Communications of the Korean Statistical Society
2009, Vol. 16, No. 6, 971–978

Maximum Likelihood Estimation Using Laplace Approximation in Poisson GLMMs

Il Do Ha[1,a]

[a]Department of Asset Management, Daegu Haany University

Abstract

Poisson generalized linear mixed models(GLMMs) have been widely used for the analysis of clustered or correlated count data. For the inference marginal likelihood, which is obtained by integrating out random effects, is often used. It gives maximum likelihood(ML) estimator, but the integration is usually intractable. In this paper, we propose how to obtain the ML estimator via Laplace approximation based on hierarchical-likelihood (h-likelihood) approach under the Poisson GLMMs. In particular, the h-likelihood avoids the integration itself and gives a statistically efficient procedure for various random-effect models including GLMMs. The proposed method is illustrated using two practical examples and simulation studies.

Keywords: H-likelihood, laplace approximation, marginal likelihood, generalized linear mixed models, random effects.

1. Introduction

Poisson GLMMs, Poisson generalized linear models (GLMs; Nelder and Wedderburn, 1972) with normally distributed random effects, have been widely used for the analysis of clustered or correlated count data (Breslow and Clayton, 1993). For the inference likelihood-based (Breslow and Clayton, 1993; Booth and Hobert, 1999) and Bayesian (Besag *et al.,* 1995; Efron, 1996) methods have been usually used.

In this paper we are interested in ML inference under Poisson GLMMs which gives a good asymptotic property (Jiang, 2007). However, the marginal likelihood involves intractable integrals whose dimension depends on the structure of random effects. To overcome this problem, various numerical approximation methods have been proposed; for example, expectation maximization(EM), Monte Carlo EM(MCEM), Gauss-Hermite quadrature(GHQ) approximation and the Laplace approximation(LA). The EM method still requires an integration for the E-step. Even though MCEM avoids integration by using a Monte Carlo method, it is also still computationally intensive (Gueorguieva, 2001). The GHQ method is not available for the models more than two random-effect terms (Huber *et al.,* 2004).

Breslow and Clayton (1993) proposed the penalized quasi-likelihood(PQL) method based on the LA of the marginal likelihood which is easy to implemented. However, the PQL methods leads to biased estimates, particularly for dispersion parameters (Jiang, 2007). In this paper we propose the use of LA using the h-likelihood method (Lee and Nelder, 1996, 2001). In particular, the h-likelihood avoids the difficult integration itself and gives a statistically efficient procedure for various random-effect models including GLMMs (Lee *et al.*, 2006). In this paper, we consider the Poisson GLMMs

This work was supported by the Korea Research Foundation Grant funded by the Korean Government(KRF-2008-521-C00057).

[1] Professor, Department of Asset Management, Daegu Haany University, Gyeongsan, 712-715, Korea.
E-mail: idha@dhu.ac.kr

with one-random component. The proposed LA method is compared with the marginal GHQ method based on SAS NLMIXED procedure.

The paper is organized as follows. In Section 2 we briefly describe the structures of data and models. In Section 3 we show how to obtain the LA estimators via the h-likelihood. The proposed method is demonstrated with two practical examples and simulation study in Section 4. Finally, some concluding remarks are given in Section 5.

2. The Model

Let y_{ij} $(i = 1, \ldots, q,\ j = 1, \ldots, n_i,\ n = \sum_i n_i)$ be the count response for the j^{th} observation of the i^{th} individual (or cluster). Denoted by v_i the unobserved random effect of the i^{th} individual. Then the Poisson GLMMs with one random-effect term v_i are described as follows:

(i) Given v_i, y_{ij} are independent and follow the Poisson distribution with mean μ_{ij} having

$$\log \mu_{ij} = x_{ij}^T \beta + v_i. \tag{2.1}$$

(ii) $v_i \sim N(0, \alpha)$ and v_i's are independent.

Here $x_{ij} = (1, x_{ij1}, \ldots, x_{ijp})^T$ is a vector of fixed covariates and β is a $(p+1) \times 1$ vector of fixed effects. The α is called dispersion parameter. Note that if $\alpha = 0$ (i.e. $v_i = 0$ for all i) then the above model becomes the Poisson GLM.

