Proc. Natl. Sci. Counc. ROC(A)
Vol. 23, No. 6, 1999. pp. 695-702

Finite Mixture Multivariate Generalized Linear Models Using Gibbs Sampling and E-M Algorithms

CHIH-CHIEN YANG*, BENGT O. MUTHÉN**, AND CHIH-CHIANG YANG***

**Graduate School of Educational Measurement and Statistics*
National Taichung Teachers' College
Taichung, Taiwan, R.O.C.
***Department of Education*
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA, U.S.A.
****The Overseas Chinese College of Commerce*
Taichung, Taiwan, R.O.C.

(Received November 10, 1998; Accepted May 24, 1999)

ABSTRACT

Finite mixture multivariate generalized linear modeling has been shown to be an important analytic tool for many research fields, for example, image recognition, astronomical data classification, biomedicine diagnosis, and biological classification. Recent statistical and computational advances have further encouraged researchers to explore the modeling possibility using the Bayesian framework. We compare Expectation (E)-Maximization (M) algorithms for maximum likelihood estimation of classical statistics with Gibbs sampling methods of Bayesian statistics in estimating finite mixture multivariate generalized linear models. A Monte Carlo study to compare the two methods is provided for practical reference. We also propose two finite mixture multivariate generalized linear models that can allow more flexibility in modeling substantive applications. The Longitudinal Study of American Youth (LSAY) data set is also analyzed as a practical application.

Key Words: finite mixture models, multivariate generalized linear models, Gibbs sampling, E-M algorithms, Bayesian theory

I. Introduction

The study of finite mixture multivariate generalized linear models is one of the significant research topics which has drawn much attention from researchers in recent years for both statistical and substantive reasons. In particular, the significance of finite mixture models has been re-emphasized recently in various practical research areas, including image recognition (Murtagh and Raftery, 1984; Campbell *et al*., 1997), astronomical data classification (Celeux and Govaert, 1995; Mukerjee *et al*., 1998), biomedicine diagnosis (Melton *et al*., 1994; Pickering and Forbes, 1984; Yang and Becker, 1997) and biological classification (Wang and Puterman, 1998). The major reason for the popularity of finite mixture modeling is that recent statistical advances have made it feasible to fit finite mixture models in a wide range of applications. Thanks to these statistical developments, these models now can deal with continuous outcomes (Jedidi *et al*., 1997; Yung, 1997), discrete outcomes (Qu *et al*., 1996; Brooks *et al*., 1997; Wang *et al*., 1996; Wang and Puterman, 1998), and combinations of continuous and discrete outcomes (Muthén *et al*., 1996; Muthén and Shedden, 1999) with more realistic modeling assumptions.

The proposed finite mixture multivariate generalized linear models can cover a wide range of overdispersed generalized linear models which can often be seen in many substantive applications, for example, longitudinal research or repeated measurement studies. These overdispersed models are usually analyzed by employing mixed-effects approaches (Breslow and Clayton, 1993). The mixed-effects approach assumes an infinite mixture of fixed and random effects which are included to account for the non-explainable random effects or the extra-variations within the sample. According to various distributional assumptions, the random effects may result in beta-binomial distributions (e.g., extra-binomial regression models), negative binomial distributions (e.g., extra-Poisson regression models), or non-tractable forms. In contrast, finite mixture models assume that a set of different distributions is present in the outcome variables. When these distributions come from the same exponential family, this is equivalent to saying that at least two different sets of distributional (scale/position) parameters are in the sample. Therefore, statistical interpretations can indicate that there are non-homogeneous classes or groups in the sample. Obviously, finite mixture models are more appealing and interpretable than mixed-effects models in many applications. More importantly, these finite mixture models have important substantive meaning in a broad range of research areas.

The popularity of finite mixture modeling can be seen from the recent literature cited above; however, relatively little research has focused on finite mixture multivariate generalized linear models using Bayesian theory. Most of the papers in the literature have used the maximum likelihood estimation theory and Expectation (E)-Maximization (M) algorithms (Dempster *et al*., 1977) for parameter estimation; therefore, most of them were within the classical statistics framework. One of the reasons for this is the computational complexity of Bayesian finite mixture models. Advances modern statistical computation, e.g., Gibbs sampling (Tanner, 1996), have helped solve these problems. It is very necessary to evaluate and compare the two estimation methods (E-M algorithm and Gibbs sampling) and provide a general reference for practical researchers. One of the two aims of this paper is to compare the E-M algorithms and Gibbs sampling from various computational and statistical viewpoints based on Monte Carlo studies.

