

NISS

Random Effects in Generalized Linear Mixed Models

Dongchu Sun, Paul Speckman,
and Robert K. Tsutakawa

Technical Report Number 97
April, 1999

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Random Effects in Generalized Linear Mixed Models

Dongchu Sun, Paul L. Speckman and Robert K. Tsutakawa

Department of Statistics, University of Missouri, Columbia, MO 65211, USA

Abstract

In this chapter, we examine the use of special forms of correlated random effects in the generalized linear mixed model (GLMM) setting. A special feature of our GLMM is the inclusion of random residual effects to account for lack of fit due to extra variation, outliers and other unexplained sources of variation. For random effects, we consider, in particular, the correlation structure and improper priors associated with the autoregressive (AR) model of Ord (1975) and the conditional autoregressive (CAR) model of Besag (1974). We give conditions for the propriety of the posterior distribution of the GLMM when the fixed effects have a constant improper prior and the random effects have a possibly improper conditional autoregressive prior. Several examples of exponential families as well as computational details for Markov chain Monte Carlo simulation are also presented.

1 Introduction

Traditional treatment of random effects in mixed linear and nonlinear models generally assumes that these effects are independent following some standard distributions such as normal or gamma. However, with the advent of Markov chain Monte Carlo (MCMC) methods and, in particular, the Gibbs sampler (cf. Gelfand and Smith, 1990), such restrictions are no longer necessary, and a much broader class of models, including those with correlated random effects, can be used in practice. (See Clayton (1996) for a general review of this recent development.)

In this chapter we consider generalized mixed linear models with random effects having the autoregressive and conditionally autoregressive properties commonly encountered in temporal and spatial covariates where one expects similarities among closely situated observations. Examples from disease mapping will be used to motivate these models.

The computational simplicity of MCMC methods enables one to extend the commonly used generalized linear mixed model (GLMM) to one that appends random residual effects

to the linear term to account for lack of fit. These extra terms allow for the minor perturbations and occasional outliers commonly encountered in practice. However, the remarkable ease of application of the Gibbs sampler does not come without a price. There is potential nonconvergence and other annoying problems when using the algorithm, especially in situations where noninformative prior distributions are employed.

In Section 2 we formally define the GLMM with residual effects. Two examples are given. One has the normal distribution and the other the gamma distribution, with the choice depending on the nature of the observed data. For example when the data are Poisson, it is more natural to use the conjugate gamma distribution, although the normal may be just as appropriate and simple to use.

In Section 3 we discuss several forms of correlated random effects including the AR process of Ord (1974) and the CAR process of Besag (1974), which are useful in describing spatial correlations. We examine the joint distributions associated with these processes to get a better understanding of the underlying association implied by these models. Of particular interest are distributions that are improper and could create problems when used in the GLMMs.

In Section 4, we consider the incorporation of these spatial random variables into the GLMM setting and emphasize the special role of the link function in a Bayesian hierarchical framework. In the case where the residual effects are normally distributed, the fixed effects have a constant prior and random effects may have an improper prior, we give sufficient conditions for the existence of a proper posterior distribution of all parameters including the fixed and random effects and variance components.

In Section 5, we summarize the computational details including the full conditional distributions required for the implementation of the Gibbs Sampler.

2 The Model

Let Y_1, \dots, Y_N be the independent random observations, where Y_i has the probability density

$$f_i(y_i|\eta_i, \phi) = \exp[A_i(\phi)^{-1}\{y_i\eta_i - B_i(\eta_i)\} + C_i(y_i; \phi)]. \quad (1)$$

The function $A_i(\phi)$ is commonly of the form $A_i(\phi) = \phi w_i^{-1}$, where the w_i are prespecified weights. It is often assumed that the scale parameter ϕ is known. Consider, for example, the case when the population size in area i is m_i with unknown mortality rate p_i , and Y_i is Poisson distributed with mean $m_i p_i$. This is a special case of (1) with $\phi = 1$, $A_i(\phi) = 1$, $\eta_i = \log(m_i p_i)$, $B_i(\eta_i) = \exp(\eta_i)$, and $C_i(y_i; \phi) = -\log(y_i!)$. When Y_i has a binomial distribution with parameters m_i and p_i , $\phi = 1$, $A_i(\phi) = 1$, $\eta_i = \log\{p_i/(1 - p_i)\}$, $B_i(\eta_i) = m_i \log\{1 + \exp(\eta_i)\}$, and $C_i(y_i; \phi) = \log[m_i!/\{y_i!(m_i - y_i)!\}]$.

