Comparison of Hierarchical and Marginal Likelihood Estimators for Binary Outcomes[1]

Sungcheol Yun[2], Youngjo Lee[2], Il-Do Ha[3], Weechang Kang[4]

Abstract

Likelihood estimation in random-effect models is often complicated because the marginal likelihood involves an analytically intractable integral. Numerical integration such as Gauss-Hermite quadrature is an option, but is generally not recommended when the dimensionality of the integral is high. An alternative is the use of hierarchical likelihood, which avoids such burdensome numerical integration. These two approaches for fitting binary data are compared and the advantages of using the hierarchical likelihood are discussed. Random-effect models for binary outcomes and for bivariate binary-continuous outcomes are considered.

Key Words: Binary data; Gauss-Hermite quadrature; Hierarchical likelihood; Joint model; Marginal likelihood; Random effects

1. Introduction

Correlated data are frequently observed in practice, for example in longitudinal studies, clinical trials and familial studies. Generalized linear models with random effects have become increasingly popular for the analysis of such correlated data, for example the hierarchical generalized linear models (HGLMs) of Lee and Nelder (1996, 2001a). In making inferences from HGLMs, a marginal-likelihood analysis is often burdened by intractable integration. To avoid this, various approximate and Bayesian inferential procedures have been proposed. The approximations include the Gauss-Hermit quadrature (GHQ) (Hinde, 1982; Crouch and Spiegelman, 1990; Longford, 1994), and EM methods. Because the EM method has often difficulty in implementing the expectation step, Monte Carlo EM (MCEM) methods have been proposed (McCulloch, 1997; Booth and Hobert, 1999). The Bayesian procedures include EM-type algorithms (Stiratelli, Laird and Ware, 1984) and the Gibbs sampler (Zeger and Karim, 1991). However, all these methods involve intensive computation.

In this paper we implement the h-likelihood based Laplace approximation for fixed effects, and compare it with the GHQ method for binary outcomes and for bivariate binary and continuous outcomes. Gueorguieva (2001) compared the MCEM and GHQ methods and recommended the GHQ when the number of random components is less than three, otherwise the MCEM algorithm, even though computationally intensive, may be the only

feasible choice. Neither time-consuming numerical integrations such as GHQ nor Monte Carlo methods to implement the expectation step are necessary to produce good estimates.

2. HGLMs

The specification of HGLMs is done in two stage: (i) Conditional on the random effects v, the data y are assumed to follow a distribution from a GLM family. (ii) Assumptions about the distribution of the random effects v complete the specification.

2.1 Hierarchical-likelihood method

For inferences on HGLMs, Lee and Nelder (1996) proposed to use a h-(log-)likelihood of the form

$$h = \log(f(y|v;\beta)) + \log(f(v;\sigma)),$$

where $f(y|v;\beta)$ and $f(v;\sigma)$ denote the conditional density function of y given v and the density function of v, respectively. Throughout the paper likelihood is used to mean log-likelihood unless otherwise stated. In forming the h-likelihood the choice of the scale of random effects is important. Note that v is the scale on which the random effects are assumed to occur linearly in the linear predictor. In this paper we focus on the estimation of the unknown parameters β and σ.

The marginal likelihood L can be obtained via the integration,

$$L = \log \int \exp(h) \, dv.$$

Analytical forms of L are available only in a few special cases, for example mixed linear models with a normal distribution.

Let l be a likelihood (either a marginal likelihood L or an h-likelihood h) with nuisance effects α. Lee and Nelder (2001a) considered a function $p_\alpha(l)$, defined by

$$p_\alpha(l) = [l - \frac{1}{2} \log\det(\frac{1}{2} D(l,\alpha)/\pi)]|_{\alpha = \hat{\alpha}},$$

where $D(l,\alpha) = -\partial^2 l/\partial\alpha^2$ and $\hat{\alpha}$ solves $\partial l/\partial\alpha = 0$. For fixed effects β the use of $p_\beta(L)$ is equivalent to conditioning on $\hat{\beta}$ (Cox and Reid, 1987), while for random effects v the use of $p_v(h)$ is equivalent to integrating them out by using the Laplace approximation.

Ideally we may use the h-likelihood h for inferences about v, the marginal likelihood L for β and the restricted likelihood $p_\beta(L)$ for σ if L is easily available. However, when L is hard to obtain, as respectively approximations to L and $p_\beta(L)$, we may use

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

where $D(h,v) = -\partial^2 h/\partial v^2$ and $\hat{v}$ solves $\partial h/\partial v = 0$ and

$$p_\delta(h) = [h - \frac{1}{2} \log\det(\frac{1}{2} D(h,\delta)/\pi)]|_{\delta = \hat{\delta}},$$

where $\delta = (v,\beta)$, $D(h,\delta) = -\partial^2 h/\partial\delta^2$ and $\hat{\delta}$ solves $\partial p_v(h)/\partial\delta = 0$; Lee and Nelder

(2001a) showed that $p_\delta(h)$ is approximately $p_\beta(p_v(h))$ and provides numerically reasonable dispersion estimators for HGLMs.

