = 1) \geq \alpha_{21},\ P(X_4 = 1|X_2 = 0) \leq \alpha_{20} \tag{3.5}$$
$$P(X_2 = 1|X_1 = 1) \geq P(X_2 = 1|X_1 = 0) \tag{3.6}$$

for real constants $\alpha_{i0}$ and $\alpha_{i1}$ with $0 \leq \alpha_{i0} < \alpha_{i1} \leq 1, i = 1, 2$. Then we have

$$l_i \leq P(X_i = 1) \leq u_i, \tag{3.7}$$

where for (i,k) = (1,3), (2,4)

$$\begin{aligned} u_i &= min(\frac{P(X_k = 1) - P(X_k = 1|X_i = 0)}{\alpha_{i1} - \alpha_{i0}}, 1) \\ l_i &= max(\frac{P(X_k = 1) - \alpha_{i0}}{P(X_k = 1|X_i = 1) - \alpha_{i0}}, 0). \end{aligned}$$

**Proof.** The proof for $P(X_1 = 1)$ is identical to that of Theorem 3.1. Since the probability for a correct response to item 2 is

$$\begin{aligned} P(X_4 = 1) &= P(X_2 = 1)P(X_4 = 1|X_2 = 1) \\ &\quad + P(X_2 = 0)P(X_4 = 1|X_2 = 0)\,, \end{aligned}$$

this is also the same form as that of Theorem 3.1. This completes the proof of the theorem. □

We may choose as initial values for $P(X_1 = 1)$ and $P(X_2 = 1)$ any values which satisfy (3.7).

**Remark 3.2.** After dividing basic structure 2 into two probability models, each having basic structure 1, the inequality (3.2) can be applied to basic structure 2.

The initial values for the conditional probability $P(X_2|X_1)$ in basic structure 2 will be selected according to Remark 3.3. The probability $P(X_2 = 1)$ is represented as

$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 1)P(X_2 = 1|X_1 = 1) \\ &\quad + \{1 - P(X_1 = 1)\}P(X_2 = 1|X_1 = 0). \end{aligned} \tag{3.8}$$

**Remark 3.3.** Let $X_1$ and $X_2$ denote the state of abilities 1 and 2, respectively. Then we may safely assume

$$P(X_2 = 1|X_1 = 0) \leq P(X_2 = 1|X_1 = 1).$$

So from (3.8), follows that

$$P(X_2 = 1|X_1 = 0) \leq P(X_2 = 1) \leq P(X_2 = 1|X_1 = 1). \tag{3.9}$$

After determining values for $P^{(0)}(X_1 = 1)$ and $P^{(0)}(X_2 = 1)$ in such a way that (3.7) is satisfied and by consulting experts for $P^{(0)}(X_2 = 1|X_1 = 1)$, the initial values for $P^{(0)}(X_2 = 1|X_1 = 0)$ is determined from (3.8).

**Theorem 3.3. (For basic structure 3)** Assume that $X_1$ through $X_4$ are related as in basic structure 3 and suppose that

$$P(X_3 = 1|X_1 = 1) \geq \alpha_{11}, \;\; P(X_3 = 1|X_1 = 0) \leq \alpha_{10} \tag{3.10}$$

$$\begin{aligned} P(X_4 = 1|X_1 = 1, X_2 = 1) &= \alpha_{22} \\ P(X_4 = 1|X_1 + X_2 = 1) &= \alpha_{21} \\ P(X_4 = 1|X_1 = 0, X_2 = 0) &= \alpha_{20} \end{aligned} \tag{3.11}$$

for real constants $\alpha_{10}$ and $\alpha_{11}$ with $0 \leq \alpha_{10} < \alpha_{11} \leq 1$ and for real constants $\alpha_{20}, \alpha_{21}$ and $\alpha_{22}$ with $\alpha_{20} \leq \alpha_{21} \leq \alpha_{22}$. Then we have

$$l_1 \leq P(X_1 = 1) \leq u_1, \tag{3.12}$$

where

$$\begin{aligned} u_1 &= min(\frac{P(X_3 = 1) - P(X_3 = 1|X_1 = 0)}{\alpha_{11} - \alpha_{10}}, 1) \\ l_1 &= max(\frac{P(X_3 = 1) - \alpha_{10}}{P(X_3 = 1|X_1 = 1) - \alpha_{10}}, 0), \end{aligned}$$

and

$$l_2 \leq P(X_2 = 1) \leq u_2, \tag{3.13}$$

where if $\alpha_{22} - 2\alpha_{21} + \alpha_{20} \geq 0$, then

$$\begin{aligned} u_2 &= min(\frac{P(X_4 = 1) - l_1(\alpha_{21} - \alpha_{20}) - \alpha_{20}}{\alpha_{21} + l_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}}, 1) \\ l_2 &= max(\frac{P(X_4 = 1) - u_1(\alpha_{21} - \alpha_{20}) - \alpha_{20}}{\alpha_{21} + u_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}}, 0) \end{aligned}$$

and if $\alpha_{22} - 2\alpha_{21} + \alpha_{20} < 0$, then $l_1$ and $u_1$ in denominators for $u_2$ and $l_2$ are interchanged.

**Proof.** See Appendix.

We may choose as the initial values for $P(X_1 = 1)$ and $P(X_2 = 1)$ any values which satisfy (3.12) and (3.13), respectively.

**Theorem 3.4. (For basic structure 4)** Assume that $X_1$ through $X_4$ are related as in basic structure 4, suppose that

$$P(X_2 = 1|X_1 = 1) \geq P(X_2 = 1|X_1 = 0). \tag{3.14}$$

And also suppose (3.10)and (3.11). Then we have

$$l_i \leq P(X_i = 1) \leq u_i, \ \ i = 1, 2, \tag{3.15}$$

where $u_i$ and $l_i$, $i = 1, 2$ are identical to those of Theorem 3.3.

**Proof.** See Appendix.

We may choose as the initial values for $P(X_1 = 1)$ and $P(X_2 = 1)$ in basic structure 4 any values which satisfy (3.12) and (3.13), respectively. Also, the initial values for $P(X_2 = 1|X_1 = 1)$ and $P(X_2 = 1|X_1 = 0)$ are determined as suggested in Remark 3.3.