Generalized Linear Models. We wish to model the variability in η_i to account for various fixed covariates. The natural parameters η_i are modeled as

$$h_i(\eta_i) = \mathbf{x}_{1i}^t \boldsymbol{\theta}, \quad (2)$$

where the h_i are known monotone functions, $\mathbf{X}_1 = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n})^t$ is an $N \times p$ design matrix and $\boldsymbol{\theta}$ is the vector of fixed effects. Such a model is commonly referred to as a generalized linear model (GLM) with canonical parameter η_i , scale parameter ϕ , and link function h_i (cf. McCullagh and Nelder, 1989). (Note that usually there is a single link function $h_i \equiv h$.)

Generalized Linear Mixed Models. We now extend the model to include random effects as follows. Let

$$h_i(\eta_i) = \mathbf{x}_{1i}^t \boldsymbol{\theta} + \mathbf{x}_{2i}^t \mathbf{Z}, \quad (3)$$

where h_i is a known monotone function, $\mathbf{X}_1 = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n})^t$ and $\mathbf{X}_2 = (\mathbf{x}_{21}, \dots, \mathbf{x}_{2n})^t$ are $N \times p$ and $N \times k$ design matrices, the $p \times 1$ vector $\boldsymbol{\theta}$ represents fixed effects, and \mathbf{Z} is a $k \times 1$ vector of random effects. Models given by (1) and (3) are often called generalized linear mixed models (GLMMs) and have been widely used in many problems such as disease mapping e.g., Breslow and Clayton (1993).

We can further extend the model to add additional residual effects by taking

$$h_i(\eta_i) = \mathbf{x}_{1i}^t \boldsymbol{\theta} + \mathbf{x}_{2i}^t \mathbf{Z} + e_i. \quad (4)$$

Here $\mathbf{e} = (e_1, \dots, e_N)^t$ are residual effects satisfying some restriction such as $\mathbb{E}(e_i) = 0$ or $\mathbb{E} \exp(e_i) = 1$. In addition, \mathbf{Z} and \mathbf{e} are assumed mutually independent. We include random residual effects e_i to account for the lack of fit of (3) due to extra variation, outliers, and other unexplained sources of variation. Note that the random effect e_i is quite

different from \mathbf{Z} in the sense that \mathbf{Z} often accounts for some special pattern such as random geographical effects and spatial correlation. In addition, the number of components of \mathbf{Z} is often much smaller than N , the number of residual effects e_i . By a suitable choice of the design matrix, (4) may be encompassed under (3), but we do not do this in order to emphasize the separate roles of \mathbf{Z} and \mathbf{e} . We will call the model given by (1) and (4) a GLMM as well.

There are many possible choices for the link functions h_i in models (2)–(4). For example, in the mortality setting cited earlier, Y_i has the Poisson distribution with mean $m_i p_i$ and $\eta_i = \log(m_i p_i)$. One possibility is to take $h_i(\eta_i) = \eta_i - \log(m_i) = \log(p_i)$, and a loglinear regression model may be applied. Alternatively, Y_i can be modeled with a binomial distribution. Then the logit link is canonical, and $\text{logit}(p_i) = \log\{p_i/(1 - p_i)\} = \eta_i - \log(m_i - e^{\eta_i}) = h_i(\eta_i)$, resulting in logistic regression.

The random effects term \mathbf{Z} in (3)–(4) is typically assumed to have a multivariate normal distribution. We will discuss in detail the choice of the distribution of \mathbf{Z} in the next section.

Distribution of Residual Effects. We will assume that the residual effects e_i or some monotone functions of e_i have distributions belonging to an exponential family (1), with known common canonical parameter η but unknown scale parameter ϕ . For illustration, we will consider the following two classes of distributions for residual effects.

- *Normal Residual Effects.* Residual effects e_i are independent and identically normal with mean 0 and variance δ_0 .
- *Gamma Residual Effects.* The $\exp(e_i)$ are iid gamma(R, R). Here a random variable W has the gamma(α, β) distribution if W has p.d.f.

$$f(w) = \alpha^\beta \{\Gamma(\alpha)\}^{-1} w^{\alpha-1} \exp(-\beta w).$$

Special cases of these models have appeared previously. Clayton and Kaldor (1987) and Waller *et al.* (1997) use a Poisson-normal model (Poisson for Y_i and normal for \mathbf{Z}) but without the residual term \mathbf{e} . This is a special case of (3). Ghosh *et al.* (1998) use \mathbf{e} in the binomial-normal model and treat spatial effects by taking $\mathbf{X}_2 \mathbf{Z} = \mathbf{U}$, with \mathbf{U} having a distribution defined by the conditional auto regressive CAR(1) model of Besag (1974). This is a special case of (4). In Sun, Tsutakawa, Kim and He (1997) and Sun, Tsutakawa and

He (1998), \mathbf{Z} consists of block-wise independent random effects, where each block contains random effects and the e_i are independent random variables with mean 0 and a common variance. West and Aguilar (1997) give another interesting example analysing hospital quality monitors with an extra residual term in (1.4).

