Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.2 pp. 227 234

Hierarchical Bayes Analysis of Longitudinal Poisson Count Data

Dal Ho Kim[1] · Im Hee Shin[2] · In Sun Choi[3]

Abstract

In this paper, we consider hierarchical Bayes generalized linear models for the analysis of longitudinal count data. Specifically we intoduce the hierarchical Bayes random effects models. We discuss implementation of the Bayes procedures via Markov chain Monte Carlo(MCMC) integration techniques. The hierarchical Baye method is illustrated with a real dataset and is compared with other statistical methods.

Key words : Random effects models, hierarchical Bayes, longitudinal count data, Gibbs sampling.

1. Introduction

There is considerably recent statistical interest in the analysis of longitudinal data. Unlike cross-sectional studies, where single outcome is measured for each individual, longitudinal studies involve repeated measurements of individuals or subjects through time. This introduces a natural correlation among the measurements within a subject which must be taken into account for any statistical analysis. Also, longitudinal studies can detect changes over time within individuals which cross-sectional studies cannot.

Diggle, Liang and Zeger(1994) provided a comprehensive frequentist analysis of longitudinal data. They extended the generalized linear model(GLM) concepts to the analysis of longitudinal data. In particular they have considered marginal,

1. Associate Professor, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea
E-mail : dalkim@knu.ac.kr
2. Assistant Professor, Division of Medical Statistics, Cathooolic University of Taegu, Taegu 705-718, Korea.
3. Lecturer, Department of Statistics, Kyungpook National University, Tegu 702-701, Korea.

random effects and transition models, and have provided a variety of frequentist analyses for all these models.

For the random effects models, Waclawiw and Liang(1994) considered empirical Bayes(EB) approach and Ghosh, Kim and Maiti(1997) have considered hierarchical Bayes(HB) approach to the analysis of longitudinal binary data based on generalized linear mixed models(GLMM). Our objective is to introduce HB GLMM for the analysis of longitudinal count data.

The random effects model introduces correlations within a subject over time. Assigning distributions to the prior hyperparameters as done in a HB analysis also builds dependence between subjects across different time periods. This enables one to "borrow strength" from other subjects as well as across time.

The outline of the remaining sections is as follow. Section 2 introduces a general HB random effects model. Section 3 provides the HB analysis of a real dataset and compares with the other analyses using other comparable methods.

2. Bayesian Generalized Linear Mixed Models

Suppose that Y_{ij} denotes the response of the ith subject (or individual) at the jth time $(j = 1, \cdots, n_i\ ;\ i = 1, \cdots, m)$. We consider the following HB models:

(I) $f(y_{ij}|\theta_{ij}) = \exp[\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\}/\phi_{ij}]h(y_{ij};\phi_{ij})$

where $\phi_{ij}(>0)$ are assumed to be known, $\theta_{ij} = \log\lambda_{ij}$, $\lambda_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{d}_{ij}^T\boldsymbol{u}_i + \log t_{ij},$, and $\boldsymbol{d}_{ij}$ is subset of $\boldsymbol{x}_{ij}$ and $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \cdots, u_{iq})^T$.

(II) $\boldsymbol{u}_i \overset{iid}{\sim} N(0, \mathbf{G})$.

(III) $\boldsymbol{\beta}$ and $\mathbf{G}$ are marginally independent with $\boldsymbol{\beta} \sim$ Uniform (R^p) and $\boldsymbol{G} \sim$ Inverse Wishart (S, k).

Note that an offset, $\log t_{ij}$, was introduced to take account of different interval lengths in the log-linear model. Here the Inverse Wishart distribution has the probability density function (pdf) of the form

$$f(\mathrm{G}) \propto |\mathrm{G}|^{-(q+k)/2}\exp(-\mathrm{trace}(\mathrm{SG}^{-1})/2).$$

The GLMM was also considered in Breslow and Clayton(1993) as well as Zeger and Karim(1991) in a related but different context.