3. Laplace Approximated ML Estimator

Following Lee and Nelder (1996), the h-likelihood for the model (2.1), denoted by h, is defined by

$$h = h(\beta, v, \alpha) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}, \tag{3.1}$$

where

$$\ell_{1ij} = \ell_{1ij}(\beta; y_{ij}|v_i) = y_{ij} \log \mu_{ij} - \mu_{ij} - \log y_{ij}!$$

is the logarithm of the conditional density function for y_{ij} given v_i, and

$$\ell_{2i} = \ell_{2i}(\alpha; v_i) = -\frac{1}{2} \log(2\pi\alpha) - \frac{1}{2\alpha} \sum_i v_i^2$$

is the logarithm of the density function for v_i. Here $v = (v_1, \ldots, v_q)^T$. The corresponding marginal likelihood m can be obtained by integrating out the random effects from the h-likelihood:

$$m = m(\beta, \alpha) = \sum_i \log \left\{ \int \exp(h_i)\, dv_i \right\}, \tag{3.2}$$

where $h_i = \sum_j \ell_{1ij} + \ell_{2i}$ is the contribution of the i^{th} individual to h in (3.1). The maximizing the marginal likelihood m gives ML estimators. However, the integration in (3.2) can't be computed explicitly except for Poisson-gamma model (Lee and Nelder, 1996). Thus, an approximation of this integral is needed. Thus, we consider the first-order Laplace approximation (LA1), $p_v(h)$, to m. Following Barndorff-Nielsen and Cox (1989, p.60), as $n^* = \min_{1 \le i \le q} n_i \to \infty$ we have

$$m = p_v(h) + O\left(n^{*-1}\right) \tag{3.3}$$

with

$$p_v(h) = \left[h - \frac{1}{2}\log\det\left\{\frac{D(h,v)}{(2\pi)}\right\}\right]\Bigg|_{v=\hat{v}},$$

where $D(h,v) = -\partial^2 h/\partial v^2$ and $\hat{v}$ solves $\partial h/\partial v = 0$. Note that $p_v(h)$ in (3.3) produces an adjusted profile h-likelihood for (β,α) after eliminating random effects v (Lee and Nelder, 2001). Thus, the LA1 estimator is obtained by maximizing $p_v(h)$; it can be expressed as

$$p_v(h) = \hat{h} - \frac{1}{2}\log\det\left(\hat{D}\right) + \frac{q}{2}\log(2\pi), \tag{3.4}$$

where $\hat{h} = h|_{v=\hat{v}}$ is a profile likelihood for (β,α) after eliminating v, $\hat{D} = D(h,\hat{v}) = Z^T\hat{W}Z + U$, $\hat{W} = \mathrm{diag}(\hat{\mu})$ is a diagonal weight matrix with $n \times 1$ main diagonal vector $\hat{\mu} = \exp(X\beta + Z\hat{v})$ whose ij^{th} element is $\mu_{ij} = \exp(x_{ij}^T\beta + v_i)$, and $U = \alpha^{-1}I_q$ with q dimensional identity matrix. Here, X is the $n \times (p+1)$ model matrix whose i^{th} row vector is x_{ij}^T and Z is the $n \times q$ group indicator matrix whose i^{th} row vector is z_{ij}^T.

3.1. Estimation of fixed effects

Given α, the LA1 estimators for fixed effects β are obtained by solving

$$\frac{\partial p_v(h)}{\partial \beta_k} = 0 \quad (k = 1,\ldots,p). \tag{3.5}$$

The equation on left hand side of (3.5) is computed by the following procedure. From (3.4) we have

$$\frac{\partial p_v(h)}{\partial \beta_k} = \frac{\partial \hat{h}}{\partial \beta_k} - \frac{1}{2}\mathrm{tr}\left(\hat{D}^{-1}\frac{\partial \hat{D}}{\partial \beta_k}\right).$$

