Communications for Statistical Applications and Methods
2013, Vol. 20, No. 3, 235–240

DOI: http://dx.doi.org/10.5351/CSAM.2013.20.3.235

Bayesian Modeling of Random Effects Covariance Matrix for Generalized Linear Mixed Models

Keunbaik Lee[1,a]

[a]Department of Statistics, Sungkyunkwan University

Abstract

Generalized linear mixed models(GLMMs) are frequently used for the analysis of longitudinal categorical data when the subject-specific effects is of interest. In GLMMs, the structure of the random effects covariance matrix is important for the estimation of fixed effects and to explain subject and time variations. The estimation of the matrix is not simple because of the high dimension and the positive definiteness; subsequently, we practically use the simple structure of the covariance matrix such as AR(1). However, this strong assumption can result in biased estimates of the fixed effects. In this paper, we introduce Bayesian modeling approaches for the random effects covariance matrix using a modified Cholesky decomposition. The modified Cholesky decomposition approach has been used to explain a heterogenous random effects covariance matrix and the subsequent estimated covariance matrix will be positive definite. We analyze metabolic syndrome data from a Korean Genomic Epidemiology Study using these methods.

Keywords: Modified Cholesky decomposition, heterogeneity, Positive definiteness.

1. Introduction

Generalized linear mixed models(GLMMs) are frequently used for the analysis of longitudinal categorical data when subject-specific effects are of interest (Breslow and Clayton, 1993). In the GLMMs, the structure of a random effects covariance matrix is important for the estimation of fixed effects and to explain subject and time variations. The estimation of the matrix is not simple because of the high dimension and the positive definiteness; subsequently, we practically use the simple structure of the covariance matrix such as AR(1) because of the positive definite constraint and the difficulty of estimation of the matrix. However, this strong assumption can result in biased estimates of the fixed effects (Heagerty and Kurland, 2001).

To release the strong assumption, Pourahmadi (1999, 2000) proposed the modified Cholesky decomposition which decomposes the random effects covariance matrix into two sets of parameters: generalized autoregressive parameters(GARPs) and the innovation variances(IVs). The GARPs are dependence parameters that are coefficients of previous random effects and the IVs are variance parameters for the current random effect. The positive definiteness restriction of the covariance matrix is that the IVs need to be positive (Pourahmadi, 1999, 2000). In addition, the modified Cholesky decomposition reduces the number of parameters in the covariance matrix and is used for the estimation of the covariance matrix to analyze longitudinal Gaussian data (Daniels and Pourahmadi, 2002; Daniels and Zhao, 2003; Pan and Mackenzie, 2003, 2006). Lee *et al.* (2012) proposed a GLMM with the

This project was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (2012-004002).

[1] Associate Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.
E-mail: keunbaik@skku.edu

heterogenous random effects covariance matrix depending on covariates via the modified Cholesky decomposition. In this paper, we propose the Bayesian version of a GLMM with a heterogenous random effects covariance matrix that is expressed through a modified Cholesky decomposition.

The paper is organized as follows. In Section 2, we describe the modified Cholesky decomposition approach and subsequently use it to present a Bayesian modeling for GLMM. In Section 3, we analyze data from the longitudinal study on metabolic syndrome. Finally, conclusions and extensions are provided in Section 4.

2. Bayesian Generalized Linear Mixed Models for Longitudinal Binary Data

We first explain the generalized linear mixed model with a heterogenous random effects covariance matrix.

2.1. Modified Cholesky decomposition for GLMM

Let Y_{it} be the binary response for subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$) and let x_{it} be the corresponding vector of covariates. We assume that each Y_{it} is conditionally independent given random effects b_{it}, the responses for different subjects are independent, and the regression model is given by

$$\text{logit}\, p_{it}(b_{it}) = x_{it}\beta + b_{it}, \tag{2.1}$$

where $p_{it}(b_{it}) = P(Y_{it} = 1; b_{it})$ amd β is the $p \times 1$ vector of regression coefficient. We assume that

$$b_i = (b_{i1}, \ldots, b_{in_i})^T \sim N(0, \Sigma_i),$$

where Σ_i is the random effects covariance matrix and b_i is a vector of random effects values for subject i.