### 3.2 Applications of the proposed selection rules

Applying the results derived in this subsection, we may obtain the reasonable initial values for the marginal probabilities and the conditional probabilities for the basic structures. We will call by "the guided selection(GS)" the selection which is implemented according to the results in this subsection; otherwise, we will call it "an unguided selection(UGS)." Because a GS is dependent on $\boldsymbol{\alpha}$'s, the opinion of experts plays a crucial role in applying a GS.

**Remark 3.4.** The initial values for $\alpha_{10}$ and $\alpha_{11}$ may be recommended as in the table below corresponding to the proportion correct of item 1, $\bar{X}_2$, when the item score variable $X_2$ depends on the latent variable $X_1$ as in basic structures 1 and 2. This table of recommendation does not stand on by a theoretic ground but is obtained by experience or experts' opinions, and in the subsequent examples in this paper this table will be used. This table could change case by case and according to experts' comments or suggestions. This table is simply an illustraction of how we select initial values for $\alpha_{10}$ and $\alpha_{11}$ with regard to basic structure 1.

| $\bar{X}_2$ | $\alpha_{11}^{(0)}$ | $\alpha_{10}^{(0)}$ |
|---|---|---|
| $> .90$ | $\geq .95$ | $\leq .30$ |
| $> .75$ | $\geq .90$ | $\leq .20$ |
| $> .50$ | $\geq .85$ | $\leq .10$ |
| $\leq .50$ | $\geq .80$ | $\leq .05$ |

**Remark 3.5.** The initial values for $\alpha_{20}$, $\alpha_{21}$ and $\alpha_{22}$ may be recommened as in the table below corresponding to the proportion correct of item 2, $\bar{X}_4$, when the item score variable $X_4$ depends on the latent variables $X_1$ and $X_2$ as in basic structures 3 and 4. This table is a simple suggestion for $\alpha_{20}$, $\alpha_{21}$, and $\alpha_{22}$ with regard to basic structures 3 and 4.

| $\bar{X}_4$ | $\alpha_{22}^{(0)}$ | $\alpha_{21}^{(0)}$ | $\alpha_{20}^{(0)}$ |
|---|---|---|---|
| $> .70$ | $\geq .90$ | $\leq .40$ | $\leq .15$ |
| $> .40$ | $\geq .85$ | $\leq .30$ | $\leq .10$ |
| $\leq .40$ | $\geq .80$ | $\leq .20$ | $\leq .05$ |

When the experts' opinions are not available, we may select the initial values for $\boldsymbol{\alpha}$'s as recommended in Remarks 3.4 and 3.5. For example, if the proportion correct of the item corresponding to $X_2$ in basic structure 1 is equal to 0.7, we may choose, by Remark 3.4, any initial values $\alpha_{11}^{(0)}$ and $\alpha_{10}^{(0)}$ so that $\alpha_{11}^{(0)} \geq 0.85$ and $\alpha_{10}^{(0)} \leq 0.10$ are satisfied. And if the proportion correct of the item corresponding to $X_4$ in basic structure 3 is equal to 0.5, we may choose by Remark 3.5 any initial values $\alpha_{22}^{(0)}$, $\alpha_{21}^{(0)}$ and $\alpha_{20}^{(0)}$ so that $\alpha_{22}^{(0)} \geq 0.85$, $\alpha_{21}^{(0)} \leq 0.30$ and $\alpha_{20}^{(0)} \leq 0.10$ are satisfied.

When we apply the EM algorithm with improper $\boldsymbol{\alpha}$'s, 'order-distortion' or 'abnormal situation' may occur in the estimates. In terms of conditional probability, we say that an order-distortion takes place if

$$P(X_l = 1|X_1 = x_1, \cdots, X_k = x_k) > P(X_l = 1|X_1 = x_1', \cdots, X_k = x_k')$$

when $x_i \leq x_i'$ for $i = 1, \ldots, k$ and $x_j < x_j'$ for at least one $j, 1 \leq j \leq k$.

We will now return to the example in Section 2 to illustrate some simple applications of the results of the previous subsection.

**Example 2.1(Continued)** We consider estimating the parameters of Model 1 in Figure 2.2. We will determine the initial values for the conditional probabilities $\boldsymbol{\alpha}$'s of the item score variable given latent variables based on Remarks 3.4 and 3.5. The selection results are summarized below.

- $\bar{X}_6 = 0.730 \;\; : \;\; \alpha_1^{(0)} = 0.85, \alpha_0^{(0)} = 0.079$
  The selection interval for $P(X_1 = 1)$ is

$$0.844 \leq P(X_1 = 1) \leq 0.947. \tag{3.16}$$

- $\bar{X}_7 = 0.344 \;\; : \;\; \alpha_2^{(0)} = 0.80, \alpha_1^{(0)} = 0.10, \alpha_0^{(0)} = 0.001$ and (3.16)
  The selection interval for $P(X_2 = 1)$ is

$$0.373 \leq P(X_2 = 1) \leq 0.428. \tag{3.17}$$

- $\bar{X}_8 = 0.822 \;\; : \;\; \alpha_1^{(0)} = 0.90, \alpha_0^{(0)} = 0.10$
  The selection interval for $P(X_3 = 1)$ is

$$0.903 \leq P(X_3 = 1) \leq 1.000. \tag{3.18}$$

- $\bar{X}_9 = 0.762 \;\; : \;\; \alpha_2^{(0)} = 0.90, \alpha_1^{(0)} = 0.20, \alpha_0^{(0)} = 0.001$ and (3.18)
  The selection interval for $P(X_4 = 1)$ is

$$0.803 \leq P(X_4 = 1) \leq 0.892. \tag{3.19}$$

- $\bar{X}_{10} = 0.399 \;\; : \;\; \alpha_1^{(0)} = 0.80, \alpha_0^{(0)} = 0.01$
  The selection interval for $P(X_5 = 1)$ is

$$0.492 \leq P(X_5 = 1) \leq .505. \tag{3.20}$$

From the selection intervals (3.16) through (3.20) we picked the values 0.86, 0.40, 0.92, 0.85, and 0.50 as the initial values for $P(X_i = 1), i = 1, \ldots, 5$, respectively. Table 3.1 shows the estimates by a GS. By Table 2.1 and Table

3.1, we know that $MSE(\hat{P})$s of the estimates by a GS are smaller than those by an UGS. □

In this example, the estimates under a GS look more efficient than those under an UGS. Though it is complicated to determine the initial values under a GS, the GS seems to produce better estimates than the UGS.