The other major aim of this paper is to introduce two relatively new finite mixture multivariate generalized linear models. We demonstrate the validity of the two models using simulated and real data sets. For the purpose of simplicity, all the finite mixture models discussed in this paper are limited to two different distribution forms only although they can be easily extended to more than two distributions.

The organization of this paper is as follows. In Section I, we explain the significance and properties of finite mixture models and discuss the aims of this paper. In Section II, we review the finite mixture generalized linear models. Section III contains mathematical derivations of the two estimation methods for a basic finite mixture model. The model will be used in later Monte Carlo studies to compare the two methods. Two Bayesian finite mixture multivariate generalized linear models are introduced in Section IV. Section V presents the designs and results of the estimation method comparison and illustrates the examples using analysis of simulated and real data sets using the two Bayesian models. In the last section, conclusions based on comparisons between the different estimation methods and suggestions for the use of the proposed models are provided.

II. Finite Mixture Generalized Linear Models

We review the recent literature on finite mixture models for binomial, binary and Poisson outcomes in this section. The finite mixture models defined here contain conditional probabilities (CP) for the outcome probabilities, given that a certain distribution is true, and mixing probabilities (MP) for the mixing rates of the two distributions. For a set of non-repeated Poisson outcomes, Wang *et al*. (1996) proposed two finite mixture models that can deal with data sets that have or do not have independent variables. The independent variables are included as covariates for the conditional probabilities. They used an E-M algorithm combined with a modified quasi-Newton algorithm in the estimation procedures.

For a set of non-repeated binomial outcomes, Follmann and Lambert (1989) suggested finite mixture models that could handle binomial outcomes with or without covariates (for conditional probability). A non-parametric estimation method was proposed in their work.

One of the major classes of finite mixture models consists of the so-called latent class analysis (LCA) models. Typical LCA models have been studied for decades; see, for example, Goodman (1974), Bartholomew (1987) and many other applications. Recently, numerous researchers have proposed several extended versions of LCA (ELCA) models. For example, Qu *et al*. (1996) proposed several ELCA models for multivariate binary outcomes. Specifically, they were interested in studying the analyses of sensitivity and specificity of some medical diagnosis criteria (multivariate binary outcomes). Their ELCA models contain possible random effects, direct effects, or fixed effects for conditional probabilities. The parameter estimation procedures employed in their paper were the E-M algorithms and Gauss-Hermit quadrature approximations (if random effects are involved in the ELCA models). Wang and Puterman (1998) suggested a similar ELCA model without random effects for a non-repeated binomial data set. In particular, covariates were for both conditional probabilities and mixing probability in this model.

It seems that several new models have been proposed in the last few years; however, very few finite mixture multivariate generalized linear models have actually been applied using Bayesian theory. Thanks to recent statistical computation advances, the above and other useful models are also possible under the Bayesian framework if Gibbs sampling is used. We shall provide a general guideline which researchers can employ in using E-M algorithms or Gibbs sampling.

III. E-M Algorithm and Gibbs Sampling

The two competing estimation methods, the use of E-M algorithm and Gibbs sampling, are outlined in this section. The main idea behind a generic E-M algorithm is to implement an expectation to replace the "missing" or "unobservable" data in a statistical model; therefore, a complete data likelihood can be obtained using the observed data. By maximizing the complete data likelihood function, the maximum likelihood estimation (MLE) of the model parameters can be obtained. To simplify the comparison, we will use a typical LCA model for both methods. We will further describe the major steps of the E-M algorithm in the following. Similar E-M algorithms were also used in several studies, for example, those of Qu *et al*. (1996), Wang and Puterman (1998), and Yang (1998). Let $\pi_{kc}=\Pr(u_{ik}=1|C=c)$ be the conditional probability of the kth binary response (k=1, …, t) being 1 given that $C=c$ (c=0, 1). The probability that the ith subject will have outcomes $U_i=(u_{i1}, \ldots, u_{it})'$ is given by

$$g(U_i)=\sum_{c=0}^{1}\lambda_c g_c(U_i),$$

where $\lambda_c=\Pr(C=c)$ and $g_c(U_i)=\Pr(U_i|C=c)$.

