International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:3, No:6, 2009

Applying Gibbs Sampler for Multivariate Hierarchical Linear Model

Satoshi Usami

***Abstract*—** Among various HLM techniques, the Multivariate Hierarchical Linear Model (MHLM) is desirable to use, particularly when multivariate criterion variables are collected and the covariance structure has information valuable for data analysis. In order to reflect prior information or to obtain stable results when the sample size and the number of groups are not sufficiently large, the Bayes method has often been employed in hierarchical data analysis. In these cases, although the Markov Chain Monte Carlo (MCMC) method is a rather powerful tool for parameter estimation, Procedures regarding MCMC have not been formulated for MHLM. For this reason, this research presents concrete procedures for parameter estimation through the use of the Gibbs samplers. Lastly, several future topics for the use of MCMC approach for HLM is discussed.

***Keywords*—** Gibbs sampler. Hierarchical Linear Model. Markov Chain Monte Carlo. Multivariate Hierarchical Linear Model.

I. Introduction

The Hierarchical Linear Model (HLM) is a regression model for hierarchical data sets and has been attracting interest in various research fields including education, psychology, sociology and marketing (e.g., [11],[13],[17],[18],[19]). As an example, hierarchical data often appear in the research field of education since students are generally nested exclusively within classes and classes are also nested exclusively within schools [7],[17].

Among various HLM techniques, it is desirable to use the Multivariate Hierarchical Linear Model (MHLM), particularly when multivariate data of criterion variables are collected [3],[4],[18],[19]. MHLM can take advantage of information concerning the covariance structure of criterion variables, allowing researchers to obtain a better and more complete description of what is affected by changes in explanatory variables [19]. Other advantages of MHLM are that while carrying out a series of univariate statistical tests through HLM inflates the type I error, it is better controlled through MHLM, and MHLM often exhibits greater statistical power.

The Markov Chain Monte Carlo (MCMC) method has been attracting an interest among researchers focusing on model construction (e.g., [2],[6],[13]). In both HLM and MHLM, Bayesian statistics has often been applied in analysis in order to reflect prior knowledge or to avoid obtaining incorrect estimates in cases of insufficient sample size and number of groups (e.g., [16]). In the Bayesian method, posterior distribution becomes complex, sometimes by non-conjugate prior distribution, and this can make the derivation of the marginal posterior distribution impossible. However, sampling methods using MCMC make the evaluation of point estimates and the associated standard errors of posterior distribution much easier [9].

Several approaches for using MCMC methods for HLM have been developed [12],[16]. However, some MCMC procedures have not been formulated for MHLM. Therefore, in this research, the concrete procedures for applying the Gibbs sampler technique to multivariate hierarchical data is presented, by using modeling techniques utilized mainly by Thum [19].

II. Model and Estimation

First, all symbols used in this paper are defined as follows.

・Number of groups : J, ・Sample size in group j : N_j

・Total sample size : $N=\sum N_j$ ・Number of criterion variables : K

・Number of explanatory variables at level 1 and level 2 : S_1 and S_2

A. Model

In group j, the model formulation for individual levels (level1) is as follows.

$$Y_j = X_j \beta_j + r_j \tag{1}$$

Here, criterion variables Y_j and regression coefficients β_j is expressed as $Y_j = (Y_{j1}^t, Y_{j2}^t, \cdots Y_{jK}^t)^t$, $Y_{jk} = (Y_{1jk}, \cdots Y_{ijk}, \cdots Y_{N_j jk})$ and $\beta_j = (\beta_{j1}^t, \beta_{j2}^t, \cdots \beta_{jK}^t)^t$, $\beta_{jk} = (\beta_{j1k}, \cdots \beta_{jS1k}, \cdots \beta_{jS1k})$, respectively. Assume that the responses matrix of member N_j to S_1 explanatory variables is expressed as an $N_j \times S_1$ matrix, $\tilde{X}_j$. Then, matrix notation for independent variables X_j is expressed by using Kronecker multiplication. That is, $X_j = I_k \otimes \tilde{X}_j$.

