Journal of the Korean
Data & Information Science Society
2001, Vol. 12, No.1, pp. 19 ~ 25

Second-Order REML for Random Effects Models

Il Do Ha · Geon-Ho Cho [1]

Abstract

Random effects models which describe the dependence via random effects in various correlated data have recently received considerable attention in the biomedical literature. They include mixed linear models (MLMs), generalized linear mixed models (GLMMs) and hierarchical generalized linear models (HGLMs). For the inference Lee and Nelder (2000) proposed the first- and second-order REML (restricted maximum likelihood) methods based on hierarchical-likelihood of Lee and Nelder (1996). In this paper, for Poisson-gamma HGLMs the new methods are theoretically compared with marginal likelihood methods and both methods are illustrated by two practical examples.

*Key Words and Phrases:*Hierarchical-likelihood, Marginal likelihood, Poisson-gamma HGLMs, Profile likelihood, Random effects, Restricted maximum likelihood.

1. Introduction

Correlated data including longitudinal or repeated measurements data are frequently encountered in biomedical research. A number of authors have proposed using the models with random effect, a common unobservable effect within the same individual, to account for the dependence between the correlated data. For the inference the Bayesian approaches (e.g. Besag et al., 1995; Efron, 1996) have been extensively studied. Many researchers have insisted (e.g. Bjornstad, 1996) that the marginal likelihood should be used unless the random effects are of inferential interest. However, problems with numerical integration associated with the use of the marginal likelihood have been a major obstacle. As an alternative, Lee and Nelder (1996) proposed the use of hierarchical-likelihood (h-likelihood) that avoids this difficulty; its procedure gives a conceptually simple and numerically efficient algorithm. It can be further applied to structured dispersion, spatial and temporal models involving

[1]Professor, Faculty of Information Science, Kyungsan University, Kyungpook, 712-240, Korea.

non-normal data, and can also be used to give an efficient inference procedure for both frailty models and MLMs with multivariate censored data: see Lee and Nelder (2000) and Ha et al. (2001a, b). Lee and Nelder (2000) also showed by numerical study that the first- and second-order REML procedures based on h-likelihood yield reasonably good parameter estimators in all the case they studied. In this paper, for Poisson-gamma HGLMs we investigate the properties of the new methods by two practical examples.

In Section 2 we review both Lee and Nelder (2000) and marginal likelihood methods for the random effects models. For Poisson-gamma HGLMs, both methods are theoretically compared in Section 3 and are illustrated by two practical examples in Section 4.

2. The first- and second-order REMLs

Let Y_{ij} be the response variable for the jth ($j = 1, \cdots, n_i$) observation of the ith ($i = 1, \cdots, q$) individual (or cluster) and U_i be the unobserved random effect of the ith individual.

The random effects models are described as follows:

(i) $Y_{ij}|U_i = u_i \sim$ a conditional density $f_1(\cdot|u_i; \eta_{ij}, \phi)$ with the linear predictor $\eta_{ij} = x_{ij}^t\beta + v_i$, where $x_{ij} = (x_{ij1}, \cdots, x_{ijp})^t$ is a vector of fixed covariates, β is a $p \times 1$ vector of fixed effects and $v_i = g(u_i)$ with some strictly monotone function $g(\cdot)$.

(ii) $U_i \sim$ a density $f_2(\cdot; \alpha)$.

Here, $\theta = (\phi, \alpha)$ are dispersion parameters. Then Lee and Nelder's (1996) h-likelihood for the models is defined by

$$h = h(\beta, \theta) = \sum_{ij} \ell_{1ij} + \sum_{i} \ell_{2i}, \qquad (1)$$

where $\ell_{1ij} = \ell_{1ij}(\beta, \phi; y_{ij}|u_i) = \log\{f_1(y_{ij}|u_i; \beta, \phi)\}$ and $\ell_{2i} = \ell_{2i}(\alpha; v_i) = \log\{f_2(v_i; \alpha)\}$. The corresponding marginal (log-)likelihood is given by

$$m = m(\beta, \theta) = \sum_{i} \log\Big\{\int \exp(h_i)\ dv_i\Big\}, \qquad (2)$$

where $h_i = \sum_j \ell_{1ij} + \ell_{2i}$ is the contribution of the ith individual to h in (1).