To solve the positive-definiteness constraint and the exponentially increasing number of parameters of Σ_i, we use the modified Cholesky decomposition. We have

$$b_{i1} = e_{i1}, \tag{2.2}$$

$$b_{it} = \sum_{j=1}^{t-1} \phi_{i,tj} b_{ij} + e_{it}, \quad \text{for } t = 2, \ldots, n_i, \tag{2.3}$$

where $e_i = (e_{i1}, \ldots, e_{in_i}) \sim N(0, D_i)$ with $D_i = \text{diag}(\sigma^2_{i1}, \ldots, \sigma^2_{in_i})$. From (2.2) and (2.3), we have the following matrix form as

$$T_i b_i = e_i, \tag{2.4}$$

where T_i is a lower triangular matrix having ones on its diagonal and $-\phi_{i,tj}$ at its $(t, j)^{th}$ position for $j < t$. Then we have

$$T_i \Sigma_i T_i^T = D_i.$$

Here, ϕ_i and σ_i^2 are called by the generalized autoregressive parameters(GARPs) and the innovation variances(IVs), respectively. The GARP/IV parametrization provides parameters that can easily be modeled without the concern of the estimator being positive definite, that have a sensible interpretation, and that allow for simple computation (Daniels and Zhao, 2003; Lee *et al.*, 2012).

The parameters, GARP and IV can be modeled using time and/or subject-specific covariate vectors $w_{i,tj}$ and $h_{i,t}$ by setting

$$\phi_{i,tj} = w_{i,tj}^T \gamma, \qquad \log\left(\sigma_{it}^2\right) = h_{i,t}^T \lambda, \tag{2.5}$$

where γ is a $a \times 1$ vector of unknown dependence parameters, λ is a $b \times 1$ vector of unknown variance parameters, design vectors $w_{i,tj}$ and $h_{i,t}$ are covariates to model the GARP/IV parameters as functions of subject-specific covariates (Pourahmadi, 2000; Pourahmadi and Daniels, 2002; Daniels and Zhao, 2003; Lee *et al.*, 2012). Therefore, the random effects covariance matrix can be heterogeneous in the subject-specific covariates. Because the positive σ_{it}^2 guarantees the positive definiteness of Σ_i, the loglinear model is used in (2.5).

2.2. Bayesian modeling

Now we specify the prior distributions of the parameters. The diffused prior distributions for β, γ, and λ are given by

$$\beta \sim N\left(0, \sigma_\beta^2 I\right), \tag{2.6}$$

$$\gamma \sim N\left(0, \sigma_\gamma^2 I\right), \tag{2.7}$$

$$\lambda \sim N\left(0, \sigma_\lambda^2 I\right), \tag{2.8}$$

where $\sigma_\beta^2 = 100$, $\sigma_\gamma^2 = 100$, and $\sigma_\lambda^2 = 100$.

From the sampling distribution (2.1) and prior distributions (2.6)–(2.8), we have the joint distribution given by

$$f(y, b, \beta, \gamma, \lambda) \propto \phi\left(\beta|0, \sigma_\beta^2\right) \phi\left(\gamma|0, \sigma_\gamma^2\right) \phi\left(\lambda|0, \sigma_\lambda^2\right) \left[\prod_{i=1}^{N} \left\{ \prod_{t=1}^{n_i} \left(p_{it}^c\left(b_{it}, \beta\right)\right)^{y_{it}} \left(1 - p_{it}^c\left(b_{it}, \beta\right)\right)^{1-y_{it}} \right\} \phi(b_i|\gamma, \lambda)\right],$$

where $p_{it}^c(b_{it}, \beta) = P(Y_{it} = 1|x_{it}, b_{it}, \beta)$ and $\phi(\cdot)$ is the multivariate normal distribution function. For the estimation of our model, Gibbs sampling is implemented using WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml). The WinBUGS code is available upon request.

2.3. Deviance information criterion

In Bayesian modeling, there are several model selection criteria such as Bayes factor, posterior predictive loss, and deviance information criterion(DIC) (Daniels and Hogan, 2008). Especially, we use DIC for the model selection criterion of this paper. The DIC is commonly used to compare competing models (Spiegelhalter *et al.*, 2002) and it is a model-based criterion composed of a goodness of fit term and a penalty term. Let θ be a vector of all parameters in (2.1). Then the fit is measured by the deviance given by

$$\text{Dev}(\theta) = -2 \log L(\theta|y),$$

where $L(\theta|y)$ is the likelihood of $y = (t_1, \ldots, y_N)^T$. Larger values of the deviance indicate poorer fit. The penalty term measures model complexity and is given by

$$p_D = E\left\{\text{Dev}(\theta)|y\right\} - \text{Dev}\left\{E(\theta|y)\right\}.$$