### 3.3 A geometric investigation on the selection of initial values

In this subsection, we restrict our attention to basic structure 1 in Figure 2.1 and to investigate geometrically the efficiency of estimates under a GS. For notational convenience, we let $\theta_1 = P(X_1 = 1)$, $\theta_2 = P(X_2 = 1|X_1 = 1)$ and $\theta_3 = P(X_2 = 1|X_1 = 0)$. We will consider a graph of $\theta_1$ and $\theta_2$ with $\theta_3$ fixed at 0.1.

**Table 3.1.** Actual probabilities and probability estimates for Model 1 under the Guided Selection.

| | | GS | | |
|---|---|---|---|---|
| | $P$ | $P^{(0)}$ | $\hat{P}$ | $MSE(\hat{P})$ |
| $P(X_1 = 1)$ | .896 | .860 | .877 | .0003 |
| $P(X_6 = 1\|X_1 = 0)$ | .100 | .079 | .085 | .0002 |
| $P(X_6 = 1\|X_1 = 1)$ | .801 | .850 | .820 | .0003 |
| $P(X_2 = 1)$ | .400 | .400 | .402 | .0000 |
| $P(X_7 = 1\|X_1 = 0, X_2 = 0)$ | .000 | .001 | .003• | .0000 |
| $P(X_7 = 1\|X_1 = 0, X_2 = 1)$ | .098 | .100 | .231 | .0181 |
| $P(X_7 = 1\|X_1 = 1, X_2 = 0)$ | .101 | .100 | .099 | .0000 |
| $P(X_7 = 1\|X_1 = 1, X_2 = 1)$ | .797 | .800 | .797 | .0000 |
| $P(X_3 = 1)$ | .900 | .920 | .909 | .0001 |
| $P(X_8 = 1\|X_3 = 0)$ | .106 | .100 | .073 | .0010 |
| $P(X_8 = 1\|X_3 = 1)$ | .901 | .900 | .896 | .0000 |
| $P(X_4 = 1\|X_3 = 0)$ | .851 | .735 | .789 | .0039 |
| $P(X_4 = 1\|X_3 = 1)$ | .898 | .860 | .869 | .0008 |
| $P(X_9 = 1\|X_3 = 0, X_4 = 0)$ | .000 | .001 | .008• | .0001 |
| $P(X_9 = 1\|X_3 = 0, X_4 = 1)$ | .206 | .200 | .171 | .0011 |
| $P(X_9 = 1\|X_3 = 1, X_4 = 0)$ | .200 | .200 | .226 | .0008 |
| $P(X_9 = 1\|X_3 = 1, X_4 = 1)$ | .899 | .900 | .914 | .0002 |
| $P(X_5 = 1)$ | .398 | .500 | .467 | .0047 |
| $P(X_{10} = 1\|X_5 = 0)$ | .100 | .010 | .048 | .0028 |
| $P(X_{10} = 1\|X_5 = 1)$ | .850 | .800 | .799 | .0026 |

NOTE: A bullet (•) is attached when $\hat{P}$ is not contained in the corresponding probability interval at the significance level 0.05.

$P(X_2 = 1)$ can be expressed as

$$P(X_2 = 1) = \theta_1\theta_2 + (1 - \theta_1)\theta_3,$$

and so $\theta_2$ is given by

$$\theta_2 = (P(X_2 = 1) - \theta_3)/\theta_1 + \theta_3.$$

Next we will have a close look at the location of the estimates by GSs and UGSs.

**Example 3.1** We will consider basic structure 1 where $P(X_2 = 1) = 0.82$ and compare the GS and the UGS in estimating $\theta_1$ and $\theta_2$ with $\theta_3$ fixed to 0.1. The shaded area in Figure 3.1 is a Guided Selection Area(GSA). The GSA is an area which represents selection intervals of $\theta_1$ under $0.1 < \theta_2 \leq 1$ and $\theta_3 = 0.1$.

Table 3.2 shows the MLEs corresponding to the GS and the UGS. In addition, Figure 3.2 shows the 3-dimensional graph of $(\theta_1, \theta_2, \theta_3)$. The MLE points $\boxed{1}$, $\boxed{2}$, and $\boxed{3}$ by the GSs are located near the actual value $\boldsymbol{\theta}' = (\theta_1, \theta_2, \theta_3)$ = (0.9, 0.901, 0.106), but the MLE points $\boxed{4}$, $\boxed{5}$, $\boxed{6}$ and $\boxed{7}$ by the UGSs are scattered around the actual values. □

**Table 3.2.** The MLEs $\hat{\theta}$ from the initial values $\theta^{(0)}$ by the GSs and the UGSs when $P(X_2 = 1) = 0.8215$ under basic structure 1. The initial values for $\theta_3$ is fixed to 0.1.

| | $\theta_1^{(0)}$ | $\theta_2^{(0)}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | Point of the Coordinates |
|---|---|---|---|---|---|---|
| | .97 | .85 | .969 | .844 | .097 | $\boxed{1}$ |
| GS | .92 | .90 | .913 | .890 | .094 | $\boxed{2}$ |
| | .87 | .95 | .857 | .943 | .090 | $\boxed{3}$ |
| | .30 | .90 | .659 | .988 | .497 | $\boxed{4}$ |
| UGS | .70 | .70 | .852 | .908 | .322 | $\boxed{5}$ |
| | .90 | .70 | .942 | .857 | .229 | $\boxed{6}$ |
| | .30 | .30 | .504 | .911 | .729 | $\boxed{7}$ |

## 3.4 Asymptotic distribution for parameter estimates

### 3.4.1 The asymptotic distribution of the parameter estimates

This subsection presents the asymptotic theory of parametric models for categorical data. The approach is well described in Rao(1973), Bishop, Fienberg and Holland(1975), and Agresti(1990). The following results are the fundamental results of the large-sample model-based inference for categorical data. The key tool is the delta method.