Here

$$\frac{\partial \hat{h}}{\partial \beta_k} = \frac{\partial h}{\partial \beta_k}\bigg|_{v=\hat{v}}$$

since $\partial\hat{h}/\partial\beta_k = \{(\partial h/\partial\beta_k) + (\partial h/\partial v)(\partial\hat{v}/\partial\beta_k)\}|_{v=\hat{v}}$ and $(\partial h/\partial v)|_{v=\hat{v}} = 0$: see also Appendix 1 of Ha, Lee and Song (2001), and $\partial h/\partial\beta_k = \sum_{ij}(y_{ij} - \mu_{ij})x_{ijk} = (y-\mu)^T X_k$ with the k^{th} column vector X_k of X, and

$$\frac{\partial \hat{D}}{\partial \beta_k} = Z^T\hat{W}_k' Z,$$

where $\hat{W}_k' = \mathrm{diag}[\hat{\mu}_{ij}\{X_k + Z(\partial\hat{v}/\partial\beta_k)\}]$. Note here that following Appendix C of Lee and Nelder (1996), we also have

$$\begin{aligned}\frac{\partial \hat{v}}{\partial \beta_k} &= -\left(\frac{-\partial^2 h}{\partial v^2}\right)^{-1}\left(\frac{-\partial^2 h}{\partial v \partial \beta_k}\right)\bigg|_{v=\hat{v}} \\ &= -\left(Z^T\hat{W}Z + U\right)^{-1}\left(Z^T\hat{W}X_k\right).\end{aligned}$$

In this paper we use the Newton-Raphson to solve (3.5), with the following negative second derivatives

$$-\frac{\partial^2 p_v(h)}{\partial\beta_k\partial\beta_l} = -\frac{\partial^2 \hat{h}}{\partial\beta_k\partial\beta_l} + \frac{1}{2}\mathrm{tr}\left(-\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\beta_k}\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\beta_l} + \frac{\partial^2\hat{D}}{\partial\beta_k\partial\beta_l}\right). \tag{3.6}$$

Here

$$-\frac{\partial^2 \hat{h}}{\partial\beta_k\partial\beta_l} = X_k^T\hat{W}X_l - X_k^T\hat{W}Z\left(Z^T\hat{W}Z + U\right)^{-1}Z^T\hat{W}X_l$$

and

$$\frac{\partial^2\hat{D}}{\partial\beta_k\partial\beta_l} = Z^T\hat{W}_{kl}''Z,$$

where $\hat{W}_{kl}'' = \mathrm{diag}[\hat{\mu}_{ij}\{X_k + Z(\partial\hat{v}/\partial\beta_k)\}\{X_l + Z(\partial\hat{v}/\partial\beta_l)\} + \{Z(\partial^2\hat{v}/\partial\beta_k\partial\beta_l)\}]$ with

$$\frac{\partial^2\hat{v}}{\partial\beta_k\partial\beta_l} = -\left(Z^T\hat{W}Z + U\right)^{-1}Z^T\hat{W}_k'\left\{X_l + Z\frac{\partial\hat{v}}{\partial\beta_l}\right\}.$$

3.2. Estimation of dispersion parameter

Similarly, the LA1 dispersion estimator for α is obtained by solving

$$\frac{\partial p_v(h)}{\partial\alpha} = 0. \tag{3.7}$$

Here

$$\frac{\partial p_v(h)}{\partial\alpha} = \left.\frac{\partial h}{\partial\alpha}\right|_{v=\hat{v}} - \frac{1}{2}\mathrm{tr}\left(\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha}\right),$$

where $\partial h/\partial\alpha = \sum_i \partial\ell_2/\partial\alpha = \sum_i\{-1/(2\alpha) + v_i^2/(2\alpha^2)\}$, and

$$\frac{\partial\hat{D}}{\partial\alpha} = Z^T\left(\frac{\partial\hat{W}}{\partial\alpha}\right)Z - \alpha^{-2}I_q,$$

where $\partial\hat{W}/\partial\alpha = \mathrm{diag}\{\hat{\mu}_{ij}Z(\partial\hat{v}/\partial\alpha)\}$ and $\partial\hat{v}/\partial\alpha = (Z^T\hat{W}Z)^{-1}(\alpha^{-2}\hat{v})$.