Given the latent classes, conditional independence of t binary responses can be assumed. Therefore, $g_c(U_i)$ has the form

$$g_c(U_i)=\prod_{k=1}^{t}\pi_{kc}^{u_{ik}}(1-\pi_{kc})^{1-u_{ik}}.$$

Thus, the likelihood function for each subject in this model based on the conditional independence assumption is given by

$$g(U_i)=\sum_{c=0}^{1}\lambda_c\prod_{k=1}^{t}\pi_{kc}^{u_{ik}}(1-\pi_{kc})^{1-u_{ik}}.$$

The posterior probability of $C=c$ for the ith subject with binary outcomes U_i is given by

$$h_{ci}=\frac{\lambda_c g_c(U_i)}{\sum_{c=0}^{1}\lambda_c g_c(U_i)}.$$

Replacing the class indicator C with its expectation value (similar to a typical E-M algorithm for missing data) and using a set of reasonable starting values, the expectation (E-step) for C and the maximization (M-step) for the conditional probabilities and class proportions alternate until convergence occurs.

In the E-step, the posterior probabilities h_{ci} are evaluated from current parameter estimates. In the M-step, the class proportions λ_c are updated as

$$\hat{\lambda}_c=\frac{K_c}{\sum_{c=0}^{1}K_c},$$

where

$$K_c=\sum_{i=1}^{N}h_{ci},\ c=0,\ 1,$$

and the parameters in the component distributions are updated as

$$\hat{\pi}_{kc}=\frac{S_{kc}}{K_c},\ c=0,\ 1,\ k=1,\ 2,\ 3,\ \ldots,\ t,$$

where

$$S_{kc}=\sum_{i=1}^{N}h_{ci}u_{ik}.$$

The E-M algorithm was programmed using the FORTRAN 90 language.

Gibbs sampling has a much complicated estimation procedure; see, e.g., Tanner (1996) and references therein. Therefore, only specific mechanisms for the LCA model will be provided here. The main objective of Gibbs sampling is to synthesize the complicated joint distribution from randomly sampled conditional distributions because such complicated joint distributions can be very difficult to integrate analytically. For a more complete introduction of LCA by using Gibbs sampling, please see Yang and Muthén (1997a, 1997b). The method employed in this paper to obtain the parameter estimates is to average the random samples acquired from Gibbs sampling procedures.

We will describe the major Gibbs sampling setups for estimating our finite mixture models in the following. Using the same notations as for the E-M algorithm, the conditional function $f(u_{ik}|c_i)$ has a Bernoulli distribution; however, instead of being latent class indicators as in the previous method, c_i are now distributed as in a Bernoulli distribution. Both Bernoulli distributions have conjugate Dirichlet prior distributions. Following are the equations for the Bayesian LCA model:

$$u_{ik}|c_i\sim Bernoulli(\pi_{kc}),$$

$$\text{where }\pi_{kc}\sim Dirichlet(\alpha_1,\alpha_2),$$

and

$c_i \sim Bernoulli(\lambda)$, where $\lambda \sim Dirichlet(\alpha_1, \alpha_2)$.

We select non-informative priors for the parameters in the LCA model by setting α_1 and α_2 equal to 1. Non-informative priors are chosen to ensure that the E-M algorithm and Gibbs sampling are used on the same platforms; therefore, the difference recorded in this study is only due to estimation algorithms themselves, not to the priors. The posteriors were obtained and parameter estimations were performed using the BUGS (Spiegelhalter *et al*., 1995a, 1995b) computer software.