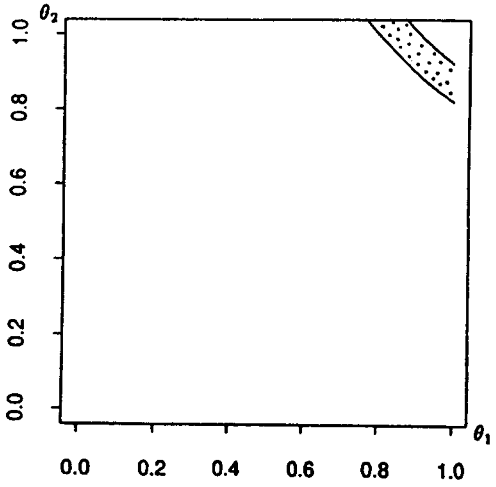


**Figure 3.1.** A GSA for $(\theta_1, \theta_2)$ with $\theta_3 = 0.1$.

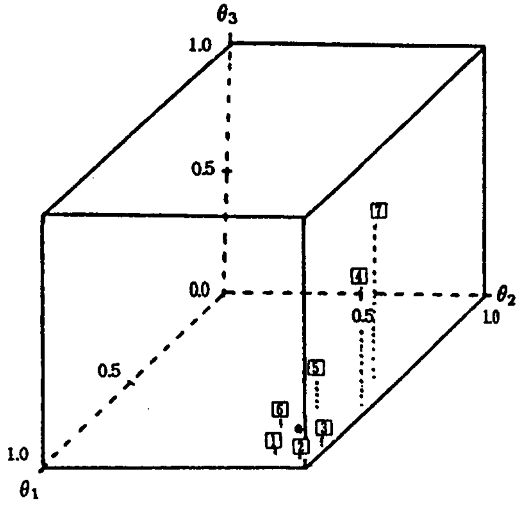


**Figure 3.2.** The 3-dimensional graph of $(\theta_1, \theta_2, \theta_3)$. The boxed labels are as in Table 3.2. The actual point (•) of $(\theta_1, \theta_2, \theta_3)$ is $(.900, .901, .106)$.

The data are counts $(n_1, \ldots, n_N)$ in $N$ cells of a contingency table. The asymptotics regard $N$ as fixed and let $n = \Sigma n_i \to \infty$. Suppose the cell counts have a multinomial distribution with the cell probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)'$, where V$'$ denotes the transpose of the matrix on vecter V. Let $\mathbf{p} = (p_1, \ldots, p_N)'$ denote the sample proportions, where $p_i = n_i/n$. The model relates $\boldsymbol{\pi}$ to a smaller number of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_t)'$. We express it as $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, and $\pi_i(\boldsymbol{\theta})$ denotes the function that relates parameters to $\pi_i, i = 1, \ldots, N$. We use $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ to denote generic parameter and probability values, and $\boldsymbol{\theta}_0 = (\theta_{10}, \ldots, \theta_{t0})'$ and $\boldsymbol{\pi}_0 = (\pi_{10}, \ldots, \pi_{N0})' = \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ to denote the true values for a particular application.

Consider the asymptotic distribution of the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and the asymptotic distribution of the model-based MLE $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})$. The kernel of the multinomial log-likelihood is

$$L(\boldsymbol{\theta}) = \log \prod_{i=1}^{N} \pi_i(\boldsymbol{\theta})^{n_i} = n \sum_{i=1}^{N} p_i \log \pi_i(\boldsymbol{\theta}).$$

Let A denote the $N \times t$ matrix

$$A = \left[ \ \pi_{i0}^{-1/2} \left( \frac{\partial \pi_i(\boldsymbol{\theta})}{\partial \theta_{j0}} \right) \ \right] = Diag(\boldsymbol{\pi}_0)^{-1/2} \left( \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\theta}_0} \right).$$

Then we obtain the important result

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, (A'A)^{-1}),$$

where $\sqrt{n}(\mathbf{p} - \boldsymbol{\pi}_0) \xrightarrow{d} N(0, Diag(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0')$
and

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \xrightarrow{d} N\left( \ 0, \left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\theta}_0}\right)(A'A)^{-1}\left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\theta}_0}\right)' \ \right).$$

### 3.4.2 Asymptotic distributions for basic structures

The asymptotic distribution of the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is given by

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N[0, (A'A)^{-1}],$$

where $\boldsymbol{\theta}_0$ is the true value for a particular application. We consider only basic structure 1 and basic structure 4.

basic structure 1

Let the cell probabilities for basic structure 1 be denoted by, for $x_1$, $x_2$ = 0,1,

$$\pi_{x_1 x_2} = P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) P(X_2 = x_2 | X_1 = x_1),$$

where $\sum_{x_1}\sum_{x_2}\pi_{x_1x_2}=1$. Then $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are given as follows :

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} P(X_1=1) \\ P(X_2=1|X_1=1) \\ P(X_2=1|X_1=0) \end{pmatrix},$$

$$\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}) = \begin{pmatrix} \pi_{00}(\boldsymbol{\theta}) \\ \pi_{01}(\boldsymbol{\theta}) \\ \pi_{10}(\boldsymbol{\theta}) \\ \pi_{11}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} (1-\theta_1)(1-\theta_3) \\ (1-\theta_1)\theta_3 \\ \theta_1(1-\theta_2) \\ \theta_1\theta_2 \end{pmatrix}.$$