Assume that the dispersion parameters θ are given. Let $h_i^{(k)} = -\partial^k h/\partial v_i^k$ and $\tilde{h}_i^{(k)} = h_i^{(k)}(\tilde{v}_i)$, where $\tilde{v}_i = \tilde{v}_i(\beta)$ is the solution from $\partial h/\partial v_i = 0$ for given β. Also, let $v = (v_1, \cdots, v_q)^t$. Following Reid (1991), the first- and the second-order Laplace approximations to m of (2) are, respectively, given by

$$m = \hat{m} + O_p(n_i^{-1}) \ \text{ and } \ m = \hat{m}_2 + O_p(n_i^{-2}), \qquad (3)$$

where $\hat{m} = [h - \frac{1}{2}\log\det\{(-\partial^2 h/\partial v^2)/2\pi\}]\,|_{v=\tilde{v}}$, $\hat{m}_2 = \hat{m} + C_2$, $C_2 = -\,\mathrm{trace}\{\mathrm{diag}(S_i)\}/24$, and $\mathrm{diag}(S_i)$ is the $q \times q$ diagonal matrix whose ith element is

$$S_i = 3\frac{\tilde{h}_i^{(4)}}{[\tilde{h}_i^{(2)}]^2} - 5\frac{[\tilde{h}_i^{(3)}]^2}{[\tilde{h}_i^{(2)}]^3}.$$

For MLMs m is exactly $\hat{m}$ because of $h_i^{(3)} = h_i^{(4)} = 0$. Note that the maximum h-likelihood estimators (MHLEs) of $\tau = (\beta, v)$ are obtained from $\partial h/\partial \tau = 0$ and that the MHLE for β becomes an approximation of the marginal MLE obtained from $\partial m/\partial \beta = 0$: see Lee and Nelder (1996).

For the estimation of the dispersion parameters θ given estimates of τ, Lee and Nelder (2000) used the first- and second-order adjusted profile h-likelihood (APHL). They are, respectively, defined by

$$h_p = [\ h - \frac{1}{2}\log\det\{(-\partial^2 h/\partial \tau^2)/2\pi\}\ ]\ |_{\tau=\hat{\tau}} \ \text{ and } \ h_{p2} = h_p + C_{p2}, \tag{4}$$

where $C_{p2} = C_2|_{\tau=\hat{\tau}}$ with C_2 in (3) and $\hat{\tau} = \hat{\tau}(\theta) = (\hat{\beta}(\theta), \hat{v}(\theta))$ with $\hat{v}(\theta) = \tilde{v}(\beta)|_{\beta=\hat{\beta}(\theta)}$. Note that APHL is an extension of REML (Patterson and Thompson, 1971) for MLMs. The first-and second-order REML estimators for θ are, respectively, obtained by solving

$$\frac{\partial h_p}{\partial \theta} = 0 \ \text{ and } \ \frac{\partial h_{p2}}{\partial \theta} = 0. \tag{5}$$

The estimated variances for θ are, respectively, obtained from inverse of $-\partial^2 h_p/\partial \theta^2$ and $-\partial^2 h_{p2}/\partial \theta^2$. Because of orthogonality between v_i and θ, Lee and Nelder (1996) ignored terms in $\partial \hat{v}_i/\partial \theta$. Thus, their original REML estimator, which maximize h_p, may lead to downward bias, particularly in some extreme cases such as binary data with small n_i. Because the number of random effects often grows with the sample size $n = \sum_i n_i$, it is important to allow for $\partial \hat{v}_i(\theta)/\partial \theta$, but not for $\partial \hat{\beta}(\theta)/\partial \theta$, in implementing two REML estimating equations (5): see Lee and Nelder (2000).

On the other hand, Cox-Reid's (1987) adjusted profile marginal likelihood is defined by

$$m_p = [\ m - \frac{1}{2}\log\det\{(-\partial^2 m/\partial \beta^2)/2\pi\}\ ]\ |_{\beta=\hat{\beta}}. \tag{6}$$

The marginal likelihood and Cox-Reid estimators for θ are obtained by solving $\partial m/\partial \theta = 0$ and $\partial m_p/\partial \theta = 0$, respectively; the latter generally improves the former in terms of bias. In general, h_p is a good approxmation to m_p: see Lee and Nelder (2000). However, both m in (2) and m_p in (6) often require numerical integration.

3. Poisson-gamma HGLMs

We consider the Poisson-gamma HGLMs with log-link, an example of the random effects models: for $i = 1, \cdots, q$ and $j = 1, \cdots, n_i$,
(i) $Y_{ij}|U_i = u_i \sim$ Poisson(μ'_{ij}) with $\mu'_{ij} = \mu_{ij}u_i$,
where $\eta_{ij} = \log(\mu'_{ij}) = \log\mu_{ij} + v_i$ with $\log\mu_{ij} = x^t_{ij}\beta$ and $v_i = \log(u_i)$.
(ii) $U_i \sim$ gamma distribution with mean 1 and variance α;
its density is of the form $\{\Gamma(1/\alpha)\alpha^{1/\alpha}\}^{-1}u_i^{(1/\alpha)-1}\exp(-u_i/\alpha)$.
The h-likelihood for the models is given by

$$h = h(\beta,\alpha) = \sum_{ij}(y_{ij}\log\mu_{ij} + y_{ij}v_i - \mu_{ij}u_i) + \sum_i\{(v_i - u_i)/\alpha + c(\alpha)\},$$

where $c(\alpha) = -\log\Gamma(\alpha^{-1}) - \alpha^{-1}\log\alpha$.