Special cases of Poisson-gamma models are found in Clayton and Kaldor (1987) and Tsutakawa (1988). Specifically, in Tsutakawa (1988), \mathbf{Z} contains independent random effects, and the $\exp(e_i)$ are independent gamma variables with mean 1 and a common variance.

The general model (1) and (2) can be used for both continuous and discrete data. A discrete example of (4), which motivated much of this work, is studied in Sun, Tsutakawa, Kim and He (1998), where a spatio-temporal model for cancer mortality data is proposed. For a given gender, let Y_{ijk} denote the frequency of deaths from some specific cause in the i th region and j th age group during the k th time period, $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$. Conditionally on the fixed and random parameters, assume the Y_{ijk} are independent and Poisson with means $m_{ijk}p_{ijk}$, where m_{ijk} is the size of the ijk th target population. The model of Sun *et al.* takes the form

$$\log(p_{ijk}) = \theta_j + Z_i + (\mu_j + W_{ij})(t_k - \bar{t}) + e_{ijk},$$

where θ_j is the effect of the j th age group, Z_i is the effect of the i th region, t_k is the midpoint of the k th time period, and $\bar{t} = \sum_{k=1}^K t_k / K$. The rate of change over time is represented by $(\mu_j + W_{ij})$ for the j th age group in the i th region. Both θ_j and μ_j are treated as fixed effects, while Z_i and W_{ij} are random. The residual effects e_{ijk} are also random. A detailed description of the distributions of the random effects and prior distributions are given in Sun *et al.* (1998), where disease mapping and interpretation of numerical results for male lung cancer in the state of Missouri can be found.

3 Random effects

3.1 Independent Random Effects

Historically, it was common to assume independent random effects for linear mixed models, i.e., Z_1, \dots, Z_N are independently and identically $N(0, \delta_1)$ distributed. (See Harville

(1977).) Typical examples include one-way ANOVA and two-way ANOVA models with random effects. Hobert and Casella (1996) gave necessary and sufficient conditions for the propriety of the posterior distribution for a class of noninformative priors for variances components assuming independence of random effects.

3.2 Correlated random effects

There are many important situations where the random effects should be modeled as correlated. Correlated models are especially appropriate for spatial effects. A number of related methods are commonly used.

Direct specification of correlation matrix. If the random effects are linearly ordered, as for example with longitudinal data, it may be convenient to specify a correlation structure directly. For example, to model correlation decreasing with distance, $\mathbf{Z} = (Z_1, \dots, Z_k)^t$ can be taken to have the $\text{MVN}(\mathbf{0}, \mathbf{\Sigma})$ distribution, where $\mathbf{\Sigma} = (\sigma_{ij})$ is the $k \times k$ matrix with elements

$$\sigma_{ij} = \tau \rho^{|i-j|}, \quad (5)$$

and $\tau > 0$ and $\rho \in (-1, 1)$ are constants. For MCMC methods with modest size k (say, $k < 100$ or so), it is sometimes feasible to generate \mathbf{Z} from the joint conditional distribution directly. A number of authors including Cressie and Chan (1989) have used the distance between area i and area j to introduce spatial correlation.

AR models. Again assuming a linear ordering for the components of \mathbf{Z} , a commonly used structure is the AR(1) model with

$$Z_i = \rho Z_{i-1} + \epsilon_i, \quad i = 2, \dots, k, \quad (6)$$

where ρ is a constant in $(-1, 1)$, and the ϵ_i are independent and identically $N(0, \delta_1)$ distributed. If $Z_1 \sim N(0, \delta_1/(1-\rho^2))$, the distribution of \mathbf{Z} is given by (5) with $\tau = \delta_1/(1-\rho^2)$.

Ord (1975) proposed a generalized AR(1) model by defining

$$Z_i = \rho \sum_{j=1}^k C_{ij} Z_j + \epsilon_i, \quad (7)$$

where the C_{ij} are fixed constants satisfying $C_{ii} = 0$, and $\epsilon_1, \dots, \epsilon_k$ are iid $N(0, \delta_1)$. Here ρ is a ‘‘correlation coefficient,’’ measuring the correlation among Z_i in the sense that the

larger $|\rho|$ is, the stronger the correlation among the components of \mathbf{Z} . For example, if the Z_i are linearly ordered, one can define their joint distribution by assuming

$$\begin{aligned} Z_1 &= \rho Z_2 + \epsilon_1, \\ Z_i &= \rho(Z_{i-1} + Z_{i+1}) + \epsilon_i, \quad i = 2, \dots, k-1, \\ Z_k &= \rho Z_{k-1} + \epsilon_k. \end{aligned} \tag{8}$$