r_j is an $(N_j \times K)\times 1$ residual vector whose element r_{ijk} is supposed to be distributed as $r_{ijk} \sim N(0, \Sigma_j)$, where Σ_j is an $(N_j \times K)\times(N_j \times K)$ residual variance-covariance matrix for group j. Generally, as data from different groups are assumed to be independent, Σ_j can be expressed as $\Sigma_j = \tilde{\Sigma} \otimes I_{Nj}$. $\tilde{\Sigma}$ is a residual matrix between the criterion variables with the assumption of equivalence in all groups.

Then, model formulation for the group level (level2) is as

follows.

$$\beta_j = Z_j\gamma + u_j \tag{2}$$

Here, Z_j is independent variable for groups, and is expressed as $Z_j = I_{S1\times K} \otimes \tilde{z}_j$, and, $\tilde{z}_j = (z_{j1,} \cdots z_{js2,} \cdots z_{jS2})$.

Assume that the S_2 regression coefficients of s_1 explanatory variables for the criterion variable of k are expressed as $\gamma_{s1k} = (\gamma_{1s1k,} \cdots \gamma_{s2s1k}, \cdots \gamma_{S2s1k})$, γ is a $(S_2 \times S_1 \times K) \times 1$ regression coefficients vector , $\gamma = (\gamma_{1,}^t \gamma_2^t, \cdots \gamma_K^t)^t$, where each element is expressed as $\gamma_k = (\gamma_{1k,}^t \cdots \gamma_{s1k}^t, \cdots \gamma_{S1k}^t)^t$.

u_j is a size $(S_1 \times K) \times 1$ residual vector, and $u_j \sim N(0,T)$. T is an $(S_1 \times K) \times (S_1 \times K)$ residual variance-covariance matrix, and some constraints can be imposed as individual levels.

B Parameter estimation

Prior distribution

In this paper, prior distributions for β, γ, $\tilde{\Sigma}$, T are set as follows.

$$\beta \propto h_1, \quad \gamma \propto h_2,$$
$$p(\tilde{\Sigma}) \sim W^{-1}(v_1, \Sigma_1), \; p(T) \sim W^{-1}(v_0, \Sigma_0) \tag{3}$$

Where, h_1 and h_2 are constants, and $\propto$ indicates a proportional relation. Additionally, v_1, v_0, Σ_1, Σ_0 are hyper-parameters, which are to be set before the analysis.

v_1, v_0 are related to the sample size and the number of groups corresponding to the prior information, and Σ_1, Σ_0 represent information for the square sum of residuals within and between groups. In addition, W^{-1} represents the inverse Wishart distribution.

Likelihood and conditional posterior distribution

As data from different groups are assumed to be independent, the likelihood for the whole data becomes,

$$L(\beta, \gamma, \tilde{\Sigma}, T / Y, X, Z) = \prod \frac{1}{(2\pi)^{p/2}\left|V_j\right|^{1/2}} \exp[(Y_j - U_j)^t V_j^{-1}(Y_j - U_j)] \tag{4}$$

where, $\beta = (\beta_1, \beta_2, \cdots \beta_J)$, $Y = (Y_1, Y_2, \cdots Y_J)$, $X = (X_1, X_2, \cdots X_J)$, $Z = (Z_1, Z_2, \cdots Z_J)$, $U_j = X_j Z_j \gamma$, $V_j = X_j^t T X_j + \Sigma_j$.