The HB method is implemented via the MCMC integration technique. (cf. Gelfand and Smith(1990)). This requires generation of samples from the following full conditionals.

$$(i)\ [\boldsymbol{\beta}|\theta, \boldsymbol{G}, \boldsymbol{y}] \sim N\left(\sum_i\sum_j \frac{(\psi(\theta_{ij}) - \log t_{ij})}{\boldsymbol{d}_{ij}^T\boldsymbol{G}\,\boldsymbol{d}_{ij}}\ \boldsymbol{x}_{ij},\ \left(\sum_i\sum_j \frac{\boldsymbol{x}_{ij}\,\boldsymbol{x}_{ij}^T}{\boldsymbol{d}_{ij}^T\boldsymbol{G}\,\boldsymbol{d}_{ij}}\right)^{-1}\right)$$

$$(ii)\ p(\theta_{ij}|\theta_{kl(k\neq i,\,l\neq j)}, \boldsymbol{G}, \boldsymbol{\beta}, \boldsymbol{y}) \propto \exp\left\{ y_{ij}\theta_{ij} - \psi(\theta_{ij}) - \frac{(\psi(\theta_{ij}) - \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} - \log t_{ij})^2}{2\,\boldsymbol{d}_{ij}^{T}\boldsymbol{G}\,\boldsymbol{d}_{ij}} \right\}$$

$$(iii)\ p(\boldsymbol{G}|\theta, \boldsymbol{\beta}, \boldsymbol{y}) \propto \Pi_i\Pi_j(\boldsymbol{d}_{ij}^{T}\boldsymbol{G}\,\boldsymbol{d}_{ij})^{-1/2} \exp\left\{ \frac{[\psi(\theta_{ij}) - \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} - \log t_{ij}]^2}{2\,\boldsymbol{d}_{ij}^{T}\boldsymbol{G}\,\boldsymbol{d}_{ij}} \right\}$$
$$\times |\boldsymbol{G}|^{-(q+k)/2} \exp\{-\operatorname{trace}(\boldsymbol{S}\boldsymbol{G}^{-1})/2\}$$

It is easy to generate samples from the full conditionals given in (i). However, (ii) and (iii) are not a standard density from which one can generate samples easily. This difficulty is overcome by employing the Metropolis-Hastings algorithm. An alternative approach to generate sample form (ii) would be to use the adaptive rejection sampling(ARS) of Gilks and Wild(1992) since $\pi(\theta_{ij}|\cdot)$ are all log-concave. But the latter is not pursued here.

3. Data Analysis

We provide in this section a HB analysis of a real dataset given Thall and Vail (1990) and by Breslow and Clayton(1993). For each patient, the number of epileptic seizures was recorded during a baseline period of eight weeks. Patients were then randomized to treatment with the anti-epileptic drug progabide, or to placebo in addition to standard chemotherapy. The number of seizures was then recorded in four consecutive two-week intervals. Let Y_{ij} denote the Poisson counts response from 0 to infinite corresponding to occur the epileptic seizures. The objective is whether or not the progabide reduces the rate of epileptic seizures. The successive seizure counts for 59 patients. Covariates are treatment (0, 1), 8-week baseline seizure counts, and age in years.

We first consider the following models. Model 1 is a log-linear model with a random intercept as follows.

(I) $\log E(Y_{ij}|u_i) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij1}x_{ij2} + u_i + \log(t_{ij})$

$i = 1, \cdots, 59, \quad j = 0, \cdots, 4$

where $x_{ij1} = \begin{cases} 1 & \text{if the ith subjects is assigned to the progabide group} \\ & \text{if the ith subjects is assigned to the placebo group} \end{cases}$

$x_{ij2} = \begin{cases} 1 & \text{if } j = 1,2,3, \text{ or } 4 \\ & \text{if } j = 0 \end{cases}$

(II) $u_i \overset{iid}{\sim} N(0, r_u^{-1})$

(III) $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and r_u are marginally independent with $\boldsymbol{\beta} \sim$ Uniform (R^4) and $r_u \sim$ Gamma(a/2, b/2).