The value of p_D is called the effective number of parameters. The DIC is defined as

$$\begin{aligned}\text{DIC} &= \text{Dev}\left\{E(\theta|y)\right\} + 2p_D \\ &= -4E\left\{\log L(\theta|y)|y\right\} + 2\log L\left\{E(\theta|y)|y\right\}.\end{aligned}$$

In practice, DIC can be expressed in different ways depending on how $E\{\text{Dev}(\theta)|y\}$ and $\text{Dev}\{E(\theta|y)\}$ are estimated or approximated. Celeux *et al.* (2006) compared various forms of DICs and suggested DIC_3 as one of the most reliable forms of DIC:

$$\text{DIC}_3 = -4E\left\{\log L(\theta|y)|y\right\} + 2\log L\left(\hat{\theta}|y\right),$$

where

$$L\left(\hat{\theta}|y\right) = \prod_{i=1}^{N} E_\theta\left\{f(y_i|\theta)|y\right\}$$

with $f(y_i|\theta)$ is a probability density function for subject i.

For our model, the first and second terms in DIC_3 are approximated by a Markov chain Monte Carlo(MCMC) algorithm as

$$E\left\{\log L(\theta|y)|y\right\} \approx \frac{1}{M}\sum_{l=1}^{M}\sum_{i=1}^{N}\sum_{t=1}^{n_i}\left[y_{it}\left(x_{it}\beta^{(l)} + b_{it}^{(l)}\right) + \log\left(1 - p_{it}^{c}\left(b_{it}^{(l)},\beta^{(l)}\right)\right)\right]$$

and

$$\log L\left(\hat{\theta}|y\right) \approx \sum_{i=1}^{N}\log\left[\frac{1}{M}\sum_{l=1}^{M}\prod_{t=1}^{n_i}\left(p_{it}^{c}\left(b_{it}^{(l)},\beta^{(l)}\right)\right)^{y_{it}}\left(1 - p_{it}^{c}\left(b_{it}^{(l)},\beta^{(l)}\right)\right)^{1-y_{it}}\right],$$

where M is the number of iterations, $p_{it}^{c}(b_{it}^{(l)},\beta^{(l)}) = \exp(x_{it}\beta^{(l)} + b_{it}^{(l)})/(1 + \exp(x_{it}\beta^{(l)} + b_{it}^{(l)}))$, and $(\beta^{(l)}, b_{it}^{(l)})$ is the set of simulated random numbers at the $l$the MCMC iteration.

3. Example

The Korean Genomic Epidemiology Study(KoGES) is a cohort study to monitor the development of metabolic syndrome for 2310 Korean adults aged 39–69 years (Kim *et al.*, 2006). Participants were examined every two years for up to eight years to monitor the development of metabolic syndrome. Metabolic syndrome is defined as three or more of the following five disorders: abdominal obesity (waist circumference > 90cm in men or > 80cm in women), high blood pressure (systolic BP levels > 130mmHg or diastolic BP levels > 85mmHg), high impaired fasting glucose (IFG > 110mg/dl), high triglyceridemia (TG > 150mg/dl), and low high-density lipoprotein cholesterol (HDL-C < 40mg/dl in men or < 50mg/dl in women). It is of primary interest how demographic factors affect metabolic syndrome. The demographic factors were sex, age, alcohol intake, and smoking.

We analyzed the effect of the demographic factors on metabolic syndrome in KoGES. In this paper, we used the final binary outcome for metabolic syndrome ($Y = 1$ for presence of metabolic syndrome; 0 for absence of metabolic syndrome). For predictors, we included sex (1 = male; 0 = female), age (log(age/10)), alcohol intake type (Drink1 = 1 if drinking in the past, 0 otherwise; Drink2 = 1 if drinking currently, 0 otherwise), and smoking types (Smoke1 = 1 if smoking in the past, 0 otherwise; Smoke2 = 1 if smoking currently, 0 otherwise).

