GENERALISED LINEAR MIXED MODELS OF INTEREST AT UNIVERSITY OF WOLLONGONG STATISTICS GROUP

Spring Session, 2008

DATE OF THIS DRAFT: 01 OCT 2008 BY M.P. WAND

1 Introduction

On Wednesday 24th September, 2008, the Statistics group at University Wollongong held a forum on the theme:

Why GLMM is relevant to you.

The speakers were: Ray Chambers, Robert Clark, Ken Russell, David Steel and Matt Wand. Ken spoke on behalf of Damian Collins, who recently submitted a PhD thesis on GLMM research. This document aims to summarise what was established during that forum.

To get started we give, in the next section, an excerpt from:

Zhao, Y., Staudenmayer, J., Coull, B.A. and Wand, M.P. (2006). General design Bayesian generalized linear mixed models. *Statistical Science*, **21**, 35–51.

in which a very general GLMM set-up is laid out. Following that we identify which subsets of this set-up corresponding to the forum presentations and other aspects that were raised.

2 A Very General GLMM Set-Up

GLMMs for canonical one-parameter exponential families (e.g. Poisson, logistic) and Gaussian random effects take the general form

$$[\mathbf{y}|\boldsymbol{\beta},\mathbf{u}] = \exp\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}) + \mathbf{1}^T c(\mathbf{y})\}, \qquad (1)$$

$$[\mathbf{u}|\mathbf{G}] \sim N(\mathbf{0},\mathbf{G}) \qquad (2)$$

where here, and throughout, the distribution of a random vector $\mathbf{x}$ is denoted by $[\mathbf{x}]$ and the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ is denoted by $[\mathbf{y}|\mathbf{x}]$.

In the Poisson case $b(x) = e^x$, while in the logistic case $b(x) = \log(1+e^x)$. A few other models (e.g. gamma, inverse Gaussian) also fit into this structure (McCullagh and Nelder, 1989). A number of extensions and modifications are possible. One is to allow for overdispersion, especially in the Poisson case. In this paper we will restrict attention to the canonical one-parameter exponential family structure.

It is important to separate out random effects structure for handling grouping. One reason is that it allows for the possibility of hierarchical centering in the MCMC implementations It also recognizes the different covariance structures used in longitudinal data modeling, smoothing and spatial statistics. Such considerations suggest the breakdown

$$\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u} = \mathbf{X}^{\mathrm{R}}\boldsymbol{\beta}^{\mathrm{R}} + \mathbf{Z}^{\mathrm{R}}\mathbf{u}^{\mathrm{R}} + \mathbf{X}^{\mathrm{G}}\boldsymbol{\beta}^{\mathrm{G}} + \mathbf{Z}^{\mathrm{G}}\mathbf{u}^{\mathrm{G}} + \mathbf{Z}^{\mathrm{C}}\mathbf{u}^{\mathrm{C}}, \qquad (3)$$

where

$$\mathbf{X}^{\text{R}} \equiv \begin{bmatrix} \mathbf{X}_1^{\text{R}} \\ \vdots \\ \mathbf{X}_m^{\text{R}} \end{bmatrix}, \qquad \mathbf{Z}^{\text{R}} \equiv \underset{1\le i\le m}{\text{blockdiag}}(\mathbf{X}_i^{\text{R}})$$

and

$$\text{Cov}(\mathbf{u}^{\text{R}}) \equiv \underset{1\le i\le m}{\text{blockdiag}}(\mathbf{\Sigma}^{\text{R}}) \equiv \mathbf{I}_m \otimes \mathbf{\Sigma}^{\text{R}}$$

correspond to random intercepts and slopes, as typically used for repeated measures data on m groups with sample sizes $n_1, \ldots, n_m$. Here $\mathbf{X}_i^{\text{R}}$ is a $n_i \times q^{\text{R}}$ matrix for the random design corresponding to the ith group, $\mathbf{\Sigma}^{\text{R}}$ is an unstructured $q^{\text{R}} \times q^{\text{R}}$ covariance matrix and $\otimes$ denotes Kronecker product.