One advantage of (7) is that the formulation generalizes easily to two or more dimensions. Taking $\mathbf{C} = (C_{il})$ to be the $k \times k$ matrix of coefficients, \mathbf{I} the $k \times k$ identity matrix, and

$$\mathbf{W}_\rho = \mathbf{I} - \rho \mathbf{C}, \tag{9}$$

model (7) is equivalent to $\mathbf{W}_\rho \mathbf{Z} = (\epsilon_1, \dots, \epsilon_k)^t$. If \mathbf{W}_ρ is nonsingular, \mathbf{Z} has a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma} = \delta_1 (\mathbf{W}_\rho^t \mathbf{W}_\rho)^{-1}$. A common choice of \mathbf{C} is the adjacency matrix $\mathbf{A} = (a_{ij})_{k \times k}$, defined by

$$a_{ij} = \begin{cases} 1, & \text{if } j \text{ is adjacent to } i, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

The class of distributions for \mathbf{Z} when $\mathbf{W}_\rho = \mathbf{I} - \rho \mathbf{A}$ has been used in modeling random regional effects in disease mapping by Sun, Tsutakawa, Kim and He (1997) and random county effects in hunting success rates from a turkey hunting survey in the State of Missouri by He and Sun (1998).

One appealing way to view the prior for \mathbf{Z} is through the conditional distributions of Z_i given $\mathbf{Z}_{-i} = (Z_j, j \neq i)$. For the simple AR(1) prior (6), it can be shown that $\mathbf{B} = \delta_1 \boldsymbol{\Sigma}^{-1}$ is a tridiagonal matrix with diagonal elements $(1, 1 + \rho^2, \dots, 1 + \rho^2, 1)$ and off-diagonal elements $-\rho$. It follows easily that \mathbf{Z} has the Markov property

$$Z_i | \mathbf{Z}_{-i} \sim N\left(\frac{\rho}{1 + \rho^2}(Z_{i-1} + Z_{i+1}), \frac{\delta_1}{1 + \rho^2}\right), \text{ for } i = 2, \dots, k-1,$$

that is, the conditional distribution of Z_i given the rest depends only on adjacent variables. Curiously, the generalized AR prior specified through the adjacency matrix in (8) does not have a similar Markov property. This follows from the fact that $\boldsymbol{\Sigma}^{-1} = \delta_1^{-1} \mathbf{W}_\rho^t \mathbf{W}_\rho$ is a banded matrix but is not tridiagonal. Instead, $Z_i | \mathbf{Z}_{-i}$ depends on $(Z_{i-2}, Z_{i-1}, Z_{i+1}, Z_{i+2})$ for $3 \leq i \leq k-2$.

CAR(1) model. In an effort to use priors with the appealing first-order Markov property in spatial modeling, many authors have adopted conditional autoregressive or CAR models, which are developed by specifying the conditional distributions directly in a (presumably) consistent manner. One popular model takes

$$Z_i | \mathbf{Z}_{-i} \sim N\left(\frac{\rho}{d_i} \sum_{j \neq i}^k C_{ij} Z_j, \frac{\delta_1}{d_i}\right), \quad (11)$$

where C_{ij} and $d_i > 0$ are constants satisfying $C_{ii} = 0$. This is a special case of Besag's (1974) model with

$$f(Z_i | \mathbf{Z}_{-i}) = \left(\frac{\alpha_i}{2\pi\delta_1}\right)^{\frac{1}{2}} \exp\left\{-\frac{\alpha_i}{2\delta_1} \left(Z_i - \sum_{j \neq i}^k \beta_{ij} Z_j\right)^2\right\}, \quad (12)$$

$i = 1, \dots, k$. Suppose \mathbf{B} is the $k \times k$ matrix with diagonal elements α_i and ij th off-diagonal elements $-\alpha_i \beta_{ij}$. Besag proved that if \mathbf{B} is symmetric and positive definite, these conditional distributions lead to the joint probability density of \mathbf{Z} ,

$$f(\mathbf{Z}) = (2\pi\delta_1)^{-k/2} |\mathbf{B}|^{1/2} \exp\left(-\frac{1}{2\delta_1} \mathbf{Z}^t \mathbf{B} \mathbf{Z}\right), \quad (13)$$

i.e. $\mathbf{Z} \sim \text{MVN}(\mathbf{0}, \delta_1 \mathbf{B}^{-1})$. In the context considered here, suppose

$$\mathbf{B} = \mathbf{B}_\rho = \mathbf{D} - \rho \mathbf{C}, \quad (14)$$

where \mathbf{D} is a $k \times k$ diagonal matrix with positive elements (d_1, \dots, d_k) , and \mathbf{C} is a symmetric matrix with $C_{ii} = 0$. If \mathbf{B}_ρ is positive definite, then the joint distribution of \mathbf{Z} is (13), and the conditional distributions of Z_i given \mathbf{Z}_{-i} are (11).