Then we obtain

$$(A'A)^{-1} = \begin{pmatrix} \theta_{10}(1-\theta_{10}) & 0 & 0 \\ 0 & \theta_{20}(1-\theta_{20})/\theta_{10} & 0 \\ 0 & 0 & \theta_{30}(1-\theta_{30})/(1-\theta_{10}) \end{pmatrix}.$$

basic structure 4

Let the cell probabilities for basic structure 4 be given by

$$\pi_{x_1x_3} = P(X_1=x_1, X_3=x_3) = P(X_1=x_1)P(X_3=x_3|X_1=x_1),$$

$$\begin{aligned} \pi_{x_1x_2x_4} &= P(X_1=x_1, X_2=x_2, X_4=x_4) \\ &= P(X_1=x_1, X_2=x_2)P(X_4=x_4|X_1=x_1, X_2=x_2), \end{aligned}$$

for $x_i=0,1,\ i=1,2,3,4$, where $\sum_{x_1}\sum_{x_3}\pi_{x_1x_3} = \sum_{x_1}\sum_{x_2}\sum_{x_4}\pi_{x_1x_2x_4} = 1$. As $\pi_{x_1x_3}$ is similar to basic structure 1, consider $\pi_{x_1x_2x_4}$ only. Then $\theta$ and $\pi$ are given as follows :

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \end{pmatrix} = \begin{pmatrix} P(X_1=1, X_2=1) \\ P(X_1=1, X_2=0) \\ P(X_1=0, X_2=1) \\ P(X_4=1|X_1=1, X_2=1) \\ P(X_4=1|X_1=1, X_2=0) \\ P(X_4=1|X_1=0, X_2=1) \\ P(X_4=1|X_1=0, X_2=0) \end{pmatrix},$$

$$\boldsymbol{\pi} \;=\; \boldsymbol{\pi}(\boldsymbol{\theta}) = \begin{pmatrix} \pi_{000}(\boldsymbol{\theta}) \\ \pi_{001}(\boldsymbol{\theta}) \\ \pi_{010}(\boldsymbol{\theta}) \\ \pi_{011}(\boldsymbol{\theta}) \\ \pi_{100}(\boldsymbol{\theta}) \\ \pi_{101}(\boldsymbol{\theta}) \\ \pi_{110}(\boldsymbol{\theta}) \\ \pi_{111}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} (1-\theta_1-\theta_2-\theta_3)(1-\theta_7) \\ (1-\theta_1-\theta_2-\theta_3)\theta_7 \\ \theta_3(1-\theta_6) \\ \theta_3\theta_6 \\ \theta_2(1-\theta_5) \\ \theta_2\theta_5 \\ \theta_1(1-\theta_4) \\ \theta_1\theta_4 \end{pmatrix}.$$

Then we obtain

$$(A'A)^{-1} = diag \begin{pmatrix} \theta_{10}(1-\theta_{10}-\theta_{20}-\theta_{30}) \\ \theta_{20}(1-\theta_{10}-\theta_{20}-\theta_{30}) \\ \theta_{30}(1-\theta_{10}-\theta_{20}-\theta_{30}) \\ \theta_{40}(1-\theta_{40})/\theta_{10} \\ \theta_{50}(1-\theta_{50})/\theta_{20} \\ \theta_{60}(1-\theta_{60})/\theta_{30} \\ \theta_{70}(1-\theta_{70})/\{1-\theta_{10}-\theta_{20}-\theta_{30}\} \end{pmatrix}.$$

### 3.4.3 Application

In the example below, we will see, by applying the asymptotic results derived above, how the GS works in comparison with the UGS confined to basic structure 1.

**Example 3.2**

Here, the true values of vectors $\boldsymbol{\theta}$ and $\boldsymbol{\pi}(\boldsymbol{\theta})$ for basic structure 1 are as follows:

$$\boldsymbol{\theta}_0 \;=\; (\theta_{10}, \theta_{20}, \theta_{30})' = (0.900, 0.901, 0.106)',$$

$$\boldsymbol{\pi}_0 \;=\; \boldsymbol{\pi}(\boldsymbol{\theta}_0) = \begin{pmatrix} (1-\theta_{10})(1-\theta_{30}) \\ (1-\theta_{10})\theta_{30} \\ \theta_{10}(1-\theta_{20}) \\ \theta_{10}\theta_{20} \end{pmatrix} = \begin{pmatrix} 0.0894 \\ 0.0106 \\ 0.0891 \\ 0.8109 \end{pmatrix}.$$

The asymptotic distribution for $\hat{\boldsymbol{\theta}}$ is given by

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N[0, (A'A)^{-1}],$$

where

$$(A'A)^{-1} = \begin{pmatrix} 0.09 & 0 & 0 \\ 0 & 0.09911 & 0 \\ 0 & 0 & 0.94764 \end{pmatrix}.$$

From the above asymptotic results, we obtain the probability intervals of $\boldsymbol{\theta}$ at the significance level $\alpha = 0.05$ and with the sample size $n = 100$. $z_\alpha$ denotes the upper $100\alpha$ percentile point.

**Table 3.3.** MLEs from GS and UGS given the initial point $\theta_3^{(0)} = 0.1$.

| | $\theta_1^{(0)}$ | $\theta_2^{(0)}$ | $\hat\theta_1$ | $\hat\theta_2$ | $\hat\theta_3$ |
|---|---|---|---|---|---|
| GS | 0.94 | 0.87 | 0.939 | 0.868 | 0.101 |
| | 0.91 | 0.90 | 0.906 | 0.895 | 0.099 |
| | 0.90 | 0.93 | 0.885 | 0.916 | 0.087 |
| | 0.94 | 0.88 | 0.936 | 0.871 | 0.095 |
| | 0.95 | 0.89 | 0.941 | 0.867 | 0.084 |
| | 0.92 | 0.94 | 0.891 | 0.913 | 0.071 |
| | 0.92 | 0.90 | 0.913 | 0.890 | 0.094 |
| | 0.86 | 0.94 | 0.859 | 0.939 | 0.101 |
| | 0.94 | 0.89 | 0.932 | 0.874 | 0.089 |
| | 0.95 | 0.93 | 0.923 | 0.884 | 0.061 |
| UGS | 0.60 | 0.80 | 0.802* | 0.944 | 0.322* |
| | 0.80 | 0.58 | 0.903 | 0.871 | 0.356* |
| | 0.77 | 0.95 | 0.824* | 0.966∗ | 0.143 |
| | 0.95 | 0.80 | 0.960* | 0.849 | 0.138 |
| | 0.30 | 0.80 | 0.650* | 0.976* | 0.533* |
| | 0.50 | 0.55 | 0.753* | 0.921 | 0.518* |
| | 0.90 | 0.25 | 0.943 | 0.832* | 0.629* |
| | 0.10 | 0.60 | 0.335* | 0.975* | 0.743* |
| | 0.70 | 0.10 | 0.699* | 0.812* | 0.822* |
| | 0.60 | 0.15 | 0.672* | 0.844 | 0.774* |

NOTE: An asterisk ($*$) is attached when $\hat{P}$ is not contained in the corresponding probability interval at the significance level 0.05.