Next, $\mathbf{X}^{\text{G}}$ and $\mathbf{Z}^{\text{G}}$ are general design matrices, usually of different form than those arising in random effects models. In many of our examples $\mathbf{X}^{\text{G}}$ contains indicator variables, or polynomial basis functions of a continuous predictor, while $\mathbf{Z}^{\text{G}}$ contains spline basis functions (e.g. Brumback *et al.*, 1999). The $\mathbf{Z}^{\text{G}}\mathbf{u}^{\text{G}}$ term may be further decomposed as

$$\mathbf{Z}^{\text{G}}\mathbf{u}^{\text{G}} = \sum_{\ell=1}^{L} \mathbf{Z}_\ell^{\text{G}}\mathbf{u}_\ell^{\text{G}}$$

with each $\mathbf{Z}_\ell^{\text{G}}$, $1 \le \ell \le L$, usually corresponding to a smooth term in an additive model. Also, in keeping with spline penalization, we only consider

$$\text{Cov}(\mathbf{u}^{\text{G}}) = \underset{1\le \ell\le L}{\text{blockdiag}}(\sigma_{u\ell}^2 \mathbf{I}).$$

Note that the decomposition (3) is not unique for a particular model. For instance, in the crossed random effects model given in the following *Example 3*, we present two ways of decomposition.

The $\mathbf{Z}^{\text{C}}\mathbf{u}^{\text{C}}$ component represents random effects with spatial correlation structure. This can be done in a number of ways (e.g. Wakefield, Best and Waller, 2000); we will just describe one of the more common approaches here. Suppose disease incidence data are available over N contiguous regions. The random effect $\mathbf{u}^{\text{C}}$ vector is of dimension N with entries $U_1^C, \ldots, U_N^C$. The conditional distribution of U_i^C given U_j^C, $j \neq i$, is a univariate normal distribution with mean equal to the average U_j^C values of U_i^C's neighboring regions, and variance equal to σ_c^2 divided by the number of neighboring regions. This is known as the intrinsic Gaussian autoregression distribution (Besag, York and Mollié, 1991). This leads to $\mathbf{u}^{\text{C}}$ having an improper density proportional to

$$\exp\left\{ -\sum_{i\sim j} \tfrac{1}{2}\sigma_c^{-2}(U_i^C - U_j^C)^2 \right\} \tag{4}$$

where $i \sim j$ denotes spatially adjacent regions.

The versatility of (3) can be appreciated by considering the following set of examples. Note that we use truncated linear basis functions for smoothing components to keep the formulations simple. In practice these may be replaced by B-splines or radial basis functions. Knots are denoted by κ_k with possible superscripting. In the examples we use $\mathbf{1}_d$ to denote a $d \times 1$ vector of ones.

Example 1: Random intercept

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_{ij} = \beta_0 + U_i + \beta_1 x_{ij}, \quad 1 \le j \le n_i,\ 1 \le i \le m,$$

$$\mathbf{X}_i^{\mathrm{R}} = \mathbf{1}_{n_i}, \qquad \mathbf{X}^{\mathrm{G}} = [x_{ij}], \qquad \mathbf{Z}^{\mathrm{G}} = \mathbf{Z}^{\mathrm{C}} = \emptyset, \qquad \boldsymbol{\Sigma}^{\mathrm{R}} = \sigma_u^2.$$

Example 2: Random intercept and slope

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_{ij} = \beta_0 + U_i + (\beta_1 + V_i)x_{ij}, \quad 1 \le j \le n_i,\ 1 \le i \le m,$$

$$\mathbf{X}_i^{\mathrm{R}} = \begin{bmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix}, \qquad \mathbf{X}^{\mathrm{G}} = \mathbf{Z}^{\mathrm{G}} = \mathbf{Z}^{\mathrm{C}} = \emptyset,$$

$$\boldsymbol{\Sigma}^{\mathrm{R}} = \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}.$$

Example 3: Crossed random effects model

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_{ii'} = \beta_0 + U_i + U'_{i'}, \quad 1 \le i \le n,\ 1 \le i' \le n',$$

$$\mathbf{X}^{\mathrm{G}} = \mathbf{1}_{nn'}, \qquad \mathbf{Z}^{\mathrm{G}} = [\mathbf{I}_n \otimes \mathbf{1}_{n'} \mid \mathbf{1}_n \otimes \mathbf{I}_{n'}], \qquad \mathbf{X}^{\mathrm{R}} = \mathbf{Z}^{\mathrm{R}} = \mathbf{Z}^{\mathrm{C}} = \emptyset,$$

$$\mathbf{u}^{\mathrm{G}} = [U_1, \ldots, U_n, U'_1, \ldots, U'_{n'}]^T, \qquad \mathrm{Cov}(\mathbf{u}^{\mathrm{G}}) = \mathrm{blockdiag}(\sigma_u^2\mathbf{I}_n, \sigma_{u'}^2\mathbf{I}_{n'}).$$