To solve (3.7), as in previous section we also use the Newton-Raphson, with the negative second derivative

$$-\frac{\partial^2 p_v(h)}{\partial\alpha^2} = -\frac{\partial^2\hat{h}}{\partial\alpha^2} + \frac{1}{2}\mathrm{tr}\left(-\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha}\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha} + \frac{\partial^2\hat{D}}{\partial\alpha^2}\right). \tag{3.8}$$

Often $p_v(h)$ does not provide sufficiently accurate approximation to m when the cluster size n_i are small. If so, the second-order Laplace approximation is recommended (Shun, 1997; Lee and Nelder, 2001). As $n^* = \min_{1\le i\le q} n_i \to \infty$ we again have

$$m = s_v(h) + O\left(n^{*-2}\right).$$

Here

$$s_v(h) = p_v(h) - \frac{F(h)}{24}$$

and

$$F(h) = \sum_{i=1}^{q} \left\{ -3\left(\frac{\partial^4 h}{\partial v_i^4}\right) b_{ii}^2 - 5\left(\frac{\partial^3 h}{\partial v_i^3}\right)^2 b_{ii}^3 \right\} \Bigg|_{v=\hat{v}},$$

where b_{ii} is the i^{th} diagonal element of $D(h, v)^{-1}$; in model (2.1) $b_{ii} = \sum_j \mu_{ij} + \alpha^{-1}$ and $\partial^3 h/\partial v_i^3 = \partial^4 h/\partial v_i^4 = -\sum_j \mu_{ij}$. In this paper, we call the dispersion estimator of α maximizing $s_v(h)$, the second-order Laplace approximation(LA2) estimator; it is also obtained by solving

$$\frac{\partial s_v(h)}{\partial \alpha} = 0.$$

We have found that for the estimation of β, the use of $p_v(h)$ is enough because it performs well. Thus, in this paper the LA2 method uses $s_v(h)$ for α, but $p_v(h)$, not $s_v(h)$, for β.

3.3. Variance estimation for parameter estimators

Following the usual ML inference, the asymptotic covariance matrix of $\hat{\beta}$ and $\hat{\alpha}$ is obtained from the inverse of information matrix, $-\partial^2 m/\partial\psi^2$ with $\psi = (\beta, \alpha)^T$. However, the integration in m is again intractable, so that we use the first-order approximation $p_v(h)$. Thus, in this paper the variances of $\hat{\psi}$ are estimated from the main diagonal elements of inverse of $H = -\partial^2 p_v(h)/\partial\psi^2$, given by

$$H = \begin{pmatrix} H_1 & H_2 \\ H_2^T & H_3 \end{pmatrix}. \tag{3.9}$$

Here, $H_1 = -\partial^2 p_v(h)/\partial\beta^2$ with entries given in (3.6), $H_3 = -\partial^2 p_v(h)/\partial\alpha^2$ is given in (3.8) and $H_2 = -\partial^2 p_v(h)/\partial\beta\partial\alpha$ with its entries

$$-\frac{\partial^2 p_v(h)}{\partial\beta_k\partial\alpha} = -\frac{\partial^2 \hat{h}}{\partial\beta_k\partial\alpha} + \frac{1}{2}\mathrm{tr}\left(-\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\beta_k}\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha} + \frac{\partial^2\hat{D}}{\partial\beta_k\partial\alpha}\right).$$

4. Illustration

The proposed method is illustrated using two examples and simulation studies. Here we focus on comparisons of LA1 and LA2 estimators. For the further comparison, we also include the marginal GHQ method using SAS NLMIXED procedure. For the model fitting and computation, we used SAS/IML.