Table 1: Posterior means for generalized linear mixed model (95% credible intervals in the parentheses)

	Model 1	Model 2
Fixed parameters: β		
Intercept	-7.070^* (-8.588, -5.653)	-7.190^* (-8.769, -5.671)
Sex (male versus female)	-0.727^* (-1.413, -0.069)	-0.734^* (-1.425, -0.111)
log(Age/10)	2.114^* (1.062, 3.005)	2.104^* (1.112, 3.174)
Drink1 (past)	0.052 (-0.729, 0.828)	0.065 (-0.739, 0.885)
Drink2 (current)	-0.572^* (-1.063, -0.152)	-0.596 (-1.113, -0.122)
Smoke1 (past)	0.590 (-0.131, 1.320)	0.605^* (-0.123, 1.313)
Smoke2 (current)	1.260^* (0.568, 1.995)	1.285^* (0.558, 2.032)
GARP: γ		
γ_0 (AR(1))	0.807^* (0.756, 0.860)	0.803^* (0.750, 0.857)
γ_1 (AR(2))		-0.006 (-2.760, 2.837)
IV parameters: λ		
λ_0	2.751^* (2.245, 3.235)	2.835^* (2.393, 3.340)
DIC	23868.88	24845.49

$*$: Indicates significance at the 5 % level of significance.

We fit two models proposed in Section 2. Model 1 is a random effects logistic regression with AR(1) structure of the random effects covariance matrix. Model 2 is the same model with AR(2) structure of the covariance matrix. The posterior means, 95% confidence intervals, and the DIC values for Models 1 and 2 are provided in Table 1. The DIC value for Model 1 (23868.88) is smaller than Model 2 (24845.49); subsequently, this indicates that Model 1 is superior to Model 2.

In the GARP, the coefficient (γ_0) for AR(1) were positive and the credible interval was above zero which implies the significant positive relationship of random effects. The coefficient for AR(2) was not significant. This indicates that the random effects covariance matrices had homogeneous AR(1) structures. In the log(IV), the intercept (λ_0) was significant and above zero.

In the fixed effects, the coefficients of gender (Sex), age (log(Age/10)), and current drinker (Drink 2), and current smoker (Smoke2) were significant because 95% credible intervals did not contain zero. This indicates that the estimated conditional probability of metabolic syndrome was lower for males than for females, was lower in current-drinking group than in non-drinking group, and was higher in current-smoking group than in non-smoking group, respectively. The conditional probability of Metabolic Syndrome increased as age increased.

4. Discussion

We proposed a Bayesian modeling of random effects covariance matrices for a generalized linear mixed model with modified Cholesky decomposition approach. This approach allows a complex structure of covariance matrix and satisfy the positive-definiteness of random effects covariance matrix.

In the analysis of KoGES, we found that participant's gender, age, current smoking group, and current smoking group had statistically significant effects on metabolic syndrome. We also found that the random effects covariance matrices had homogeneous AR(1) structures.

Acknowledgement

We would like to thank Dr. Jungbok Lee of Korea University for providing the data and for their help in data collection and clarifying some issues with data.

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 125–134.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006). Deviance information criteria for missing data models, *Bayesian Analysis*, **1**, 651–674.

Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Chapman & Hall/CRC.

Daniels, J. M. and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data, *Biometrika*, **89**, 553–566.

Daniels, J. M. and Zhao, Y. D. (2003). Modelling the random effects covariance matrix in longitudinal data, *Statistics in Medicine*, **22**, 1631–1647.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalized linear mixed models, *Biometrika*, **88**, 973–985.

Kim, J., Kim, E., Yi, H., Joo, S., Shin, K., Kim, J., Kim, K. and Shin, C. (2006). Short-term incidence rate of hypertension in Korea middle-aged adults, *Journal of Hypertension*, **24**, 2177–2182.

Lee, K., Yoo, J. K., Lee, J. and Hagan, J. (2012). Modeling the random effects covariance matrix for the generalized linear mixed models, *Computational Statistics & Data Analysis*, **56**, 1545–1551.

Pan, J. and Mackenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies, *Biometrika*, **90**, 239–244.

Pan, J. and Mackenzie, G. (2006). Regression models for covariance structures in longitudinal studies, *Statistical Modelling*, **6**, 43–57.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation, *Biometrika*, **86**, 677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix, *Biometrika*, **87**, 425–435.

Pourahmadi, M. and Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal data, *Biometrics*, **58**, 225–231.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **64**, 583–616.

Received April 14, 2013; Revised May 14, 2013; Accepted May 21, 2013