An alternative representation of this model is

$$\mathbf{X}_i^{\mathrm{R}} = \mathbf{1}_{n'\times 1}, \qquad \mathbf{Z}^{\mathrm{G}} = [\mathbf{1}_n \otimes \mathbf{I}_{n'}], \qquad \mathbf{X}^{\mathrm{G}} = \mathbf{Z}^{\mathrm{C}} = \emptyset,$$

$$\boldsymbol{\Sigma}^{\mathrm{R}} = \sigma_u^2, \qquad \mathrm{Cov}(\mathbf{u}^{\mathrm{G}}) = \sigma_{u'}^2\mathbf{I}_{n'}.$$

This allows for implementation of hierarchical centering if MCMC is used for fitting.

Example 4: Nested random effects model

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_{ijk} = \beta_0 + U_i + V_{j(i)} + \beta_1\, x_{ijk}, \quad 1 \le i \le m,\ 1 \le j \le n,\ 1 \le k \le p,$$

$$\mathbf{X}^{\mathrm{G}} = [1\ x_{ijk}]_{1\le i\le m, 1\le j\le n, 1\le k\le p}, \quad \mathbf{Z}^{\mathrm{G}} = [\mathbf{I}_m \otimes \mathbf{1}_{np} \mid \mathbf{I}_m \otimes (\mathbf{I}_n \otimes \mathbf{1}_p)], \quad \mathbf{X}^{\mathrm{R}} = \mathbf{Z}^{\mathrm{R}} = \mathbf{Z}^{\mathrm{C}} = \emptyset,$$

$$\mathbf{u}^{\mathrm{G}} = [U_1, \ldots, U_m, V_{1(1)}, \ldots, V_{n(1)}, \ldots, V_{1(m)}, \ldots, V_{n(m)}]^T,$$

$$\mathrm{Cov}(\mathbf{u}^{\mathrm{G}}) = \mathrm{blockdiag}(\sigma_u^2\mathbf{I}_m, \sigma_v^2\mathbf{I}_{np}).$$

Example 5: Generalized scatterplot smoothing

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_i=\beta_0+\beta_1 x_i+\sum_{k=1}^{K}u_k(x_i-\kappa_k)_+,$$

$$\mathbf{X}^{\text{G}}=[1\ x_i]_{1\le i\le n},\qquad \mathbf{Z}^{\text{G}}=[\underset{1\le k\le K}{(x_i-\kappa_k)_+}]_{1\le i\le n},\qquad \mathbf{X}^{\text{R}}=\mathbf{Z}^{\text{R}}=\mathbf{Z}^{\text{C}}=\emptyset,$$

$$\text{Cov}(\mathbf{u}^{\text{G}})=\sigma_u^2\mathbf{I}_K.$$

Example 6: Generalized additive model

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_i=\beta_0+\beta_s s_i+\sum_{k=1}^{K^s}u_k^s(s_i-\kappa_k^s)_++\beta_t t_i+\sum_{k=1}^{K^t}u_k^t(t_i-\kappa_k^t)_+,$$

$$\mathbf{X}^{\text{G}}=[1\ s_i\ t_i]_{1\le i\le n},\qquad \mathbf{Z}^{\text{G}}=[\underset{1\le k\le K^s}{(s_i-\kappa_k^s)_+}\quad \underset{1\le k\le K^t}{(t_i-\kappa_k^t)_+}]_{1\le i\le n},\qquad \mathbf{X}^{\text{R}}=\mathbf{Z}^{\text{R}}=\mathbf{Z}^{\text{C}}=\emptyset,$$

$$\text{Cov}(\mathbf{u}^{\text{G}})=\text{blockdiag}(\sigma_{us}^2\mathbf{I}_{K^s},\ \sigma_{ut}^2\mathbf{I}_{K^t}).$$