We now derive C_2 in (3), which is the term for the second-order Laplace approximation to m using h. Assume that the dispersion parameter α is given.
From

$$\partial h/\partial v_i = \sum_j(y_{ij} - \mu_{ij}u_i) \ + \ \alpha^{-1} - \alpha^{-1}u_i = 0,$$

we have

$$\tilde{u}_i = \tilde{u}_i(\beta) = \frac{\alpha^{-1} + \ y_{i+}}{\alpha^{-1} + \ \mu_{i+}}\ , \tag{7}$$

which also becomes $E(U_i|y_i)$ from the fact that the conditional distribution of U_i given the ith observed data $y_i = (y_{i1}, \cdots, y_{in_i})$ is gamma. Here, $y_{i+} = \sum_j y_{ij}$ and $\mu_{i+} = \sum_j\mu_{ij}$. Since $h_i^{(2)} = h_i^{(3)} = h_i^{(4)} = (\alpha^{-1} + \ \mu_{i+})u_i$, from (7) we obtain $\tilde{h}_i^{(2)} = \tilde{h}_i^{(3)} = \tilde{h}_i^{(4)} = \alpha^{-1} + \ y_{i+}$ and $S_i = -2(\alpha^{-1} + \ y_{i+})^{-1}$. Thus we have that

$$C_2 = \frac{1}{12}\text{trace}[\text{diag}\{(\alpha^{-1} + \ y_{i+})^{-1}\}]. \tag{8}$$

¿From (4) and (8) the second-order APHL is given by

$$h_{p2} = h_p + C_2,$$

where $C_2 = C_{p2}$ because C_2 in (8) does not depend on τ. We use the Newton-Raphson method to solve the second-order REML score equation $\partial h_{p2}/\partial\alpha = 0$. Here, from (7) we have $\partial\hat{v}_i/\partial\alpha = (\partial\hat{u}_i/\partial\alpha)/\hat{u}_i$ with $\partial\hat{u}_i/\partial\alpha = \{\alpha^{-2}(\hat{u}_i - 1)\}/(\alpha^{-1} + \hat{\mu}_{i+})$.

¿From (2) the corresponding marginal likelihood is explicitly given by

$$m = m(\beta,\alpha) = \sum_{ij} y_{ij}\log\mu_{ij} - \sum_i(\alpha^{-1} + y_{i+})\log(\alpha^{-1} + \mu_{i+}) + \sum_i\{\log\Gamma(\alpha^{-1} + y_{i+}) + c(\alpha)\}.$$

Whenever marginal likelihood is explicitly obtainable we find the following three facts. Firstly, the maximum marginal likelihood (MML) estimating equations for β for given α,

$$\frac{\partial m}{\partial\beta_k} = \sum_{ij}\left\{y_{ij} - \mu_{ij}\left(\frac{\alpha^{-1} + \ y_{i+}}{\alpha^{-1} + \ \mu_{i+}}\right)\right\} x_{ijk} = 0 \ (k = 1, \ldots, p),$$

are equivalent to the MHL estimating equations

$$\frac{\partial h}{\partial \beta_k} = \sum_{ij} (y_{ij} - \mu_{ij} u_i) x_{ijk} = 0 \ (k = 1, \dots, p)$$

with $\tilde{u}_i = (\alpha^{-1} + y_{i+})/(\alpha^{-1} + \mu_{i+})$ in (7). That is, the MHLE for β given α is the same as the maximum marginal likelihood estimator (MMLE). Secondly, given α, the variance estimate of $\widehat{\beta}$ calculated from derivatives of h agrees with that from derivatives of m. That is, our estimate for var($\hat{\beta}$) is the same as $(-\partial^2 m/\partial \beta^2)^{-1}|_{\beta=\hat{\beta}}$. Finally, it is easily shown that h_p becomes m_p, after replacing some gamma functions by their Stirling approximations. Though not reported here, for the Poisson-gamma models we have found by simulation that the second-order estimator using h_{p2} is reasonably good and is about the same as the Cox-Reid estimator using m_p.

4. Examples

To illustrate both Lee and Nelder (2000) and marginal likelihood methods, we now present two practical examples. For each example we fit the Poisson-gamma HGLMs using five methods; ML (marginal likelihood), C-R (Cox-Reid's adjusted marginal likelihood), old (Lee and Nelder, 1996), the first-order and second-order REMLs. In these examples we focus on comparisons of five methods for inferences of the parameters, rather than finding a reasonable model. For the model fittings we used SAS/IML.