(1) Probability interval of $\theta_1$ is $0.8412 < \hat\theta_1 < 0.9588$,

where $|\hat\theta_1 - \theta_{10}| < z_{\frac{\alpha}{2}}\sqrt{\frac{\theta_{10}(1-\theta_{10})}{n}}$.

(2) Probability interval of $\theta_2$ is $0.8393 < \hat\theta_2 < 0.9627$,

where $|\hat\theta_2 - \theta_{20}| < z_{\frac{\alpha}{2}}\sqrt{\frac{\theta_{20}(1-\theta_{20})}{n\theta_{10}}}$.

(3) Probability interval of $\theta_3$ is $0 \le \hat\theta_3 < 0.2968$,

where $|\hat\theta_3 - \theta_{30}| < z_{\frac{\alpha}{2}}\sqrt{\frac{\theta_{30}(1-\theta_{30})}{n(1-\theta_{10})}}$. □

Based on the above probability intervals, we will discuss the efficiency of the estimates under the GSs and the UGSs with $\theta_3^{(0)}$ fixed at 0.1. When $P(X_2 = 1) = 0.8215$, the values between 0.8215 and 0.95 look reasonably good choices for the initial values of $\theta_2$ under a GS. From the GSA in Figure 3.1, we randomly selected the 10 initial points $(\theta_1^{(0)}, \theta_2^{(0)})$, given $\theta_3^{(0)}$=0.1. Table 3.3 shows that the MLEs with the GS are efficient, while it is not the case as for the UGS.

**Table 3.4.** MLEs from the initial values in regions $R_1$ through $R_4$ of Figure 3.3.

| Region | $\theta_1^{(0)}$ | $\theta_2^{(0)}$ | $\hat\theta_1$ | $\hat\theta_2$ | $\hat\theta_3$ |
|---|---|---|---|---|---|
|  | 0.77 | 0.95 | 0.824* | 0.966* | 0.143 |
|  | 0.80 | 0.93 | 0.842 | 0.949 | 0.138 |
| $R_1$ | 0.85 | 0.83 | 0.885 | 0.909 | 0.145 |
|  | 0.90 | 0.83 | 0.923 | 0.878 | 0.144 |
|  | 0.95 | 0.80 | 0.960* | 0.849 | 0.138 |
|  | 0.50 | 0.99 | 0.747* | 0.997* | 0.301* |
|  | 0.55 | 0.90 | 0.773* | 0.972 | 0.305* |
|  | 0.60 | 0.80 | 0.802* | 0.944 | 0.322* |
|  | 0.65 | 0.75 | 0.827* | 0.926 | 0.319* |
| $R_2$ | 0.70 | 0.70 | 0.852 | 0.908 | 0.322* |
|  | 0.75 | 0.60 | 0.880 | 0.883 | 0.364* |
|  | 0.80 | 0.58 | 0.903 | 0.871 | 0.356* |
|  | 0.85 | 0.55 | 0.927 | 0.857 | 0.359* |
|  | 0.90 | 0.50 | 0.952 | 0.843 | 0.381* |
|  | 0.95 | 0.47 | 0.960* | 0.832* | 0.389* |
|  | 0.25 | 0.90 | 0.621* | 0.990* | 0.545* |
|  | 0.30 | 0.80 | 0.650* | 0.976* | 0.533* |
|  | 0.40 | 0.65 | 0.702* | 0.947 | 0.523* |
|  | 0.50 | 0.55 | 0.753* | 0.921 | 0.518* |
| $R_3$ | 0.60 | 0.45 | 0.800* | 0.893 | 0.534* |
|  | 0.70 | 0.35 | 0.843* | 0.867 | 0.576* |
|  | 0.80 | 0.30 | 0.892 | 0.848 | 0.595* |
|  | 0.90 | 0.25 | 0.943 | 0.832* | 0.629* |
|  | 0.10 | 0.60 | 0.335* | 0.975* | 0.743* |
|  | 0.20 | 0.45 | 0.456* | 0.948* | 0.714* |
|  | 0.30 | 0.35 | 0.532* | 0.920* | 0.708* |
| $R_4$ | 0.40 | 0.30 | 0.606* | 0.899* | 0.701* |
|  | 0.50 | 0.20 | 0.629* | 0.866* | 0.745* |
|  | 0.60 | 0.15 | 0.672* | 0.844* | 0.774* |
|  | 0.70 | 0.10 | 0.699* | 0.812* | 0.822* |

NOTE: An asterisk(∗) is attached when $\hat{P}$ is not contained in the corresponding probability interval at the significance level 0.05.

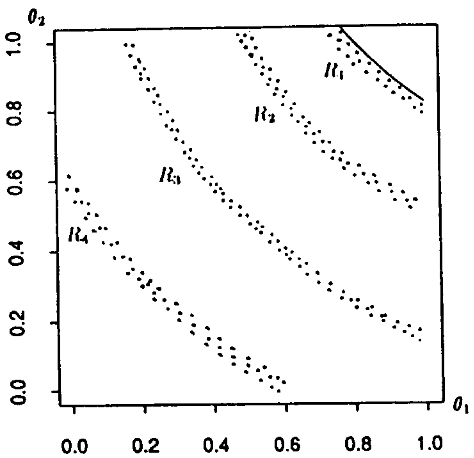


**Figure 3.3.** The plot of the four regions $R_1, R_2, R_3$, and $R_4$. The solid line is the graph of $\theta_2 = 0.7215/\theta_1 + 0.1$.

Next, we will look at the MLEs for each of the regions in Figure 3.3. Table 3.4 shows estimates obtained with the initial values lying in regions $R_1$ through $R_4$ in Figure 3.3, respectively. The region $R_1$ shows that, though the initial values $\theta_1^{(0)}$ and $\theta_2^{(0)}$ stay in the probability interval $0.05 \leq |\theta_1 - \theta_1^{(0)}| \leq 0.13$ of $\theta_1$ and the probability interval $0.049 \leq |\theta_2 - \theta_2^{(0)}| \leq 0.101$ of $\theta_2$, respectively, some of the estimates fall beyond the corresponding probability intervals.