IV. New Bayesian Models

In this section, we will describe two finite mixture multivariate generalized linear models using the theory of Bayesian statistics. The first model is an extended latent class analysis model for repeated measurements with two sets of independent variables. The independent variables are used as covariates for mixing and conditional probabilities. This model is described based on Bayesian statistics in the following. Assume that y_{ik} is the kth binary response for subject i, and that, C_i are the latent class indicators. π_{ikc} are the conditional probabilities for y_{ik} given C_i; that is,

$$y_{ik}|C_i \sim Bernoulli(\pi_{ikc}),$$

where

$$\text{logit}(\pi_{ikc}) = \beta_{0kc} + \beta_{1kc} w_{1i} + \beta_{2kc} w_{2i} + \ldots + \beta_{pkc} w_{pi}.$$

Moreover, in an LCA model, C_i have a Bernoulli distribution with parameter λ:

$$C_i \sim Bernoulli(\lambda).$$

In other words, λ is the mixing rate of the two distributions. In an ELCA model, we can further assume that each subject's mixing probability is to be predicted by some covariates. This can also be described in the form of an equation as follows:

$$\text{logit}(\lambda_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_m x_{mi}.$$

Further, we select conjugate prior distributions for parameters; that is,

$$\underset{\sim}{\beta} \sim Normal(\mu_\beta, \sigma_\beta^2) \text{ and } \underset{\sim}{\alpha} \sim Normal(\mu_\alpha, \sigma_\alpha^2),$$

where μ_α, σ_α, μ_β, and σ_β are the constant parameters for their corresponding prior distributions.

Similarly, we can extend the previous ELCA model to analyze multivariate Poisson outcomes. The following equation describes the second model:

$$y_{ik}|c_i \sim Poisson(\kappa_{ikc}),$$

where

$$\log(\kappa_{ikc}) = \beta_{0kc} + \beta_{1kc} w_{1i} + \beta_{2kc} w_{2i} + \ldots + \beta_{pkc} w_{pi}.$$

Moreover,

$$C_i \sim Bernoulli(\lambda_i),$$

where

$$\text{logit}(\lambda_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_m x_{mi}.$$

Again, we select conjugate prior distributions for parameters; that is,

$$\underset{\sim}{\beta} \sim Normal(\mu_\beta, \sigma_\beta^2) \text{ and } \underset{\sim}{\alpha} \sim Normal(\mu_\alpha, \sigma_\alpha^2),$$

where μ_α, σ_α, μ_β, and σ_β are the constant parameters for their corresponding prior distributions. The two Bayesian finite mixture multivariate generalized linear models will be discussed and applied in Section V.

V. Monte Carlo and Real Data Illustrations

To compare the two estimation methods and show the plausibility of the proposed Bayesian models, brief simulation studies and real data analyses were conducted.

1. Comparison of the Methods

Monte Carlo comparisons of the E-M algorithms and Gibbs sampling in estimating the finite mixture models are conducted using the typical LCA model. The design of this study was as follows. Two sample sizes, N=300 and N=800, were used. In the LCA model, we used a four-item-test with two latent classes model. The "true" values for generating simulated data sets are listed in Table 1.

Because there is still a lack of a robust diagnostic method for detecting convergence, we had to consider several methods for a single replication in order to determine whether a Gibbs sampling estimation would converge. Nevertheless, we found, from practical experience, that a Gibbs sampling chain with discarding of 5,000 burn-in samples and recording of every 5th of 25,000 samples could usually satisfy most of the

Table 1. LCA Model with N=300 and N=800, Averages of 20 Replications

	True	N=300				N=800			
		E-M	Gibbs Sampling	E-M	Gibbs Sampling	E-M	Gibbs Sampling	E-M	Gibbs Sampling
		Estimates		S. D.		Estimates		S. D.	
P_c=1	0.6	0.6042	0.6008	0.0452	0.0452	0.6029	0.6012	0.0282	0.0279
π_{10}	0.1419	0.1545	0.1637	0.0461	0.0455	0.1396	0.1436	0.0280	0.0279
π_{20}	0.1419	0.1311	0.1405	0.0448	0.0446	0.1386	0.1427	0.0280	0.0278
π_{30}	0.7311	0.7409	0.7328	0.0548	0.0544	0.7268	0.7237	0.0342	0.0339
π_{40}	0.7311	0.7483	0.7420	0.0553	0.0547	0.7343	0.7311	0.0346	0.0345
π_{11}	0.7311	0.7133	0.7093	0.0415	0.0416	0.7317	0.7310	0.0258	0.0257
π_{21}	0.7311	0.7259	0.7234	0.0417	0.0414	0.7307	0.7302	0.0258	0.0257
π_{31}	0.1419	0.1444	0.1503	0.0348	0.0348	0.1419	0.1431	0.0211	0.0211
π_{41}	0.1419	0.1361	0.1390	0.0343	0.0343	0.1354	0.1365	0.0207	0.0208

diagnostic methods employed in this paper. We conducted twenty replications for data simulation and estimation using both methods and summarize the results in Table 1.