Lee and Nelder (1996) observed that deviance differences constructed from h and $p_v(h)$ are often very similar, so they (1996, 2001a) used h for estimating β and observed that the resulting estimators generally work very well; see the simulation studies of Poisson and binomial HGLMs (Lee and Nelder, 2001a), frailty models (Ha, Lee and Song, 2001) and mixed linear models with censoring (Ha, Lee and Song, 2002). We call the resulting estimators the maximum h-likelihood estimators (MHLEs). However, they observed that the MHLEs can have non-negligible biases for some binary data. In this paper we implement the fitting algorithm of β using $p_v(h)$ and compare the resulting estimators with other estimators.

2.2 The marginal-likelihood method

The marginal likelihood L often has no explicit form. A commonly used method to approximate L would be GHQ. It replaces the integral in (2) by a finite weighted sum. When the dimension (the number of random components) of the integrals is small, it can be used to approximate the likelihood closely (Crouch and Spiegelman, 1990). However, the error induced by replacing the integral with finite sum becomes more and more difficult to control and exponentially time consuming as the dimension increases. The MCEM method for finding the marginal-likelihood estimate provides an alternative to the GHQ method (Gueorguieva, 2001). These iterative methods can handle high-dimensional integrals better than the GHQ method. However, they require fairly sophisticated computer programs and are very computationally intensive. Compared with methods using marginal likelihood those using h-likelihood are analytically straightforward and computationally easy.

3. Simulation Studies

Numerical studies, based upon 200 replication of simulated data, are presented to evaluate the performance of the proposed estimation procedures. For each parameter combination of the simulated data we compute the bias and mean square error (MSE). We compare three estimators, MHLE, an estimator maximizing adjust profile likelihood $p_v(h)$, and an estimator maximizing the marginal likelihood L via an adaptive GHQ approximation. For the adaptive GHQ method we use the procedure NLMIXED of SAS (Wolfinger, 1999) with 20 quadrature points. The resulting estimators are denoted by MHLE, $p_v(h)$ and GHQ, respectively. For methods using h-likelihood, MHLE or $p_v(h)$ for estimating the fixed effects, we maximize $p_\delta(h)$ to obtain the estimates of dispersion parameters.

4. Example

Price et al. (1985) presented data from a study on the developmental toxicity, ethylene glycol (EG), in mice. The rates of fetal malformation increase with dose, ranging from 0.3%

in the control group to 57% in the highest dose (3g/kg/day) group. Fetal weight decreases with increasing dose, with the average weight ranging from 0.972g in the control group to 0.704g in the highest dose group. Gueorguieva (2001) introduced a joint HGLM for the analysis of the EG data. He used the GHQ and MCEM methods. All the covariates (Dose and $Dose^2$) of this study are between-subject covariates. Gueorguieva (2001) neglected the quadratic trend of dose in the HGLM for binary outcomes because of its insignificance. He considered the quadratic trend only for the HGLM for continuous outcomes. However, we found that it is necessary for both HGLMs i.e. the quadratic trend becomes significant if it appears in both HGLMs.

We present the results from the $p_v(h)$ and the GHQ method. These data have I=94 with varying cluster sizes whose average $\bar{J}$=10.9. The SAS NLMIXED procedure for the GHQ takes more than 20 minutes on a PC with Pentium 3 processor and 128 megabytes of RAM, while our $p_v(h)$ procedure takes less than 5 minutes. The implementation of the method in Lee and Nelder (2001a) uses the Aitken (1926) extrapolation procedure, which greatly reduces the number of iterations. We do not use the Aitken procedure here in order to have a fair comparison, but time for the $p_v(h)$ procedure could be further reduced by its use.