Example 1: *Pump failure data.* Gaver and O'Muircheartaigh (1987) presented a small data set about failures of 10 pumps. We fit the Poisson-gamma models where the fixed effects are group effects, the offset is the logarithm of the period of operation and the random effects are one for each pump. The results in Table 1 show that for the fixed effects β_0 and β_1 the five methods give about the same results, but that for dispersion parameter α both the ML and old methods lead to smaller estimates than those of the other three methods. In particular, we find that the second-order REML method gives substantially the same results as the Cox and Reid method.

Example 2: *Data on epileptics.* This example is based on the longitudinal seizure count data from a clinical trial of 59 epileptics, presented by Thall and Vail (1990). For these data we perform two analyses. For the first analysis, we fit the Poisson-gamma models with only a single covariate, indicating a new drug (Trt=1) or placebo (Trt=0). In the second analysis we consider two covariates, the Trt and age. Other covaiates are also available but are omitted. The results are presented in Table 2(a) and Table 2(b), respectively. Overall, the trends of the results are similar to those given in Example 1.

Table 1. *Analyses for the pump failure data of Gaver & O'Muircheartaigh* (1987)

Method	$\hat{\beta}_0$	SE	$\hat{\beta}_1$	SE	$\hat{\alpha}$	SE
ML	-1.603	0.472	1.671	0.640	0.770	0.372
C-R	-1.594	0.513	1.666	0.692	0.933	0.468
old	-1.599	0.492	1.668	0.666	0.848	0.380
1st	-1.595	0.509	1.666	0.687	0.915	0.458
2nd	-1.594	0.513	1.666	0.692	0.933	0.468

ML, marginal likelihood; C-R, adjusted marginal likelihood of Cox and Reid (1987); old, Lee and Nelder (1996); 1st, the first-order REML; 2nd, the second-order REML. $\hat{\beta}_0$ is the estimate of intercept β_0, $\hat{\beta}_1$ is that of group(2) effect β_1, and $\hat{\alpha}$ is that of dispersion parameter α. SE, the estimated standard error.

Table 2. *Analyses for the data on epileptics of Thall & Vail* (1990)

	(a) Single covariate			(b) Two covariates			
Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$
ML	2.152	-0.077	0.901	2.630	-0.146	-0.016	0.894
	(0.182)	(0.252)	(0.157)	(0.666)	(0.267)	(0.021)	(0.155)
C-R	2.152	-0.077	0.928	2.630	-0.146	-0.016	0.935
	(0.185)	(0.255)	(0.163)	(0.681)	(0.273)	(0.021)	(0.165)
old	2.152	-0.077	0.883	2.630	-0.146	-0.016	0.889
	(0.180)	(0.249)	(0.147)	(0.665)	(0.266)	(0.021)	(0.149)
1st	2.152	-0.077	0.925	2.630	-0.146	-0.016	0.932
	(0.185)	(0.255)	(0.162)	(0.680)	(0.273)	(0.021)	(0.164)
2nd	2.152	-0.077	0.928	2.630	-0.146	-0.016	0.935
	(0.185)	(0.255)	(0.163)	(0.681)	(0.273)	(0.021)	(0.165)

ML, marginal likelihood; C-R, adjusted marginal likelihood of Cox and Reid (1987); old, Lee and Nelder (1996); 1st, the first-order REML; 2nd, the second-order REML. $\hat{\beta}_0$ is the estimate of intercept β_0, $\hat{\beta}_1$ is that of a new drug effect β_1, $\hat{\beta}_2$ is that of age effect β_2, and $\hat{\alpha}$ is that of dispersion parameter α. (·) indicates the estimated standard error (SE).

References

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

2. 2. Bjornstad, J. F. (1996). On the generalization of the likelihood function and likelihood principle. *Journal of the American Statistical Association*, 91, 791-806.

3. 3. Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society* B, 49, 1-39.

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

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

6. 6. Ha, I. D., Lee, Y. and Song, J.-K. (2001a). Hierarchical likelihood approach for frailty models. to appear in *Biometrika.*

7. 7. Ha, I. D., Lee, Y. and Song, J.-K. (2001b). Hierarchical likelihood approach for mixed linear models with censored data. to appear in *Lifetime Data Analysis.*

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

9. 9. Lee, Y. and Nelder, J. A. (2000). Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions. Technical Report, Imperial College, London.

10. 10. Patterson, H. D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, 58, 545-554.

11. 11. Reid, N. (1991). Approximations and asymptotics. *Statistical Theory and Modelling*, ed. by D. V. Hinkley, N. Reid and E. J. Snell, Chapman & Hall, London.

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