4.1. Examples

Example 1. **(Pump failure data)** Gaver and O'Muircheartaigh (1987) presented a small data set about failures of 10 pumps. The number of failures and the period of operation were recorded for each of 10 pumps. We fit the Poisson GLMM where the fixed effect is group effect, the offset is the logarithm of the period of operation and the random effect is each pump. The results in Table 1 show that for the fixed effects β_0 and β_1 the LA1 and LA2 estimates are about the same as GHQ estimates,

Table 1: Results on the estimation of parameters for the pump failure data

Method	$\hat{\beta}_0$ (SE)	$\hat{\beta}_1$ (SE)	$\hat{\alpha}$ (SE)
LA1	−2.047 (0.507)	1.687 (0.696)	0.901 (0.510)
LA2	−2.048 (0.510)	1.686 (0.700)	0.916 (0.524)
GHQ	−2.048 (0.510)	1.689 (0.700)	0.914 (0.523)

Note: LA1, the first-order Laplace approximation; LA2, the second-order Laplace approximation; GHQ, the marginal Gauss-Hermite quadrature method using SAS PROC NLMIXED; β_0, intercept; β_1, group effect; α, variance of random effect; SE, the estimated standard error.

Table 2: Results on the estimation of parameters for the epileptics data (single covariate)

Method	$\hat{\beta}_0$ (SE)	$\hat{\beta}_1$ (SE)	$\hat{\alpha}$ (SE)
LA1	1.772 (0.182)	−0.294 (0.253)	0.877 (0.178)
LA2	1.772 (0.183)	−0.294 (0.254)	0.880 (0.179)
GHQ	1.772 (0.183)	−0.294 (0.254)	0.880 (0.179)

Note: β_0, intercept; β_1, new drug effect; α, variance of random effect.

Table 3: Results on the estimation of parameters for the epileptics data (two covariates)

Method	$\hat{\beta}_0$ (SE)	$\hat{\beta}_1$ (SE)	$\hat{\beta}_2$ (SE)	$\hat{\alpha}$ (SE)
LA1	2.867 (1.927)	−0.311 (0.255)	−0.327 (0.573)	0.874 (0.177)
LA2	2.867 (1.931)	−0.312 (0.255)	−0.327 (0.574)	0.877 (0.179)
GHQ	2.867 (1.931)	−0.312 (0.255)	−0.327 (0.574)	0.877 (0.178)

Note: β_0, intercept; β_1, new drug effect; β_2, age effect; α, variance of random effect.

but that for the dispersion parameter α the LA2 estimate is very closer to the GHQ estimate. However, the LA1 estimate for α is smaller than the corresponding LA2 and GHQ estimates.

Example 2. **(Data on epileptics)** This example is based on the longitudinal seizure count data from a clinical trial which consists of four repeated measures of 59 epileptics, presented by Thall and Vail (1990). For the data we perform two analyses. For the first analysis, we fit the Poisson GLMM with a single covariate only, indicating a new drug (Trt = 1) or placebo (Trt = 0). In the second analysis, we consider two covariates, the Trt and age. Other covariates are also available but are omitted. The corresponding results are presented in Tables 2 and 3, respectively. Overall, the trends of the results are similar to those evident in Table 1. However, the three methods lead to similar estimates for both β and α. A possible reason is that the data set has a larger sample size $n = 236$ with $q = 59$ and $n_i = 4$.

4.2. Simulation studies

Simulated studies, based on 200 replications of simulated data, are presented to compare the LA1 and LA2 methods. Data are generated from the Poisson GLMM (2.1) assuming fixed effects $\beta = (\beta_0, \beta_1) = (1, -1)$ and dispersion parameter (i.e. variance of normal random effect) $\alpha = 1$. Here, we set a single covariate x_{ij} to be 0 for the first $q/2$ individuals (control group), and x_{ij} to be 1 for the remaining $q/2$ individuals (treatment group). We also set the sample size $n = \sum_{i=1}^{q} n_i = 40$ and 200, which correspond to $q = 20$ with $n_i = 2$ and $q = 50$ with $n_i = 4$, respectively. For the 200 replications we computed the mean, standard deviation(SD), the mean(SEM) of the estimated standard errors(SEs) for $\hat{\beta}$ and $\hat{\alpha}$. Here, the SEs are obtained from (3.9).