Then the joint pdf of θ, $\boldsymbol{\beta}$, and r_u given y_{ij} is given by

$$f(\theta, \boldsymbol{\beta}, r_u | y_{ij}) \propto \prod_i \prod_j [\exp\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\}]$$
$$\times \prod_i \prod_j r_u^{1/2} \exp[-\frac{r_u}{2}(\theta_{ij} - \log t_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta})^2]$$
$$\times \prod_i \prod_j r_u^{b/2} \exp\{-(r_u + a)/2\}.$$

To implement the MCMC integration techniques in Model 1, we need to generate samples from the full conditional distributions like below:

$$(i)\ [r_u | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{y}] \sim \text{Gamma}(\{a + (\sum_i \sum_j \theta_{ij} - \sum_i \sum_j \log t_{ij} - \sum_i \sum_j \boldsymbol{x}_{ij}^T \boldsymbol{\beta})^2\}/2, \{\sum u_i + b\}/2)$$

$$(ii)\ [\boldsymbol{\beta} \mid \boldsymbol{\theta}, \boldsymbol{y}, r_u] \sim N\left(r_u \sum_i \sum_j (\theta_{ij} - \log t_{ij})\ \boldsymbol{x}_{ij}^T,\ r_u^{-1} \sum_i \sum_j (\boldsymbol{x}_{ij}\ \boldsymbol{x}_{ij}^T)^{-1}\right)$$

$$(iii)\ p(\theta_{ij} | \theta_{kl}((k,l) \neq (i,j)), \boldsymbol{\beta}, r_u, \boldsymbol{y}) \propto \exp\{y_{ij}\theta_{ij} - \psi(\theta_{ij}) - r_u(\theta_{ij} - \log t_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta})/2\}$$

To complete the hierarchical Model 1, we assign a Uniform (R^4) prior for $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$, and Gamma $(0.0005/2, 0.0005/2)$ prior for r_u.

In Model 2, we add a second random effect for the pre/post-treatment indicator (x_2) as follows.

$$\log E(Y_{ij} | u_i) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij1} x_{ij2} + u_{i1} + x_{ij2} u_{i2} + \log(t_{ij})$$

where $\boldsymbol{u}_i = (u_{i1}, u_{i2})^T$ is assumed to follow a bivariate normal distribution with mean $\boldsymbol{0}$ and variance matrix $\boldsymbol{G}$ with elements $\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$ Moreover, the hyperpriors are assumed that $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and $\boldsymbol{G}$ are marginally independent with $\boldsymbol{\beta} \sim$ Uniform (R^4) and $\boldsymbol{G} \sim$ Invers Wishart $(\boldsymbol{S}, k)$. Denote $\boldsymbol{d}_{ij} = (1, x_{ij2})^T$. The inclusion of u_{i2} allows us to address the concern that there might be heterogeneity among subjects in the ratio of the expected seizure counts before and after the randomization. The degree of heterogeneity can be measured by the magnitude of G_{22}, the variance of u_{i2}.

To implement the MCMC integration techniques in Model 2, we generate samples from the full conditional distributions as follows:

$$(i)\ [\boldsymbol{\beta} | \theta, \boldsymbol{G}, \boldsymbol{y}] \sim N\left(\sum_i \sum_j \frac{\psi(\theta_{ij}) - \log t_{ij}}{\boldsymbol{d}_{ij}^T \boldsymbol{G}\ \boldsymbol{d}_{ij}}\ \boldsymbol{x}_{ij},\ \left(\sum_i \sum_j \frac{\boldsymbol{x}_{ij}\ \boldsymbol{x}_{ij}^T}{\boldsymbol{d}_{ij}^T \boldsymbol{G}\ \boldsymbol{d}_{ij}}\right)^{-1}\right)$$

$$(ii)\ p(\theta_{ij} | \theta_{kl(k \neq i, l \neq j)}, \boldsymbol{G}, \boldsymbol{\beta}, \boldsymbol{y}) \propto \exp\{y_{ij}\theta_{ij} - \psi(\theta_{ij}) - (\psi(\theta_{ij}) - \boldsymbol{x}_{ij}^T \boldsymbol{\beta} - \log t_{ij})^2 / \{2\ \boldsymbol{d}_{ij}^T \boldsymbol{G}\ \boldsymbol{d}_{ij}\}\}$$

$$(iii)\quad p(G_{11}|G_{12},G_{22},\theta,\boldsymbol{\beta},\boldsymbol{y}) \propto \prod_i\prod_j (G_{11}+2x_{ij2}G_{12}+x_{ij2}^2G_{22})^{-1/2} \times \exp\left\{\frac{(\psi(\theta_{ij})-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}-\log t_{ij})^2}{2(G_{11}+2x_{ij2}G_{12}+x_{ij2}^2G_{22})}\right\} \times |G_{11}G_{22}-G_{12}^2|^{-(q+k)/2}\exp\left\{\frac{1}{2}\frac{G_{11}+G_{22}}{G_{11}G_{22}-G_{12}^2}\right\}$$