In practice, these models are important because the simple conditional distributions depending only on neighboring values for the Z_i are desirable for Bayesian analysis using Markov chain Monte Carlo methods. Here are two important cases.

Case 1. Assume that $\mathbf{C} = \mathbf{A}$, the adjacency matrix, and $d_i = \sum_j C_{ij}$. If $\rho \in (-1, 1)$, then \mathbf{B} is positive definite and the conditional distribution of Z_i given \mathbf{Z}_{-i} is $N(\rho \bar{Z}_i, \delta_1/n_i)$, where n_i is the number of neighbors of location i , and \bar{Z}_i is the mean of the n_i neighboring Z_j s. (This corresponds to $\alpha_i = n_i$ and $\beta_{ij} = \rho/n_i$ if j is adjacent to i and zero otherwise.) This model was studied in Besag (1975) and Ripley (1981).

Case 2. Assume that $\mathbf{C} = \mathbf{A}$, the adjacency matrix, and $\mathbf{D} = \mathbf{I}$. Let λ_1 and λ_k be the smallest and largest eigenvalues of \mathbf{C} . If $\lambda_1^{-1} < \rho < \lambda_k^{-1}$, then \mathbf{B} is positive definite and

the conditional distribution of Z_i given \mathbf{Z}_{-i} is $N(\rho \sum_{j \neq i} a_{ij} Z_j, \delta_1)$. This model was used in Ripley (1988).

However, there are potential problems in modeling the dependence among the Z_i through the choice of (α_i, β_{ij}) in (12). One problem, singularity of \mathbf{B} , is addressed further in the next section. Another possible problem is with specifying a symmetric matrix \mathbf{B} . The specification in Case 1 with conditional variance δ_1/n_i seems unrealistic when ρ is small, since in the limiting case $\rho = 0$, the Z_i are independent but the variances still depend on the number of neighbors. This may not make sense near boundaries or in nonregular cases. On the other hand, in Case 2 the conditional variance of Z_i does not depend at all on the number of neighbors. As an alternative, suppose we let the conditional distribution of Z_i given \mathbf{Z}_{-i} be $N(\rho \bar{Z}_i, \delta_1(1 + \rho/n_i))$ and assume $0 \leq \rho \leq 1$. Formally, this is equivalent to a CAR model with $\alpha_i = (1 + \rho/n_i)^{-1}$ and

$$\beta_{ij} = \begin{cases} \rho/n_i, & \text{if } j \text{ is adjacent to } i, \\ 0, & \text{otherwise.} \end{cases}$$

Unfortunately, the ij th off-diagonal element of \mathbf{B} , $\alpha_i \beta_{ij} = \rho/(\rho + n_i)$, is not equal to $\alpha_j \beta_{ji}$ unless $n_i = n_j$. Even in the linearly ordered case, this fails at the boundary where $n_1 = 1$ and $n_j = 2$ for $1 < j < k$. Thus care must be taken in specifying a CAR model.

3.3 Strongly correlated random effects

If the determinant $|\mathbf{B}|$ is zero, the set of full conditional distributions given by (12) is not “compatible,” a definition used by Arnold and Press (1989), in the sense that there is no joint density of \mathbf{Z} consistent with the corresponding conditional densities. However, there are situations in practice where a nonpositive definite \mathbf{B} is desirable. For example, if $\rho \rightarrow 1$ in (14) when $\mathbf{C} = \mathbf{A}$, the adjacency matrix, and \mathbf{D} is the diagonal matrix of row sums of \mathbf{C} , the model is a Markov random field. Clearly \mathbf{B} is singular.

When the matrix \mathbf{B} is nonpositive definite, there are two possible interpretations. One way is to consider a lower dimensional distribution, in the sense that it is proper in certain directions but degenerate in some other directions. For example, let r be the rank of \mathbf{B} , and let $\lambda_1, \dots, \lambda_r$ be the positive eigenvalues of \mathbf{B} . Write $\mathbf{B} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^t$, where $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_k)$ is an orthogonal matrix, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$. Let $\boldsymbol{\gamma}_1 = (\gamma_{11}, \dots, \gamma_{r1})$ and