From the Bayes theorem, the joint posterior distribution becomes

$$P(\beta, \gamma, \tilde{\Sigma}, T / Y, X, Z) \propto L(\beta, \gamma, \tilde{\Sigma}, T / Y, X, Z) P(\beta) P(\gamma) p(\tilde{\Sigma}) p(T) \tag{5}$$

In addition, the full conditional posterior distribution, for example β, becomes as follows.

$$P(\beta / \gamma, \tilde{\Sigma}, T, Y, X, Z) \propto L(\beta, \gamma, \tilde{\Sigma}, T / Y, X, Z) P(\beta) \tag{6}$$

That is, the full conditional distribution is proportional to the product of likelihood and prior distribution of each parameter.

C Gibbs sampler algorithm

The Gibbs sampler method is feasible when sampling from conditional posterior distribution is possible, and when sampling from marginal posterior distribution is difficult. In the Gibbs sampler, iterative calculation is repeated by using full conditional distribution.

Referring to [8],[12],[14],[15],[16], who studied sampling procedures in the framework of linear model, the detailed algorithm through the Gibbs sampler for MHLM is derived as follows. These results can be easily obtained through the evaluation of full conditional distribution for each parameter and basic matrix operation [5],[10]. For the brief notation, detail derivation is omitted here.

$$\tilde{\Sigma}^{-1} = W(N + v_1, (\Sigma_1 + \sum_j E_j E_j^t)^{-1})) \tag{7}$$

$$\beta_j = N((I_{S1\times K} - \Lambda_j)G_j + \Lambda_j H_j, (T^{-1} + X_j^t \tilde{\Sigma}^{-1} X_j)^{-1}) \tag{8}$$

$$T^{-1} = W(J + v_0, (\Sigma_0 + \sum_j F_j F_j^t)^{-1}) \tag{9}$$

$$\gamma = N((\sum_j Z_j^t T^{-1} Z_j)^{-1} \sum_j Z_j^t T^{-1} \beta_j, (\sum_j Z_j^t T^{-1} Z_j)^{-1}) \tag{10}$$

where $E_j = (E_{j1}, E_{j2}, \cdots E_{jK})^t$, and

$$E_{jk} = Y_{jk} - \tilde{X}_j \beta_{jk} \tag{11}$$

$$F_j = \beta_j - Z_j \gamma \tag{12}$$

$$G_j = Z_j \gamma \tag{13}$$

$$H_j = (X_j^t \tilde{\Sigma}^{-1} X_j)^{-1} X_j^t \tilde{\Sigma}^{-1} Y_j \tag{14}$$

$$\Lambda_j = (X_j^t \tilde{\Sigma}^{-1} X_j + T)^{-1} X_j^t \tilde{\Sigma}^{-1} X_j \tag{15}$$

That is, E_j and F_j has information regarding the residual of individual levels and group levels, respectively. Λ_j is generally referred to as multivariate reliability matrix [13]. Therefore, β_j is an empirical Bayes estimator that uses weights for the ratio concerning the residual variance-covariance matrices at the individual levels and the group levels.

In the Gibbs sampler, as with the Metropolis-Hastings algorithm, appropriate initial values should be elected for stable

convergence of estimates.

D Parameter estimation from sampling

In MCMC application, after T times sampling for each parameter, the obtained samples can be regarded from the point of view of marginal distributions. However, the first $V(\leq T)$'s samples are excluded from the calculation for parameter estimation in order to obtain more stable and plausible estimates. This V period is commonly referred to as burn-in.

With respect to the parameter estimates, for instance, regarding γ, the point estimates $\bar{\gamma}$ and the estimates of the standard error $\sqrt{V(\bar{\gamma})}$ can be calculated by using the average and the standard deviation of samples through $T-V$ times by removing samples corresponding to the burn-in period. That is,

$$\bar{\gamma} = \frac{1}{T-V}\sum_{t=V+1}^{T}\gamma^t \tag{16}$$

$$\sqrt{V(\bar{\gamma})} = \sqrt{\frac{1}{T-V}\sum_{t=V+1}^{T}(\gamma^t - \bar{\gamma})^2} \tag{17}$$

γ^t is the value of γ at t-th sampling.