For each replication and method, we used three sets of starting values: one set contained "true values", another set was one estimated standard deviation below the true values, and the last set was one estimated standard deviation above the true values. We selected the solution that was the closest to the true values as our final solution; therefore, we could avoid local modes that might trap the estimation methods and give less accurate results. The practical strategy of trying different initial values to avoid local modes has been useful for estimating finite mixture models in many studies (Muthén and Shedden, 1999; Spiegelhalter *et al*., 1995b).

2. Monte Carlo Illustration

In this section, we will demonstrate the Bayesian plausibility of the finite mixture multivariate Poisson model by analyzing artificial data sets. The finite mixture multivariate Poisson model contains one continuous covariate (W) for predicting conditional probabilities (κ_{ikc}) and one continuous covariate (X) for predicting mixing probabilities (λ_i). The values of the intercepts (α_0,β_{0kc}) and the slopes (α_1,β_{1kc}) for W and X used to generate Poisson data sets are listed in Table 2.

Similarly, we conducted twenty replications for this model and used the same rules for selecting initial values. Because of our success with Bayesian LCA models as described in Section V.1, we selected a sample size of 300 for data simulation. The same numbers of updating samples for Gibbs sampling as described in Section V.1 were used in estimating the model. We summarize the parameter estimation results in Table 2. The results confirm that the Bayesian model is plausible in practice and can be reliably estimated.

Table 2. Estimation Results for Mixture Poisson, N=300, Averages of 20 Replications

	True	Mean	S.D.	2.5%	97.5%	Median
$P(C$=1)	0.386	0.3427	0.0568	0.2319	0.4544	0.3423
α_0	-0.7	-0.7801	0.3843	-1.6080	-0.1042	-0.7584
α_1	1.8	1.5960	0.4138	0.9184	2.5320	1.5510
β_{010}	-0.5	-0.3769	0.1063	-0.5894	-0.1698	-0.3762
β_{020}	-0.5	-0.5407	0.1129	-0.7640	-0.3202	-0.5416
β_{030}	-0.8	-0.8191	0.1228	-1.0710	-0.5874	-0.8168
β_{040}	-0.8	-0.6448	0.1132	-0.8665	-0.4247	-0.6458
β_{050}	-0.8	-0.8360	0.1224	-1.0860	-0.6036	-0.8336
β_{011}	-0.8	-1.3550	0.4984	-2.4320	-0.5265	-1.2940
β_{021}	-0.8	-1.1500	0.5173	-2.4220	-0.3509	-1.0940
β_{031}	-1	-1.0890	0.4809	-2.1700	-0.2783	-1.0440
β_{041}	-1	-1.8750	0.6786	-3.5110	-0.8282	-1.7730
β_{051}	-1	-0.6943	0.3620	-1.4780	-0.0520	-0.6748
β_{110}	0.5	0.5138	0.0769	0.3613	0.6613	0.5141
β_{120}	0.5	0.6386	0.1149	0.4140	0.8632	0.6406
β_{130}	0.3	0.1619	0.1120	-0.0599	0.3813	0.1618
β_{140}	0.3	0.3184	0.0972	0.1238	0.5032	0.3212
β_{150}	0.3	0.3230	0.1357	0.0567	0.5841	0.3230
β_{111}	0.8	0.8994	0.3097	0.3698	1.5660	0.8737
β_{121}	0.8	0.8683	0.4194	0.0514	1.7630	0.8634
β_{131}	1.2	1.5150	0.2973	0.9781	2.1490	1.5030
β_{141}	1.2	1.6160	0.3350	1.0440	2.3760	1.5810
β_{151}	1.2	1.1110	0.2225	0.6833	1.5520	1.1090

Note: 2.5% and 97.5% are the lower and upper highest density regions (H.D.R.), respectively.

