Bayesian Analysis (2009) **4**, Number 2, pp. 213–216

Comments on Article by Yin

Ciprian M. Crainiceanu*

Inferential methods for Generalized Linear Mixed Models (GLMMs) are under intense methodological development because they: 1) are widely applicable; and 2) raise non-trivial technical and inferential challenges. The Generalized Method of Moments (GMM) (Hansen (1982); Newey and West (1987)) provides a powerful and robust set of inferential tools for GLMMs, especially when the likelihood formulation is difficult and interest is centered on the fixed effects parameters.

The paper by Yin (2009) is an important contribution to this literature. The main idea of the paper is to provide a simple Bayesian framework for what I considered to be a frequentist method, par excellence. I found the paper thought provoking, fresh and definitely worthy of discussion. Below I summarize my reactions and comments and provide a set problems that could be, but are not currently, addressed by this methodology.

1 Why?

The most important question in my mind after reading the paper was "Why should we use Bayesian GMM instead of GMM?" Simulations seem to indicate that both methods produce similar results, with the Bayesian methodology requiring more computational effort. One answer that I do not particularly like is "Because we can". Another possible answer could be that in some data sets with a smaller number of clusters the posterior distribution $\widetilde{\pi}(\boldsymbol{\beta}|\mathbf{y}) \propto \widetilde{L}(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta})$ might not be well approximated by a normal. In such a context, the next natural step would be to consider the sampling variability of the data by conducting a nonparametric bootstrap of the clusters. Pooled analyses using Bayesian GMM and GMM could then be compared. Some applications and simulations supporting these ideas would add credibility to the proposed methods.

2 What?

The approach proposed by Yin is to treat the quadratic objective function

$$Q_n(\boldsymbol{\beta}) = \mathbf{U}_n^T(\boldsymbol{\beta})\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})\mathbf{U}_n^T(\boldsymbol{\beta})$$

as an approximation of minus twice the log of the conditional likelihood $L(\mathbf{y}|\boldsymbol{\beta})$. More precisely, the author replaced the unknown $L(\mathbf{y}|\boldsymbol{\beta})$ by the approximate likelihood $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta}) = \exp\{-Q_n(\boldsymbol{\beta})/2\}$. When observations are not clustered $\mathbf{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^n U_i(\boldsymbol{\beta})/n$, where $U_i(\boldsymbol{\beta}) = \mathbf{D}_i v_i^{-1}(y_i - \mu_i)$, $\mathbf{D}_i = \partial\mu_i/\partial\boldsymbol{\beta}$ is the vector of derivatives of the subject i-specific mean with respect to the model parameters, and $v_i = \text{var}(y_i|\mathbf{Z}_i)$ is the conditional vari-

*Department of Biostatistics, Johns Hopkins University, Baltimore, MD, mailto:ccrainic@jhsph.edu

 DOI:10.1214/09-BA407B

ance of observations given the covariates. When observations are clustered v_i is replaced by the matrix $\boldsymbol{V}_i = \mathbf{A}_i^{1/2}\mathbf{C}_i\mathbf{A}_i^{1/2}$, where $\mathbf{A}_i = \text{diag}\{h(\mu_i)\}$, $\theta_i = h(\boldsymbol{\beta}^T\mathbf{Z}_i)$, and θ_i is the location parameter. The matrix $\mathbf{C}_i$ contains the information about the correlation structure. The author uses the idea in Qu et al. (2000) to replace $\mathbf{C}_i$ by the mixture $\alpha_1\mathbf{C}_{i,(1)} + \ldots + \alpha_J\mathbf{C}_{i,(J)}$, where $\mathbf{C}_{i,(1)}, \ldots, \mathbf{C}_{i,(J)}$ are known basis matrices and $\alpha_1, \ldots, \alpha_J$ control the importance of a particular correlation structure. For example, if $\mathbf{C}_{i,(1)}$ is the identity matrix and $\alpha_1 = 1$, $\alpha_2 = \ldots = \alpha_J = 0$ then the model assumes working independence. Given standard generalized estimating equation results (GEE) Liang and Zeger (1986), one would expect that the choice of α will affect the posterior distribution of $\boldsymbol{\beta}$, but not its posterior mean. This structure raises several questions that would require further discussion.