The simulation results are summarized in Table 4. As expected from Section 4.1, three methods give about the same results for the estimation of β, and LA2 method provides almost identical results to the GHQ method, particular for a larger sample size $n = 200$. Overall, the LA2 method works well

Table 4: Simulation results about the estimation of parameters using 200 replications under true fixed effects $\beta = (\beta_0, \beta_1) = (1, -1)$ and true variance of random effect $\alpha = 1$

		$\hat{\beta}_0$			$\hat{\beta}_1$			$\hat{\alpha}$		
n	Method	Mean	SD	SEM	Mean	SD	SEM	Mean	SD	SEM
40	LA1	0.992	0.361	0.335	−0.942	0.483	0.497	0.902	0.479	0.420
	LA2	0.989	0.363	0.339	−0.943	0.479	0.502	0.927	0.495	0.439
	GHQ	0.990	0.362	0.338	−0.942	0.478	0.501	0.923	0.490	0.435
200	LA1	1.011	0.225	0.207	−1.014	0.311	0.303	0.952	0.268	0.244
	LA2	1.010	0.225	0.208	−1.015	0.311	0.305	0.964	0.274	0.249
	GHQ	1.010	0.225	0.208	−1.014	0.311	0.305	0.965	0.273	0.248

Note: Mean and SD, the mean and standard deviation for $\hat{\beta}$ and $\hat{\alpha}$, respectively; SEM, the mean of estimated standard errors for $\hat{\beta}$ and $\hat{\alpha}$.

compared to the LA1 method. In particular, LA1 method gives more biases for $\hat{\alpha}$. In Table 4, for $\psi = (\beta, \alpha)^T$ the SD is the estimate of the true $\{\text{var}(\hat{\psi})\}^{1/2}$ and SEM is the average of SE estimate for $\hat{\psi}$. The proposed SE estimates perform well as judged by the good agreement between SD and SEM as sample size n increases.

5. Concluding Remarks

We have shown that the LA method is very useful for the ML estimation of parameters in Poisson GLMMs where the marginal likelihood is directly not obtained. In particular, we have found that the LA2 estimates are about the same as the resulting GHQ estimates from SAS NLMIXED procedure. However, the LA1 method gives similar estimates to the LA2 method as sample size increases, but it leads to more biases for α in a small sample. Thus, in this paper we recommend the use of the LA2 method for the model (2.1).

The proposed method can be straightforwardly extended to GLMMs (2.1) with binary or binomial responses. Even though our method was developed for the model with one random-effect term, extension to GLMMs with more than one random component will be possible because the proposed method avoids intractable integrations. Furthermore, the development of the LA methods for semi-parametric frailty models (Ha *et al.,* 2001; Ha *et al.,* 2007) would be also an interesting future work.

References

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic techniques for use in Statistics*, Chapman and Hall, New York.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3–66.

Booth, J. G. and Hobert, J. P. (1999). Maximum generalized linear mixed model likelihood with an automated Monte Carlo EM algorithm, *Journal of the Royal Statistical Society B*, **61**, 265–285.

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

Efron, B. (1996). Empirical Bayes methods for combining likelihoods (with discussion). *Journal of the American Statistical Association*, **91**, 538–565.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates, *Technometrics*, **29**, 1–15.

Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family, *Statistical Modelling*, **1**, 177–193.

Ha, I. D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models, *Biometrika*, **88**, 233–243.

Ha, I. D., Lee, Y. and MacKenzie, G. (2007). Model selection for multi-component frailty models, *Statistics in Medicine*, **26**, 4790–4807.

Huber, P., Ronchetti, E. and Victoria-Feser, M.-P. (2004). Estimation of generalized linear latent variable models, *Journal of the Royal Statistical Society B*, **66**, 893–908.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*, Springer, New York.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society B*, **58**, 619–678.

Lee, Y. and Nelder, J. A. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions, *Biometrika*, **88**, 987–1006.

Lee, Y., Nelder, J. A. and Pawitan (2006). *Generalized linear models with random effects*, Chapman and Hall, New York.

Nelder, J. A. and Wedderburn (1972). Generalized linear models (with discussion), *Journal of the Royal Statistical Society A*, **135**, 370–384.

Shun, Z. (1997). Another look at the salamander mating data: a modified Laplace approximation approach, *Journal of the American Statistical Association*, **92**, 341–349.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion, *Biometrics*, **46**, 657–671.

Received July 2009; Accepted September 2009