In applying Gibbs Sampler, there are cases where it is not feasible, especially when derivation of conditional posterior distribution difficult. However, Gibbs Sampling is power tool even when some expansions are needed for future researches.

III. DISCUSSION

Among the various HLM techniques, it is desirable to use MHLM, particularly when multivariate criterion variables are collected and the covariance structure contains valuable information for data analysis.

The Bayes method has often been used in hierarchical data analysis in order to reflect prior information or to obtain stable results when the sample size and the number of groups are not large enough. In these cases, although parameter estimation through MCMC is a powerful tool, the concrete procedures have not yet been derived for MHLM, especially in the framework of Thum [19]. Therefore, in this research, the concrete procedure for parameter estimation through Gibbs sampler was shown.

These methods use iterative algorithms, and as the size of the explanatory variable matrix and error variance-covariance matrix generally become large in MHLM, MCMC may be a bit computationally inconvenient for MHLM. However, MCMC promotes a powerful framework for hierarchical data analysis, and the severity of this problem will gradually decrease as the processing speed of computers improves.

An intriguing topic for future research, for example, is to construct a framework for performing factor analysis and multivariate hierarchical data analysis simultaneously.

Due to the fact that when the number of criterion variables is large in hierarchical data, integrating the results to fewer factor scores is useful for summarizing the information. Regarding this topics, some combinations of Structural Equation Models (SEM), Hierarchical Linear Models and Bayesian analysis undoubtedly helps to offer powerful tools [1].

What is more, in this approach, to construct an algorithm which determines a correct number of factors effectively is another attractive direction of expansion, as well as the development of computationally convenient algorithms. As for this problem, Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm must be useful and should be considered in future researches. Extension for applying some missing values and ordinal variables are also intriguing topics for MHLM approach.

Additionally, on a parallel with these theoretical researches, developing useful software for MHLM through the MCMC algorithm is also desired.

[1] Fahrmeir, L., & Raach, A, (2007). A Bayesian Semiparametric Latent Variable Model for Mixed response. *Psychometrica ,72*(3), 327–346
[2] Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.
[3] Goldstein, H. (1987). *Multilevel models in education and social research*. London: Oxford University Press.
[4] Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). New York: Oxford University Press.
[5] Harville, D. A. (2000). *Matrix algebra from a statistician's perspective*. Springer.
[6] Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
[7] Miyazaki, Y. (2007). Application of hierarchical linear models to educational research and viewpoints of utilizing the results to educational policies, *Japanese Journal for Research on Testing 3* (1), 123-146
[8] Okumura, T. (2007). Sample size determination for hierarchical linear models considering uncertainty in parameter estimates. *Behaviormetrika, 34*(2), 79-94.
[9] Omori, Y. (2001). Recent developments in Markov Chain Monte Carlo. *Japan Statistical Society*, 31, 305-344.
[10] Magnus, J. R. & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Wiley.
[11] Plewis, I. (2005). Modeling behaviour with multivariate multilevel growth curves. *Methodology*, 1(2), 71-80.
[12] Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika* 69, 167-190.
[13] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. (2nd ed.). Sage.
[14] Seltzer, M. H. (1991). The use of data augmentation in fitting hierarchical linear models to educational data. Unpublished doctoral dissertation, University of Chicago.
[15] Seltzer, M. H. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational Statistics, 18*, 207-235.
[16] Seltzer, M. H., Wong, W. H., & Bryk, A. S. (1996). Bayesian analysis in applications of hierarchical models: issues and methods. *Journal of Educational and Behavioral Statistics, 21*, 131-167.
[17] Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
[18] Tate, R. L., & Pituch, K. A. (2007). Multivariate hierarchical linear modeling in randomized field experiments. *The Journal of Experimental Education, 73*(4), 317-337
[19] Thum, Y. M. (1997). Hierarchical linear models for multivariate outcomes. *Journal of Educational and Behavioral Statistics, 22*, 77-108.