1. Under what conditions is $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$ a proper likelihood? It seems that in standard examples the integral $c^{-1}(\boldsymbol{\beta}) = \int \exp\{-Q_n(\boldsymbol{\beta})/2\}f(\mathbf{y}|\mathbf{Z},\boldsymbol{\beta})d\mathbf{y}$ is finite. However, having a sufficient condition for $c^{-1}(\boldsymbol{\beta}) < \infty$ would help in non-standard applications, that is when the method is most likely to provide additional insight.
2. What is the effect of ignoring the normalizing constant $c(\boldsymbol{\beta})$ in the definition of $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$? Clearly, $c^{-1}(\boldsymbol{\beta})$ cannot be calculated if the likelihood function $f(\mathbf{y}|\mathbf{Z},\boldsymbol{\beta})$ is not specified. However, the author argues convincingly that $c(\boldsymbol{\beta})$ is well approximated by $(2\pi)^{-p/2}|\boldsymbol{\Sigma}_n(\boldsymbol{\beta})|^{-1/2}$. Thus, it seems clear that the effect of ignoring $c(\boldsymbol{\beta})$ should be minimal when $\partial|\boldsymbol{\Sigma}_n(\boldsymbol{\beta})|/\partial\boldsymbol{\beta} \approx \mathbf{0}$, as it is the case in simulations. However, this may not always be the case.
3. Does including $(2\pi)^{-p/2}|\boldsymbol{\Sigma}_n(\boldsymbol{\beta})|^{-1/2}$ in the definition of $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$ lead to sampling algorithms that are significantly harder to implement?
4. In the clustered case, how are the parameters $\alpha_1, \ldots, \alpha_J$ included in the Bayesian GMM? As I mentioned above their choice (or distribution) is likely to affect the posterior distribution of $\boldsymbol{\beta}$. However, the substituted likelihood $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$ (Section 2.2) does not contain them. Moreover, the author mentions that α_j's "do not need to be sampled in the Bayesian GMM procedure".

3 When?

Standard parametric correlation matrices for random effects are those corresponding to independent, exchangeable or AR(1) structures. These correlation structures are very popular in practice because they are often *sufficient* to capture observed variability. Moreover, they are easy to implement either using Bayesian analysis, likelihood inference or GEEs. Thus, a good practice would be to start with these methods and investigate discrepancies, if any, between data and assumptions. This can be done in many ways, but here are two ideas. First, fit GEEs with various parametric structures and compare the length of the confidence intervals. Second, calculate and plot nonparametric estimates of the covariance/correlation structure and investigate what parametric assumption best fits the data. These two simple steps would provide a lot of additional information in general and in the Nursing Intervention Study in particular.

4 Where?

The author develops the Bayesian GMM methodology for standard GLMMs with a large number of clusters, where the GMM methodology was successfully applied. Thus, one would expect the current methodology to also do well on the beaten path. A challenging alternative problem is estimating a smooth, but otherwise unspecified, population mean function $f(\mathbf{Z})$ instead of $\boldsymbol{\beta}^T\mathbf{Z}$. Another set of problems where the Bayesian GMM methodology could have an impact are generated by emerging literature on multilevel functional data analysis (Crainiceanu et al. (2009); Di et al. (2009)), where random effects are replaced by random processes.

References

Crainiceanu, C., Staicu, A., and Di, C. (2009). "Generalized Multilevel Functional Regression." *Journal of the American Statistical Association, under review*. 215

Di, C., Crainiceanu, C., Caffo, B., and Naresh, P. (2009). "Multilevel Functional Principal Component Analysis." *Annals of Applied Statistics, online access 2008*, 3(1): 458–488. 215

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators." *Econometrica*, 50: 1029–1054. 213

Liang, K.-Y. and Zeger, S. L. (1986). "Longitudinal data analysis using generalized linear models." *Biometrika*, 73: 13–22. 214

Newey, W. K. and West, K. D. (1987). "Hypothesis testing with efficient method of moments estimation." *International Economic Review*, 28: 777–787. 213

Qu, A., Lindsay, B. G., and Li, B. (2000). "Improving generalised estimating equations using quadratic inference functions." *Biometrika*, 87: 823–836. 214

Yin, G. (2009). "Bayesian Generalized Method of Moments." *Bayesian Analysis*, 4(1): 1–17. 213

Acknowledgments

Crainiceanu's research was supported by Award Number R01NS060910 from the National Institute Of Neurological Disorders And Stroke. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute Of Neurological Disorders And Stroke or the National Institutes of Health.