To sum up, we have seen that the estimates starting from the initial values which stay away from the actual values fall outside the corresponding probability intervals. Hence, we may safely say that the selection of the initial values near the region of the curve ($\theta_2 = 0.7215/\theta_1 + 0.1$) yield good final estimates, and that the initial values by the GS yield even better estimates.

## 4. CONCLUDING REMARKS

We explored a methodology for selecting the initial values for the EM algorithm. It is assumed that the proportions correct of items are available. The general outline of the methodology follows. First of all, the initial values for the conditional probabilities, $\alpha$'s, of the item score variable given latent variables are determined by using the proportions correct of items and by consulting experts or from experience. Secondly, we derive selection rules for the initial values for probabilities of abilities using those initial values for $\boldsymbol{\alpha}$'s and the proportions correct of items. The simulation results strongly indicate that the GS deserves our attention when we deal with the initial values for EM. The geometric investigation of GSAs has given us an insight into how the initial values affect the final estimates. Structures such as basic structure 1 are more sensitive to the initial values than the other basic structures, so special cares are in need in selecting initial values for such structures.

In a nutshell, we recommend that, when fitting a recursive model with latent variables, several EM repetitions be tried with different initial values that are obtained by the GS and that we can then choose the most reasonable-looking set from the collection of sets of the final estimates. When the model is relatively small or the structure is relatively simple, a single EM may be enough; but when the model structure is relatively complex we have to exercise our discretion in applying the GS to obtain a set of reasonably good final estimates in as few number of EM trials as possible. The number of EM trials until success seems to depend mainly upon the model complexity and the appropriateness of the initial values.

In applying the EM for the recursive models, the experts' opinions play important roles, as illustrated in Remarks 3.4 and 3.5, in selecting the initial values. The guided selection can result in a success only when the data and the experts' opinions are fully incorporated in the selection process for the initial values.

In this paper we have not concerned ourselves about the convergence rate of EM. The convergence rate seems to be more related to the shape of a given likelihood function as a multivariate real valued function. The slope of the graph near a local maximum point to which the EM estimate points approach determines the convergence rate. Interested readers in this issue are refered to Van Dyk and Meng (1997) and literature cited therein.

In the research area of educational testing and evaluation, test item developers are usually requested to give their anticipated proportions of correct responses to the items they made. Experienced item developers would give their proportions correct close to the true. And so eliciting their opinions on the $\alpha$-values of section 3 is simply a refined version of eliciting for the proportions correct.

A main idea behind the guided selection is that we try to incorporate experts' opinions in the estimation process so that the final estimates are a balanced reflection of experts' opinions and data.

## APPENDIX

Theorems 3.3 and 3.4 are proved in this section.

**Proof of Theorem 3.3** :

The proof for $P(X_1 = 1)$ is identical to that of Theorem 3.1. Since the probability of a correct response to item 2 is

$$P(X_4 = 1) = \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} P(X_1 = x_1)P(X_2 = x_2)P(X_4 = 1|X_1 = x_1, X_2 = x_2),$$

$P(X_2 = 1)$ can be written as

$$P(X_2 = 1) = A/B, \tag{A.1}$$

where

$$\begin{aligned} A \;=\; & P(X_4 = 1) - P(X_1 = 1)\{P(X_4 = 1|X_1 = 1, X_2 = 0) \\ & - P(X_4 = 1|X_1 = 0, X_2 = 1)\} - P(X_4 = 1|X_1 = 0, X_2 = 0) \end{aligned}$$

and

$$\begin{aligned} B \;=\; & P(X_4 = 1|X_1 = 0, X_2 = 1) + P(X_1 = 1)\{P(X_4 = 1|X_1 = 1, X_2 = 1) \\ & - P(X_4 = 1|X_1 = 0, X_2 = 1) - P(X_4 = 1|X_1 = 1, X_2 = 0) \\ & + P(X_4 = 1|X_1 = 0, X_2 = 0)\} - P(X_4 = 1|X_1 = 0, X_2 = 0). \end{aligned}$$

Let

$$\delta_1 = P(X_4 = 1|X_1 = 1, X_2 = 1) - P(X_4 = 1|X_1 = 0, X_2 = 1), \tag{A.2}$$

$$\delta_2 = P(X_4 = 1|X_1 = 1, X_2 = 0) - P(X_4 = 1|X_1 = 0, X_2 = 0). \tag{A.3}$$

Substituting (A.2) and (A.3) into (A.1) and using the 3 equations in (3.11), we have

$$P(X_2 = 1) = \frac{P(X_4 = 1) - \delta_2 P(X_1 = 1) - \alpha_{20}}{\alpha_{21} + (\delta_1 - \delta_2)P(X_1 = 1) - \alpha_{20}}.$$

Then we have, if $\alpha_{22} - 2\alpha_{21} + \alpha_{20} \geq 0$, by inequality (3.12) for $P(X_1 = 1)$,

$$\begin{aligned} \frac{P(X_4 = 1) - \delta_2 u_1 - \alpha_{20}}{\alpha_{21} + (\delta_1 - \delta_2)u_1 - \alpha_{20}} \;\leq\; & P(X_2 = 1) && \text{(A.4)} \\ \leq\; & \frac{P(X_4 = 1) - \delta_2 l_1 - \alpha_{20}}{\alpha_{21} + (\delta_1 - \delta_2)l_1 - \alpha_{20}}, && \text{(A.5)} \end{aligned}$$

and if $\alpha_{22} - 2\alpha_{21} + \alpha_{20} < 0$, then $l_1$ and $u_1$ in denominators for (A.4) and (A.5) are interchanged. From the conditions (3.11), we have our result.