3. Real Data Illustration

The National Science Foundation of the United States approved a project called Longitudinal Study of American Youth (LSAY) in 1986 to measure American youth's mathematics and science achievement. We selected five mathematics test items and some background variables from the LSAY project to demonstrate

Table 3. Descriptive Statistics of the LSAY Data Set

Variable	Mean	Std. Dev.	Minimum	Maximum	Label
CK201101	0.79	0.41	0	1	GIVE 90,000,000 IN SC
CK202501	0.59	0.49	0	1	CONVERT .425 TO PERCE
CK206701	0.7	0.46	0	1	AT 6 DOZ./MONTH, COOK
CK286501	0.54	0.5	0	1	FIND THE QUOTIENT: −1
CK286502	0.53	0.5	0	1	FIND THE QUOTIENT: +1
GENDER	0.47	0.5	0	1	STUDENT GENDER
HSCRE	3.86	1.38	0	6	HOME SCIENCE RESOURCE
MOTHED	2.73	1.11	1	5	COMPOSITE MOTHER'S ED
SMHAT	11.79	2.48	2.25	16	STUDENT MATH ATTITUDE
SMHAX	2.27	1.69	0	8	STUDENT MATH ANXIETY

Note: Number of valid observations=685.

the proposed ELCA model. There were 685 Grade Seven students in our analysis, and the descriptive statistics are listed in Table 3.

We used subjects' gender, mother's education and home science resources as covariates for the latent class indicators (mixing probabilities). The latent class status can be interpreted as students' true latent mastery ability in mathematics. Furthermore, students' math attitude and math anxiety are predictors for conditional probabilities. We then have the following expression:

$$y_{ik}|C_i \sim Bernoulli(\pi_{ikc}),$$

where i=1, …, 685, k=1, …, 5, c=0, 1,

$$\text{logit}(\pi_{ikc})=\beta_{0kc}+\beta_{1c}SMHAT_{1i}+\beta_{2c}SMHAX_{2i}$$

$$\text{and } \underset{\sim}{\beta} \sim Normal(\mu_\beta,\sigma_\beta^2).$$

To simplify the interpretations, we constraint slopes for math attitude and anxiety equal across the test items and set intercepts estimated freely between different items. The estimated results are listed in Table 4.

Moreover,

$$C_i \sim Bernoulli(\lambda_i),$$

where

$$\text{logit}(\lambda_i)=\alpha_0+\alpha_1 GENDER_{1i}+\alpha_2 MOTHED_{2i}$$
$$+\alpha_2 HSCRE_{3i},$$

and

$$\underset{\sim}{\alpha} \sim Normal(\mu_\alpha,\sigma_\alpha^2).$$

We used non-informative priors for parameters; i.e., we let $\mu_\alpha=\mu_\beta=0$ and $\sigma_\alpha=\sigma_\beta=100$. Further, a Gibbs sampling chain with discarding of 10,000 burn-in samples and recording of every 10th of 50,000 samples was used for this example. The Gibbs sampling chain satisfied the convergence diagnostic tests.

Table 4. Bayesian ELCA Model for the LSAY Data Set

	Estimates	S. D.	2.5%[b]	97.5%[b]
P_c=1	0.4813	0.0501	0.3831	0.5795
π_{10}	0.7449	0.0253	0.6942	0.7919
π_{20}	0.4497	0.0322	0.3851	0.5109
π_{30}	0.6046	0.0301	0.5429	0.6621
π_{40}	0.1960	0.0488	0.0989	0.2870
π_{50}	0.1799	0.0425	0.0958	0.2629
π_{11}	0.8364	0.0240	0.7869	0.8809
π_{21}	0.7482	0.0309	0.6885	0.8076
π_{31}	0.8029	0.0267	0.7494	0.8543
π_{41}	0.9154	0.0398	0.8388	0.9944
π_{51}	0.8977	0.0464	0.8112	0.9965
α_0	−0.0775[a]	0.2161	−0.5024	0.3416
α_1	−0.2640[a]	0.1936	−0.6477	0.1047
α_2	0.3700	0.0886	0.2016	0.5465
α_3	0.1714	0.0716	0.0319	0.3128
β_{010}	1.0790	0.1342	0.8220	1.3410
β_{020}	−0.2034[a]	0.1314	−0.4685	0.0437
β_{030}	0.4281	0.1267	0.1730	0.6762
β_{040}	−1.8390	6.6970	−2.2160	−0.9128
β_{050}	−1.5510	0.3150	−2.2510	−1.0320
β_{011}	1.6640	0.1786	1.3240	2.0240
β_{021}	1.1110	0.1683	0.8019	1.4560
β_{031}	1.4330	0.1732	1.1090	1.7890
β_{041}	2.6150	0.8970	1.6750	5.2010
β_{051}	2.4310	1.0360	1.4730	5.7250
β_{10}	0.0439[a]	0.0344	−0.0243	0.1119
β_{11}	0.1206	0.0516	0.0122	0.2172
β_{20}	0.0575[a]	0.0479	−0.0382	0.1495
β_{21}	0.1122[a]	0.0812	−0.0458	0.2753