$\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_r)$. Then $\mathbf{B} = \mathbf{\Lambda}_1 \mathbf{U}_1^t$. Now let $\mathbf{U}_1 = (U_1, \dots, U_r)^t$ be a vector of independent random variables where $U_i \sim N(0, \delta_1 \lambda_i^{-1})$. Then $\mathbf{Z} = \mathbf{\Lambda}_1 \mathbf{U}_1$ has a singular normal distribution with mean $\mathbf{0}$ and covariance matrix $\delta_1 \mathbf{B}^-$, where \mathbf{B}^- is a pseudo-inverse of \mathbf{B} . We often write this distribution as $\text{MVN}(\mathbf{0}, \delta_1 \mathbf{B}^-)$. The joint distribution has the form

$$f(\mathbf{Z}) \propto (2\pi \delta_1)^{\frac{r}{2}} |\mathbf{B}|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2\delta_1} \mathbf{Z}^t \mathbf{B} \mathbf{Z}\right), \quad (15)$$

where $|\mathbf{B}|_+$ is defined to be $\prod_{i=1}^r \lambda_i$, the product of all positive eigenvalues of \mathbf{B} . Note that based on such a singular normal distribution, the full conditional distribution of Z_i given \mathbf{Z}_{-i} is degenerate instead of a normal distribution. The distribution of \mathbf{Z} is essentially proper on a lower dimensional space, so \mathbf{Z} is a vector of strongly correlated random effects.

Alternatively, we can sample an additional random sample $\mathbf{U}_2 = (U_{r+1}, \dots, U_k)^t$ from a flat constant density over a $k-r$ dimensional Euclidian space. Now define $\mathbf{Z} = \mathbf{\Lambda}_1 (\mathbf{U}_1^t, \mathbf{U}_2^t)^t$. We can see that the joint density of \mathbf{Z} has the form (15), which is improper because \mathbf{B} is singular. However, we can formally relate (15) to (12) by noting that

$$f(Z_i | \mathbf{Z}_{-i}) = f(Z_1, \dots, Z_k) / \int_{-\infty}^{\infty} f(Z_1, \dots, Z_k) dZ_i.$$

Hobert and Casella (1998) have called this type of relationship “functionally compatible,” in contrast to one being “compatible.”

Markov random field models. In (14), if $\mathbf{C} = \mathbf{A} = (a_{ij})$, $d_i = \sum_{j \neq i} a_{ij}$, and $\rho = 1$, then the distribution of \mathbf{Z} is often called a Markov random field model (cf. Kindermann and Snell (1980)). Such models have been used for modeling spatial correlations in disease mapping and other contexts by Besag, York, and Mollié (1991) and used by Bernardinelli and Montomoli (1992), Bernardinelli *et al.* (1995), Carlin and Louis (1996), Waller *et al.* (1997) and Ghosh *et al.* (1998) among others.

Autocorrelated random effects. We next give a class of strongly correlated distributions. Define the backwards difference operator \mathbf{H}_k to be the $k \times k$ matrix

$$\mathbf{H}_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdot & \cdots & -1 & 1 \end{pmatrix},$$

and let

$$\mathbf{G}_{kd} = [\mathbf{0}_{(k-d) \times d} \mid \mathbf{I}_{k-d}]$$

be the lower $(k - d)$ rows of the k -dimensional identity matrix \mathbf{I}_k . Now let the structural matrix \mathbf{B} in (15) be

$$\mathbf{B}_{kd} = (\mathbf{G}_{kd} \mathbf{H}_k^d)^t \mathbf{G}_{kd} \mathbf{H}_k^d. \quad (16)$$

Because \mathbf{B}_{kd} has rank $k-d$, \mathbf{Z} has a singular distribution, which we write as $\text{MVN}(\mathbf{0}, \delta_1 \mathbf{B}_{kd}^-)$, where \mathbf{B}_{kd}^- is a pseudo-inverse of \mathbf{B}_{kd} .

AR(1) (The first order difference model). When $d = 1$, the prior on \mathbf{Z} is called the first order difference or random walk prior. See Clayton (1996). In this case the structural matrix has the form

$$\mathbf{B}_{k1} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}_{k \times k}.$$

AR(2) (The second order difference) Model. When $d = 2$, the prior on \mathbf{Z} is called the stochastic-trend or second order difference prior. See Clayton (1996). In this case the structural matrix has the form

$$\mathbf{B}_{k2} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ -2 & 5 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 & 1 \end{pmatrix}_{k \times k}.$$

Note that this model can be obtained by the iterated formula

$$Z_i = 2Z_{i-1} - Z_{i-2} + \epsilon_i, \quad i = 3, \dots, k.$$

Such a second order random effects prior has been used for patient monitoring by Berzuini (1996).

3.4 Some Examples of the AR(d) model

To see the differences among the AR(d) models when we change d , in Figure 1 we graph three sample paths of the AR(d) process prior for \mathbf{Z} using $k = 100$, $\delta_1 = 1$ and $d = 1, 2, 3$. Note that the rank of \mathbf{B}_{kd} is $k - d$. When $d = 1$, the sample paths are simple random walks and locally rough, but sample paths are smoother when $d \geq 2$.