$$(iv)\quad p(G_{12}|G_{11},G_{22},\theta,\underline{\beta},\underline{y}) \propto \prod_i\prod_j (G_{11}+2x_{ij2}G_{12}+x_{ij2}^2G_{22})^{-1/2} \times \exp\left\{\frac{(\psi(\theta_{ij})-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}-\log t_{ij})^2}{2(G_{11}+2x_{ij2}G_{12}+x_{ij2}^2G_{22})}\right\} \times |G_{11}G_{22}-G_{12}^2|^{-(q+k)/2}\exp\left\{\frac{1}{2}\frac{G_{11}+G_{22}}{G_{11}G_{22}-G_{12}^2}\right\}$$

$$(v)\quad p(G_{22}|G_{12},G_{22},\theta,\underline{\beta},\underline{y}) \propto \prod_i\prod_j (G_{11}+2x_{ij2}G_{12}+x_{ij2}^2G_{22})^{-1/2} \times \exp\left\{\frac{(\psi(\theta_{ij})-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}-\log t_{ij})^2}{2(G_{11}+2x_{ij2}G_{12}+x_{ij2}^2G_{22})}\right\} \times |G_{11}G_{22}-G_{12}^2|^{-(q+k)/2}\exp\left\{\frac{1}{2}\frac{G_{11}+G_{22}}{G_{11}G_{22}-G_{12}^2}\right\}$$

To complete the hierarchical Model 2, we assign a Uniform (R^4) prior for $\underline{\beta}=(\beta_0,\beta_1,\beta_2,\beta_3)^T$ and Invers Wishart $\left(\begin{pmatrix}0.0005 & 0\\ 0 & 0.0005\end{pmatrix},7\right)$ prior for $\boldsymbol{G}$.

In our setting G_{ij}, $i=1,2$; $j=1,2$ is a positive real number, but M-H algorithm has all real number space, i.e. the range for M-H algorithm is $(-\infty,\infty)$. Hence we need to modify the M-H algorithm like bellow.

Table 1: Estimate and standard errors (in parentheses) for the progabide data with 49th observation.

	Model 1			Model 2		
Variable	AML	HB_MH	HB_BUGS	AML	HB_MH	HB_BUGS
Intercept(β_0)	1.0	1.3012	1.5012	1.1	1.0344	1.0182
	(0.15)	(0.3302)	(0.1302)	(0.14)	(0.2322)	(0.1134)
Treatment(β_1)	-0.023	-0.0225	-0.0325	0.050	0.0574	0.0636
	(0.20)	(0.1502)	(0.1120)	(0.18)	(0.1134)	(0.1526)
Time(β_2)	0.11	0.1102	0.1623	0.002	0.0030	0.00189
	(0.047)	(-0.0417)	(0.0547)	(0.11)	(0.0838)	(0.0875)
Trt by time(β_3)	-0.10	-0.0986	-0.1092	-0.31	-0.2938	-0.2961
	(0.065)	(0.0584)	(0.0658)	(0.015)	(0.01175)	(0.2152)
G_{11}	0.62	0.6145	0.6614	0.51	0.5117	0.5242
	(0.12)	(0.1364)	(0.2364)	(0.10)	(0.1499)	(0.1089)
G_{12}	-	-	-	0.054	0.0667	0.0623
				(0.056)	(0.0462)	(0.0489)
G_{22}	-	-	-	0.24	0.2189	0.2395
				(0.062	(0.0761)	(0.1034)

Modified Metropolis-Hasting Algorithm : Since G_{ij} $(i = 1,2 : j = 1,2)$ is a variable with range in positive real line, we can use a transformation such as $G'_{ij} = \log G_{ij}$, to map $(0, \infty)$ into $(-\infty, \infty)$, then use the transition kernel and applying of the M-H algorithm to the density of G'_{ij}. After one transition of the Metropolis-Hasting algorithm is done then we transform G'_{ij} back to the original scale means of $G_{ij} = \exp G'_{ij}$.