Notes: α_0: intercept; α_1: slope for Gender; α_2: slope for Mother's Education; α_3: slope for Home Science Resources; β_{0xx}: Intercepts; β_{1x}: slopes for Math Attitude; β_{2x}: slopes for Math Anxiety

[a] Non-significant.

[b] 2.5% and 97.5% are the lower and upper highest density regions (H.D.R.), respectively.

Some interesting results are revealed by our study of this model and are summarized as follows. The inferences from mother's education and home science resources to students' latent mastery status are found to be statistically significant while the inferences from gender have no significant effects. Details of the parameter estimates can be found in Table 4.

VI. Conclusion

Our primary results show that Gibbs sampling gives more accurate estimates and smaller standard deviations than the E-M algorithm does although the differences are small in our examples. Nevertheless, a comparison (Table 1) was made given that Gibbs sampling has non-informative priors. Practical experience shows that Gibbs sampling may perform much better when informative priors are available.

For the LCA model considered in this paper, we found that Gibbs sampling took much more time than the E-M algorithm did. For example, Gibbs sampling needed about 50 minutes to complete a 5k/25k run in order to pass most of the convergence diagnostic tests. This length of time (50 minutes) did not include convergence diagnosis time. On the other hand, the E-M algorithm only took a few seconds to finish a run. The above computational related comments were made using computers with Pentium-100 CPU's, 32 megabytes RAM's, and Windows95 operating systems.

The E-M algorithm uses a traditional convergence detecting method that can obtain convergence when the E-M iterations' absolute change is less than a certain small value. The method can be easily implemented and requires no extra effort. Gibbs sampling requires special care in convergence diagnosis since a robust method for detecting convergence is still lacking. Several diagnostic methods for Gibbs sampling convergence have been proposed and discussed in many papers. For a good summary, see the paper by Kass *et al.* (1998), who discussed many practical aspects of Gibbs sampling estimation and convergence. In our study, we monitored the convergence statistics and obtained plots using the computer software CODA (Best *et al.*, 1997). However, this became extremely difficult when the number of replications increased and this is a limitation of this paper. A more complete Monte Carlo study will be possible when a more robust diagnostic method for Gibbs sampling convergence is available.

In addition to the two different sample sizes (300 and 800), we also tried to use a sample size of 100 for data simulations and used the two methods to recover designed parameters. Interestingly, we found that both methods had less stable parameter recovery behavior for a sample size of 100 than for sample sizes of 300 or 800. An anonymous referee pointed out that in this case, prior information becomes more important, and the sensitivity analysis in the prior specification has to be examined. More studies are needed to explore the issues of small sample sizes and effects due to prior information for both methods.

Further research is needed in several areas. For example, it may be possible to generate useful models by adding random effects to ELCA models. More investigations on convergence problems where estimation involves small sample sizes are especially needed. Statistical selection between different models may also be an important future research topic.

References

Bartholomew, D. J. (1987) *Latent Variable Models and Factor Analysis*. Oxford University Press, New York, NY, U.S.A.

Best, N. G., M. K. Cowles, and S. K. Vines (1997) *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*, Ver. 0.4. MRC Biostatistics Unit, IPH, Cambridge, U.K.

Breslow, N. E. and D. G. Clayton (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.

Brooks, S. P., B. J. T. Morgan, M. S. Ridout, and S. E. Pack (1997) Finite mixture models for proportions. *Biometrics*, **53**, 1097-1115.

Campbell, J. G., C. Fraley, F. Murtagh, and A. E. Raftery (1997) Linear flaw detection in woven textiles using models-based clustering. *Pattern Recognition Letters*, **18**, 1539-1548.