5. Concluding Remarks

Up to now many researches have been devoted to implementing methods based upon marginal-likelihood. However, fitting models using methods such as GHQ and MCEM is often computationally very intensive. This computational burden becomes heavier as the number of random components increases. For example, the GHQ method in SAS NLMIXED procedure is best suited for models with a single random effect, which currently does not generally handle nested or crossed random effects (Wolfinger, 1999). This difficulty limits the wide application of these methods. By applying the h-likelihood, one can avoid such troublesome integration because there is an explicit analytical form for the likelihood. The h-likelihood, we believe, is a natural extension of the Fisher likelihood to models with random parameters. It provides a simple unified framework for random-effect models that offers a numerically and statistically efficient fitting algorithm (Lee and Nelder, 2001a). It also provides new solutions for various problems. For example, it can be applied to predict future observations (Pawitan, 2001, Chapter 16), to generate a new sandwich variance estimator (Lee, 2002), to analyze survival data (Ha, Lee and Song, 2001; Ha, Lee and Song, 2002) and to carry out meta analysis (Lee and Nelder, 2002); it offers unified simple inferential tools for HGLMs (Lee and Nelder, 1996, 2001a), extendable to the analysis of temporally and spatially correlated data (Lee and Nelder, 2001b). One major objection has been the occurrence of large biases in parameter estimators for binary data. In this paper, we see that a modification, such as the use of the Laplace approximation, gets rid of most of the biases. H-likelihood will be the basis of an important method for the analysis of random-effect models for correlated data. The latest GenStat release (Payne et al., 2002) has an implementation of the MHLE estimators for arbitrarily numbers of random components.

References

Aitken, A.C., 1926. On Bernoulli's numerical solution of algebraic equations. Proc. Roy. Soc. Edinburgh, 46, 289-305.

Booth, J. and Hobert, J., 1999. Maximizing generalized linear mixed model likelihoods with an automated monte carlo EM algorithm. J. R. Statist. Soc. B, 62, 265-285.

Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Am. Statist. Ass., 88, 9-25.

Breslow, N.E. and Lin, X., 1995. Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika, 82, 81-91.

Cox, D.R. and Reid, N., 1987. Parameter orthogonality and approximate conditional inference. J. R. Statist. Soc. B, 32, 1-18.

Crouch, E.A.C. and Spiegelman, D., 1990. The evaluation of integrals of the form $\int_{-\infty}^{+\infty} f(t)\exp(-t^2)dt$: application to logistic-normal models. J. Am. Statist. Ass., 85, 464-469.

Gueorguieva, R.V., 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 Song, J.-K., 2002. Hierarchical likelihood approach for mixed linear models with censored data. Lifetime Data Anal., 8, 163-176.

Hinde, J., 1982. Compound poisson regression models. In R Gilchrist (Ed.), GLIM 82, New York, 109-121. Springer-Verlag.

Johnson, R.A. and Wichern, D.W., 1992. Applied Multivariate Statistical Analysis, 3rd edn. Englewood Cliffs:Prentice Hall.

Lee, Y., 2002. Robust variance estimators for fixed-effect estimates with hierarchical likelihood. Statistics and Computing, 12, 201-207.

Lee, Y. and Nelder, J.A., 1996. Hierarchical generalized linear models (with discussion). J. R. Statist. Soc. B, 58, 619-678.

Lee, Y. and Nelder, J.A., 2001a. Hierarchical generalized linear models : A synthesis of

generalized linear models, random-effect model and structured dispersion. Biometrika, 88, 987-1006.

Lee, Y. and Nelder, J.A., 2001b. Modelling and analysing correlated non-normal data. Statistical Modelling, 1, 3-16.

Lee, Y. and Nelder, J.A., 2002. Analysis of the ulcer data using hierarchical generalized linear models. Statist. Med., 21, 191-202.

Lesaffre, E. and Spiessens, B., 2001. On the effect of the number of quadrature points in a logistic random-effects model: an example. Appl. Statist., 50, 325-335.

Liu, Q. and Pierce, D.A., 1993. Heterogeneity in Mantel-Haenszel-type models. Biometrika, 80, 543-546.

Longford, N.T., 1994. Logistic regression with random coefficients. Comput. Stat. Data. Anal., 49, 975-988.

McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. J. Am. Statist. Ass., 92, 162-170.

Payne, R.W., Baird, D.B., Cherry, M., Gilmour, A.R., Harding, S.A., Lane, P.W., Morgan, G.W., Murray, D.A., Soutar, D.M., Thompson. R., Todd, A.D., Tunnicliffe Wilson, G., Webster, R. and Welham, S.J., 2002. The Guide to GenStat, Part 2: Statistics. Oxford: VSN International.

Pawitan, Y., 2001. In all likelihood : statistical modelling and inference using likelihood. Oxford: Clarendon Press.

Price, C.J., Kimmel, C.A., Tyle, R.W. and Marr, M.C., 1985. The developmental toxicity of Ethylene Glycol in rates and mice. Toxicological Applications in Pharmacololgy., 81, 113-127.

Stiratelli, R., Laird, N. and Ware, J., 1984. Random effect model for serial observations with binary response. Biometric, 40, 961-971.

Wolfinger, R.D., 1999. Fitting nonlinear mixed models with the new NLMIXED procedure. Paper 287, Proceedings of the 99 Joint Statistical Meetings.

Zeger, S.L. and Karim, M.R., 1991. Generalized linear models with random effects: A Gibbs sampling approach. J. Am. Statist. Ass., 86, 79-86.