For both models, a burn-in of 7500 iterations was followed by a further 15000 iteration. Table 1 is based on the complete data and Table 3 is based on the data without 49th observation. These results are based on Gibba sampling with 5 chain and 15000 replication following Gelman and Rubin(1992).

The BUGS provides a declarative language for straightforward specification of statistical models based on the assumed graphical structure, although there are some restrictions on the class of models that can be analysed currently. A compiler then processes the model and data and sets up the sampling distributions required for the Gibbs sampling. Finally, appropriate sampling algorithms are implemented to simulate values of th unknown quantities in the model. To implement BUGS for both Model 1 and Model 2, we assign a $N(\mathbf{0}, 10^4 \boldsymbol{I})$ prior instead of a Uniform (R^4) prior for $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$. Also both models were fitted using the approximaye maximum likelihood(AML) algorithm. See Diggle, Liang and Zeger(1994) for detailed calculations.

Table 2 : Estimate and standard errors (in parentheses) for the progabide data without 49th observation.

	Model 1			Model 2		
Variable	AML	HB_MH	HB_BUGS	AML	HB_MH	HB_BUGS
Intercept(β_0)	1.0	1.0012	1.2022	1.1	1.1222	1.0843
	(0.14)	(0.1302)	(0.1252)	(0.13)	(0.1532)	(0.1341)
Treatment(β_1)	-0.009	-0.0121	-0.0134	-0.029	-0.0194	-0.0342
	(0.19)	(0.1975)	(0.0455)	(0.19)	(0.1248)	(0.0109)
Time(β_2)	0.11	0.1100	0.1346	0.10	0.0135	0.0531
	(0.047)	(0.0724)	(0.2746)	(0.11)	(0.0854)	(0.0567)
Trt by time(β_3)	-0.30	-0.3181	-0.2897	-0.34	-0.3326	-0.3453
	(0.070)	(0.06243)	(0.07423)	(0.15)	(0.1141)	(0.2133)
G_{11}	0.53	0.5497	0.6030	0.46	0.5512	0.5762
	(0.10)	(0.2186)	(0.2314)	(0.10)	(0.1614)	(0.0945)
G_{12}				0.014	0.0657	0.0532
				(0.053)	(0.0437)	(0.0185)
G_{22}				0.22	0.2349	0.2384
				(0.059)	(0.0782)	(0.0228)

In the result from Table 1 and Table 2, three methods are quite comparable. For Model 1, there is modest evidence that progabide is more effective than the placebo in reducing the occurrence of seizures ($\widehat{\beta}_3 = -0.0982 \pm 0.0584$). With possible outlier patient number 49 deleted, stronger treatment effect is suggested ($\widehat{\beta}_3 = -0.3181 \pm 0.06242$). Focusing on the results of Model 2 fitted to the complete data, subjects in the placebo group have expected seizure rate after treatment which are estimated to be roughly the same as before treatment ($\exp(\widehat{\beta}_2) = \exp(0.003) = 1.003$). For the progabide group, the seizure rates are reduced after treatment by about 25.5 percent ($1 - \exp(0.003 - 0.293) = 0.255$). Hence, the treatment seems to have a modest effect. The estimated effect is $\widehat{\beta}_3 = -0.293$ with a standard error of.117. Finally, if we set aside of the moment patient number 49, who had unusually high seizure rates, ($\widehat{\beta}_3 = -0.3326 \pm 0.1141$). The analysis without patient 49 is only exploratory, and is carried out in order to understand this patients's influence on th seizure counts and perhaps has special medical problem.

References

1. Breslow, N.E. and Clayton, D.G. (1993) Approximate in generalized linear mixed model. *Journal of the American Statistical Association*, 88, 9-25.
2. Ghosh, M., Kim, D.H., and Maiti T. (1997). Hierarchical Bayesian analysis of longitudinal data. *Sankhya, Ser. B*, 59, 326-334.
3. Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.
4. Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.
5. Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. Applied Statistics, 41, 337-348.
6. Gelfan, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
7. Thall, P.F. and Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*. 46, 657-671.
8. Walawiw, M.A. and Liang, K.Y. (1994). Empirical Bayes estimation and inference for the random effects model with binary response. *Statistics in Medicine*, 13, 542-551.
9. Zeger, S.L. and Karim, M.R(1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79-86.

[2002년 3월 접수, 2002년 10월 채택]