Celeux, G. and G. Govaert (1995) Gaussian parsimonious clustering models. *Pattern Recognition*, **28**, 781-793.

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

Follmann, D. A. and D. Lambert (1989) Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, **84**, 295-300.

Goodman, A. L. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215-231.

Jedidi, K., H. S. Jagpal, and W. DeSarbo (1997) Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, **16**, 39-59.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal (1998) Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, **52**, 93-100.

Melton, B., K. Y. Liang, and A. E. Pulver (1994) Extended latent class approach to the study of familial/sporadic forms of a disease: its application to the study of the heterogeneity of schizophrenia. *Genetic Epidemiology*, **11**, 311-327.

Mukerjee, S., E. D. Feigelson, G. J. Babu, F. Murtagh, C. Fraley, and A. Rafery (1998) *Three Types of Gamma Ray Bursts*. Technical Report, Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA, U.S.A.

Murtagh, F. and A. E. Raftery (1984) Fitting straight lines to point patterns. *Pattern Recognition*, **17**, 479-483.

Muthén, B. and K. Shedden (1999) Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, (in press).

Muthén, B., H. Brown, S. Khoo, C. C. Yang, and B. Jo (1996) General growth mixture modeling of latent trajectory classes: perspectives and prospects. The Prevention Science and Methodology Groups Meeting, Tempe, AZ, U.S.A.

Pickering, R. M. and J. F. Forbes (1984) A classification of Scottish infants using latent class analysis. *Statistics in Medicine*, **3**, 249-259.

Qu, Y., M. Tan, and H. M. Kunter (1996) Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, **72**, 797-810.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks (1995a) *Bayesian Inference Using Gibbs Sampling* (*BUGS*), Ver. 0.5. MRC Biostatistics Unit, IPH, Cambridge, U.K.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks (1995b) *BUGS Examples*, Ver 0.5, Vol. 1-2. Technical Report, MRC Biostatistics Unit, IPH, Cambridge, U.K.

Tanner, M. A. (1996) *Tools for Statistical Inference*, 3rd Ed. Springer-

Verlag, New York, NY, U.S.A.
Wang, P. and M. L. Puterman (1998) Mixed logistic regression models. *Journal of Agricultural, Biological and Environmental Statistics*. **3**, 1-26.
Wang, P., M. L. Puterman, I. Cockburn, and N. Le (1996) Mixed Poisson regression models with covariate dependent rates. *Biometrics*, **52**, 381-400.
Yang, C. C. (1998) *Finite Mixture Model Selection with Psychometrics*. Ph.D. Dissertation. University of California, Los Angeles, Los Angeles, CA, U.S.A.
Yang, C. C. and B. Muthén (1997a) Latent class analyses using Gibbs sampling and E-M algorithm. The 1997 IMS Asian and Pacific Regional Meeting, Academia Sinica, Taipei, Taiwan, R.O.C.
Yang, C. C. and B. Muthén (1997b) Finite mixture of generalized linear models using Gibbs sampling and E-M algorithm. The 1997 American Statistical Association (ASA) Joint Statistics Meeting (JSM), Anaheim, CA, U.S.A.
Yang, I. and M. P. Becker (1997) Latent variable modeling of diagnostics accuracy. *Biometrics*, **53**, 948-958.
Yung, Y. F. (1997) Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, **62**, 297-330.

以吉氏取樣與E-M算法估算之有限混合型廣義線性模式

楊志堅*　Bengt O. Muthén　楊志強*****

*國立臺中師範學院教育測驗統計研究所
**Department of Education
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA, U.S.A.
***僑光商專

摘　要

有限混合型的多變量廣義線性模式在許多的研究領域中如，影像辨認、天文資料分類、生物醫學診斷與分類皆佔有極重要的地位，近幾年來統計與計算科學的突飛猛進更鼓舞了研究者利用貝氏理論做各式該類模式的探究。本文主要針對E-M算法與吉氏取樣的估算特性做模擬研究以供研究參考，除此之外，本文並提出兩個有限混合型的貝氏多變量廣義線性模式，該模式可提供更有彈性的模式化功能，更能接近真實的研究環境，最後本文並佐以模擬試驗及實例來說明如何以吉氏取樣來做有限混合型的貝氏多變量廣義線性模式分析。