Example 7: Generalized additive semiparametric mixed model

$$\begin{aligned}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_{ij} &= \beta_0+U_i+(\beta_q+V_i)q_{ij}+(\beta_r+W_i)r_{ij}+\beta_1 x_{ij}\\ &\quad +\beta_s s_{ij}+\sum_{k=1}^{K^s}u_k^s(s_{ij}-\kappa_k^s)_++\beta_t t_{ij}+\sum_{k=1}^{K^t}u_k^t(t_{ij}-\kappa_k^t)_+,\end{aligned}$$

$$\mathbf{X}_i^{\text{R}}=\begin{bmatrix}1 & q_{i1} & r_{i1}\\ \vdots & \vdots & \\ 1 & q_{in_i} & r_{in_i}\end{bmatrix},\qquad \mathbf{X}^{\text{G}}=[s_{ij}\ t_{ij}\ x_{ij}]_{1\le j\le n_i,1\le i\le m},$$

$$\mathbf{Z}^{\text{G}}=[\underset{1\le k\le K^s}{(s_{ij}-\kappa_k^s)_+}\quad \underset{1\le k\le K^t}{(t_{ij}-\kappa_k^t)_+}],\qquad \mathbf{Z}^{\text{C}}=\emptyset,$$

$$\boldsymbol{\Sigma}^{\text{R}}=\text{unstructured }3\times 3\text{ covariance matrix},\ \text{Cov}(\mathbf{u}^{\text{G}})=\text{blockdiag}(\sigma_{us}^2\mathbf{I}_{K^s},\ \sigma_{ut}^2\mathbf{I}_{K^t}).$$

Example 8: Generalized bivariate smoothing/low-rank kriging

$$(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u})_i=\beta_0+\boldsymbol{\beta}_1^T\mathbf{x}_i+\sum_{k=1}^{K}u_kC(\|\mathbf{x}_i-\boldsymbol{\kappa}_k\|),$$

$$\mathbf{X}^{\mathrm{G}}=[1\ \mathbf{x}_i^T]_{1\le i\le n}\qquad \mathbf{Z}^{\mathrm{G}}=[C(\|\mathbf{x}_i-\boldsymbol{\kappa}_k\|)]_{\substack{1\le i\le n,\\ 1\le k\le K}}\qquad \mathbf{X}^{\mathrm{R}}=\mathbf{Z}^{\mathrm{R}}=\mathbf{Z}^{\mathrm{C}}=\emptyset,$$

$$\mathrm{Cov}(\mathbf{u}^{\mathrm{G}})=\sigma_u^2\mathbf{I}.$$

Here $\|\mathbf{v}\|\equiv\sqrt{\mathbf{v}^T\mathbf{v}}$, and $C(r)=r^2\log|r|$ corresponding to low-rank thin plate splines with smoothness parameter set to 2; $C(r)=\exp(-|r/\rho|)(1+|r/\rho|)$ corresponding to Matérn low-rank kriging with range of $\rho>0$ and smoothness parameter set to $3/2$ (Kammann and Wand, 2003). Several more examples could be added, including some where $\mathbf{Z}^{\mathrm{C}}\neq\emptyset$.

3 Wollongong Statistics Group GLMM Research

3.1 Steel

Steel's interest in GLMM comes from research on

- interview-based surveys,
- split-questionnaire designs,
- aggregation of survey data
- a medical study with acronym DYOPTA.

Steel said that most of his GLMMs use random effects to account for geographical and time effects. An example is in his aggregation research where smoking rates are available only at the post code level and a random effect for post code is needed.

In his interview-based survey research Steel uses random effects to account for interview effects. There is also the issue of confounding between interviewer and suburb that requires some delicate GLMM modelling.

3.2 Clark

Clark's GLMM research is driven by problems arising in the analysis of complex survey data. A typical example of a a response variable is

$$y_{ij}=\text{voting intention of } j\text{th household in suburb } i.$$

The simplest GLMM considered by Clark for data such as these would put in a random effect for suburb. This corresponds to Example 1 in Section 2. One of his interests is testing

$$H_0:\sigma_U^2=0$$

where σ_U^2 is the variance component attached to the suburb random effect.

He also mentioned that there are some circumstances where random intercept and slope models (Example 2 in Section 2) are useful, and that there is interest in testing this model versus random intercept models.

Clark's work also involves higher order multi-level structure. An example is

$$\text{measurement} \subset \text{person} \subset \text{household} \subset \text{area}.$$

3.3 Chambers

Chambers' interest in GLMM goes back at least to his time in the UK where his work on social surveys (e.g. with responses such as indicator for employed) and he had to deal with local authority districts. Like Clark and Steel he usually uses random effects for geographical regions. One of the models presented by Chambers had a combination of random effects and indicator variables – so probably corresponds to a random intercept and slope model (Example 2 in Section 2).

Chambers also talked about multinomial logistic versions of GLMM arising in his research.

These days much of his GLMM research comes from his work with members of the Australian Bureau of Statistics.

3.4 Collins/Russell

Russell supervised a recently submitted PhD thesis by Collins on GLMM. Collins works for the Department of Primary Industries. Russell briefly described the types of GLMMs that arise Collins' workplace.