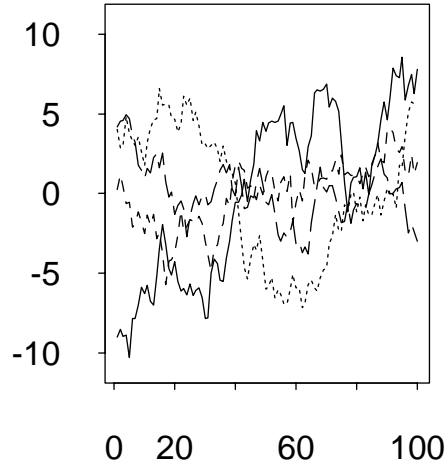
4 Hierarchical GLMMs

Bayesian analysis for the GLMs (1)–(2) and the GLMMs given by (1) and (3) is studied in Clayton (1996). We will discuss the GLMMs given by (1) and (4). For illustration, we will consider normal residual effects e_i . A full hierarchical Bayesian approach requires the specification of prior distributions for $\boldsymbol{\theta}$, the variance δ_0 of the distribution of e_i , the variance δ_1 of the distribution of random effects \mathbf{Z} , and the scale parameter ϕ .

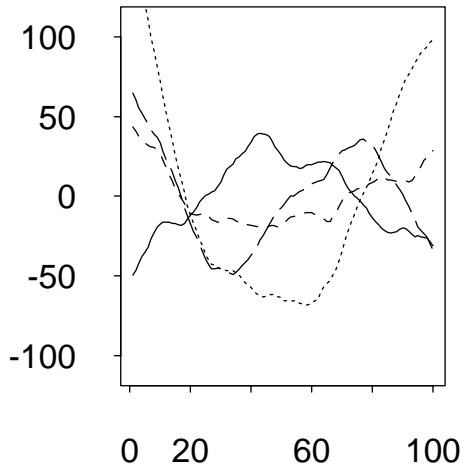
Although the commonly used prior of the fixed effects $\boldsymbol{\theta}$ is normal, we will assume a noninformative prior for $\boldsymbol{\theta}$, in particular, one having a constant density. We will not give a specific form for the priors of $(\delta_0, \delta_1, \phi)$. Since the prior for $\boldsymbol{\theta}$ is improper, and the prior for \mathbf{Z} is also improper for a singular \mathbf{B} , the joint posterior distribution may still be improper. As noted by Hobert and Casella (1996) and Sun, Tsutakawa and Speckman (1997), the propriety of the posterior is very important in Bayesian computation, especially when Markov chain Monte Carlo methods are used. Sun, Tsutakawa and Speckman (1997) considered a one-parameter distribution family where the prior for the parameters follows a linear mixed model and found conditions for a proper posterior distribution. Here we extend the results to the GLMM model, where the observations follow the densities (1) with canonical parameters η_i and a common scale parameter ϕ . We will only consider the case where there is a common variance component δ_1 for the whole vector of \mathbf{Z} . Some generalizations to block random effects can be found in Sun *et al.* (1998). We use the following notation. Note that for $B_i(\cdot)$ defined in (1), the first derivative B'_i is a strictly increasing function. Let H_i be the inverse function of B'_i . Note that for any fixed ϕ , the likelihood function $f_i(y_i|\eta_i, \phi)$ is bounded by

$$M_i(\phi) \equiv \sup_{\eta_i} f_i(y_i|\eta_i, \phi)$$

(a) $d=1$



(b) $d=2$



(c) $d=3$

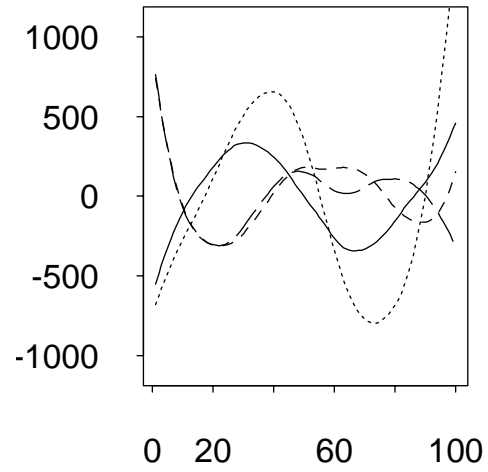


Figure 1: sample paths of CAR(d) models for Z when $n = 100$.

$$= \exp[A_i(\phi)^{-1}\{y_i H_i(y_i) - B_i(H_i(y_i))\} + C_i(y_i; \phi)]. \quad (17)$$

Theorem 4.1 Consider the GLMMs (1) and (4) with normal residual effects e_i *iid* $\sim N(0, \delta_0)$. Assume that