**Proof of Theorem 3.4** :

The proof for $P(X_1 = 1)$ is the same as that of Theorem 3.1. The probability of a correct response to item 2 is

$$\begin{aligned} P(X_4 = 1) &= \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} P(X_1 = x_1)P(X_2 = x_2 | X_1 = x_1) \\ & \quad \times P(X_4 = 1 | X_1 = x_1, X_2 = x_2). \qquad \text{(A.6)} \end{aligned}$$

Let $\varepsilon_1 = P(X_2 = 1|X_1 = 1) - P(X_2 = 1)$. By applying (3.9), (3.10), (3.11) and (3.12) for $P(X_1 = 1)$ to equation (A.6), we have

$$\begin{aligned} P(X_4 = 1) &\leq P(X_2 = 1)\{\alpha_{21} + u_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}\} \\ & \quad + u_1(\alpha_{21} - \alpha_{20}) + \alpha_{20} + \varepsilon_1 u_1(\alpha_{22} - \alpha_{21}). \end{aligned}$$

Then we have

$$\frac{P(X_4 = 1) - u_1(\alpha_{21} - \alpha_{20}) - \alpha_{20} - \varepsilon_1 u_1(\alpha_{22} - \alpha_{21})}{\alpha_{21} + u_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}} \leq P(X_2 = 1). \qquad \text{(A.7)}$$

Let $\varepsilon_2 = P(X_2 = 1) - P(X_2 = 1|X_1 = 0)$. Then by applying (3.9), (3.10), (3.11) and (3.12) for $P(X_1 = 1)$ to equation (A.6), we have

$$\begin{aligned} P(X_4 = 1) &\geq P(X_2 = 1)\{\alpha_{21} + l_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}\} \\ & \quad + l_1(\alpha_{21} - \alpha_{20}) + \alpha_{20} - \varepsilon_2(1 - l_1)(\alpha_{21} - \alpha_{20}). \end{aligned}$$

So the following holds:

$$P(X_2 = 1) \leq \frac{P(X_4 = 1) - l_1(\alpha_{21} - \alpha_{20}) - \alpha_{20} + \varepsilon_2(1 - l_1)(\alpha_{21} - \alpha_{20})}{\alpha_{21} + l_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}}. \qquad \text{(A.8)}$$

Now from (A.7) and (A.8) follows that

$$C \leq P(X_2 = 1) \leq D,$$

where

$$C = \frac{P(X_4 = 1) - u_1(\alpha_{21} - \alpha_{20}) - \alpha_{20} - \varepsilon_1 u_1(\alpha_{22} - \alpha_{21})}{\alpha_{21} + u_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}}$$

$$D = \frac{P(X_4 = 1) - l_1(\alpha_{21} - \alpha_{20}) - \alpha_{20} + \varepsilon_2(1 - l_1)(\alpha_{21} - \alpha_{20})}{\alpha_{21} + l_1(\alpha_{22} - 2\alpha_{21} + \alpha_{20}) - \alpha_{20}}.$$

The interval $(l_2, u_2)$ of (3.13) is contained in the interval $(C, D)$. If $\varepsilon_1 = 0$ and $\varepsilon_2 = 0$, $C$ equals $l_2$ and $D$ equals $u_2$, where $l_2$ and $u_2$ are the values in (3.13). If $P(X_2 = 1)$ belongs to the interval $(l_2, u_2)$, then it belongs to the interval $(C, D)$. Thus the result follows.

## ACKNOWLEDGEMENT

The authors are grateful to the referees for their thoughtful and invaluable comments and suggestions.

## REFERENCES

(1) Agresti, A.(1990), *Categorical Data Analysis.* New York: John Wiley & Sons.

(2) Bentler, P.M.(1985), *Theory and Implementation of EQS: A Structural Equations Program.* Los Angeles: BMDP Statistical Software.

(3) Bishop, Y.M., Fienberg, S.E., and Holland, P.W.(1975), *Discrete Multivariate Analysis: Theory and Practice.* Sixth printing. Cambridge, MA: MIT Press.

(4) Bollen, K.A.(1989), *Structural Equations with Latent Variables.* NY: John Wiley & Sons.

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

(6) Efron, B.(1978). The geometry of exponential families. *The Annals of Statistics*, **6**, 2, 362-376.

(7) Fienberg, S. E.(1980), *The Analysis of Cross-Classified Categorical Data.* 2nd ed. Cambridge, MA: MIT Press.

(8) Fisher, R. A.(1925), Theory of statistical estimation. *Proc. Cambridge Philos. Trans.* **122**, 700-725.

(9) Greeno, J.G. and Simon, H.A.(1988), Problem Solving and Reasoning. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey, and R.D. Luce(Eds.), *Stevens' handbook of experimental psychology*(2nd ed.), Vol. II. New York: John Wiley. 1988, 589-672.

(10) Jöreskog, K.G. and Sörbom, D.(1986), *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Method.* Mooresville, IN: Scientific Software, Inc.

(11) Kim S.- H.(1994), An Approach to Statistical Modelling for Task Ability and Task Performance(in Korean). *Korean Educational Evaluation Research*, 7, 2.

(12) Kim S.- H.(1997), Iterative Proportional Fitting for Nonhierarchical Loglinear Models. *Communications in Statistics*, **26**, 6. 1443-1460.

(13) Lauritzen, S.L.(1995). The EM Algorithm for graphical association models with missing Data. *Comp. Stat. & Data Anal.* **19**, 191-201.

(14) Lauritzen, S.L. and Wermuth, N.(1983), Graphical and Recursive Models for Contingency Tables. *Biometrika*, **70**, 3, 537-552.

(15) Oliver, R.M. and Smith, J.Q.(1990), *Influence Diagrams, Belief Nets and Decision Analysis*(edition). NY: John Wiley & Sons.

(16) Pearl, J.(1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* CA: Morgan Kaufmann Publishers, Inc.

(17) Rao, C.R.(1973), *Linear Statistical Inference and Its Applications.* 2nd ed. New York: Wiley.

(18) Van Dyk, D. and Meng, X.L.(1997). On the ordering and groupings of conditional maximizations within ECM-type Algorithms. *Journal of Computational and Graphical Statistics*, **6**, 2, 202-223.

(19) Whittaker, J.(1990), *Graphical Models in Applied Multivariate Statistics.* Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.

(20) Wu, C.F.(1983), On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**, 1, 95-103.