One use of random effects is to account for areal differences in agricultural experiments – e.g. a random paddock effect.

Some of the Collins/Russell GLMM research involves the classical animal breeding situation where sire effects need to be handled.

Some more modern uses of GLMMs involve genetics in grain crops.

In one stream of work the response is ordinal – e.g. on a scale of 1–5.

Equation 1.1 of Collins thesis and the $\mathbf{Z}$ matrix on page 14 indicates that the random effects structure of interest to Collins/Russell is limited to $\mathbf{X}^{\mathrm{R}}\boldsymbol{\beta}^{\mathrm{R}} + \mathbf{Z}^{\mathrm{R}}\mathbf{u}^{\mathrm{R}}$ structure of Section 2, although this hasn't yet been formally established.

3.5 Wand/Ormerod

Much of Wand's GLMM research is encapsulated in Section 2 where penalised splines are used in combination with random effects to allow for smooth (mainly univariate, but sometimes bivariate) functional effects. These GLMMs are such that intractable integrals have dimension corresponding to the number of spline basis functions. For univariate smoothing this is about 20–30; but for bivariate smoothing it could be 100–150.

A more recent use of GLMM is for building parsimonious and interpretable classifiers in situations where there are several (dozens of) candidate predictors. This research was done jointly with John Ormerod. A feature of this use of GLMM.

Some more recent GLMM-type research involving Wand has dealt with non-exponential family distributions such as Students-t and generalised extreme value.

4 Summary of GLMM Issues for Wollongong Statistics Group

All in the group are interested in the most basic GLMM structure: the special case of equation (3) where only the $\mathbf{X}^{R}\boldsymbol{\beta}^{R} + \mathbf{Z}^{R}\mathbf{u}^{R}$ part is present, corresponds to Examples 1 and 2 in Section 2.

Some (esp. Clark and possibly Collins/Russell) are also interested in the nested random effects structure exemplified by Example 4 at the end of Section 2. I don't believe that this fits into the $\mathbf{X}^{R}\boldsymbol{\beta}^{R} + \mathbf{Z}^{R}\mathbf{u}^{R}$ framework and I see that Zhao *et al.*(2006) used the general design matrix $\mathbf{Z}^{G}$ to handle this type of GLMM. This part of our 2006 paper was figured out by co-author Brent Coull, so I'm a bit fuzzy on this. In particular, I don't know right now the sort of dimensions of integrals that one gets stuck with for models of these types.

None of the group who work in sample survey research (Chambers, Clark, Steel) use geographical models beyond random effects for geographical regions (e.g. suburbs, post codes). So models with the $\mathbf{Z}^{C}\mathbf{u}^{C}$ structure of the previous section (e.g. Wakefield, Best and Waller, 2000), that has become popular in spatial epidemiology since the early 1990s, are not currently used. Chambers said that he tried it at least once on one of his problems and it did not make much difference.

Therefore it seems that the GLMMs of current interest in the Wollongong statistics group have the following breakdown:

1. The special case of equation (3) where only the $\mathbf{X}^{R}\boldsymbol{\beta}^{R} + \mathbf{Z}^{R}\mathbf{u}^{R}$ part is present. This means that the dimensions of intractable integrals are quite low – usually dimension 1, but sometimes 2 or 3. All GLMM researchers in the group have an interest in this situation.

2. The nested random effects situation corresponding to Example 4. At present the dimensionality of the intractable integrals should be pinned down soon. This is relevant to the research of Clark and co-researcher and possibly Collins/Russell. The work Wand and Ormerod potentially has some overlap with this type of GLMM.

3. The general design GLMMs of Section 2 that allow splines and more general spatial models. Here much higher-dimensional intractable integrals arise. Currently Wand and Ormerod are the only members of the group working on this general type of GLMM.

In terms of the response distribution, everyone seems to have to deal with binary response situation. Some have problems involving ordinal response and multi-category response. Poisson response arises in some of the Wand/Ormerod work, as does some non-exponential family distributions.

References

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–20.

Kammann, E.E. & Wand, M.P. (2003). Geoadditive models. *Applied Statistics*, **52**, 1–18.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models (Second Edition)*. London: Chapman and Hall.

Wakefield, J.C., Best, N.G. and Waller, L. (2000). Bayesian approaches to disease mapping. In Spatial Epidemiology, eds. Elliott, P., Wakefield, J.C., Best, N.G. and Briggs, D.J. Oxford: Oxford University Press. 104–127.