(a) there exists a subset of $\{1, \dots, N\}$, say $\mathcal{J}_n = (i_1, \dots, i_n)$, such that

$$\int \prod_{j \notin \mathcal{J}_n} M_j(\phi) \left\{ \prod_{j \in \mathcal{J}_n} \int f_j(y_j | \eta_j, \phi) h'_j(\eta_j) d\eta_j \right\} F(d\phi) < \infty, \quad (18)$$

where $F(\cdot)$ is the prior distribution for ϕ ;

(b) the design matrix $\mathbf{X}_1^* = (\mathbf{x}_{1,i_1}, \dots, \mathbf{x}_{1,i_n})^t$ has full rank p , and $\mathbf{X}_2^* = (\mathbf{x}_{2,i_1}, \dots, \mathbf{x}_{2,i_n})^t$ has the same rank as the matrix $\mathbf{X}_2 = (\mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,N})^t$;

(c) the prior for $\boldsymbol{\theta}$ is a constant and \mathbf{Z} follows the density (15);

(d) the rank of $(\mathbf{X}_2^{*t} \mathbf{R}_1 \mathbf{X}_2^* + \mathbf{B})$ is k , where $\mathbf{R}_1 = \mathbf{I}_n - \mathbf{X}_1^* (\mathbf{X}_1^{*t} \mathbf{X}_1^*)^{-1} \mathbf{X}_1^{*t}$;

(e) the prior for (δ_0, δ_1) satisfies the moment condition,

$$\mathbb{E}\{\delta_0^{-\frac{1}{2}(n-p-k)} \delta_1^{-\frac{1}{2}k} + \delta_0^{-\frac{1}{2}(n-p)}\} < \infty. \quad (19)$$

Then the posterior distribution of $(\boldsymbol{\eta}, \phi, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ given $\mathbf{Y} = (y_1, \dots, y_N)$ is proper.

Proof. Without loss of generality, assume that $\mathcal{J}_n = \{1, \dots, n\}$ in assumption (a) and that δ_i has a prior density g_i . The posterior density of $(\boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ given (\mathbf{Y}, ϕ) is

$$p(\boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1 | \mathbf{Y}, \phi) \propto \prod_{i=1}^N f_i(Y_i | \eta_i, \phi) h'_i(\eta_i) \delta_0^{-\frac{1}{2}(N-n)} \times \\ \prod_{i=n+1}^N \exp\left[-\frac{1}{2\delta_0}\{h_i(\eta_i) - x_{1i}^t \boldsymbol{\theta} - x_{2i}^t \mathbf{Z}\}^2\right] \prod_{j=0}^1 g_j(\delta_j) G.$$

where

$$G = \frac{1}{\delta_0^{\frac{n}{2}} \delta_1^{\frac{k}{2}}} \exp\left\{-\frac{(\mathbf{V}^* - \mathbf{X}_1^* \boldsymbol{\theta} - \mathbf{X}_2^* \mathbf{Z})^t (\mathbf{V}^* - \mathbf{X}_1^* \boldsymbol{\theta} - \mathbf{X}_2^* \mathbf{Z})}{2\delta_0} - \frac{\mathbf{Z}^t \mathbf{B} \mathbf{Z}}{2\delta_1}\right\}.$$

Here $\mathbf{V}^* = (h_1(\eta_1), \dots, h_n(\eta_n))^t$. Let $\boldsymbol{\eta}^* = (\eta_1, \dots, \eta_n)^t$. Using inequality (17) and integrating with respect to $(\eta_{n+1}, \dots, \eta_N)^t$,

$$p(\boldsymbol{\eta}^*, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1 | \mathbf{Y}, \phi) \propto \prod_{i=n+1}^N M_i(\phi) \prod_{j=1}^n f_j(y_j | \eta_j, \phi) h'_j(\eta_j) G.$$

Using arguments similar to those in Sun, Tsutakawa and Speckman (1997), we get

$$\int_{\mathbb{R}^p} \int_{\mathbb{R}^k} G d\boldsymbol{\theta} d\mathbf{Z} \leq \{\delta_0^{-\frac{1}{2}(n-p-k)} \delta_1^{-\frac{1}{2}k} + \delta_0^{-\frac{1}{2}(n-p)}\}.$$

Therefore, from assumption (e),

$$\begin{aligned} & \int \int p(\boldsymbol{\eta}^*, \phi | \mathbf{Y}) d\boldsymbol{\eta}^* F(d\phi) \\ & \propto \int \prod_{i=n+1}^N M_i(\phi) \left\{ \prod_{j=1}^n \int f_j(y_j | \eta_j, \phi) h'_j(\eta_j) d\eta_j \right\} F(d\phi), \end{aligned}$$

which is finite by (18). □

Remark 4.1 A common prior for the variance components δ_i is inverse gamma(a_i, b_i), whose density is

$$g_i(\delta_i) \propto \frac{1}{\delta_i^{a_i+1}} \exp(-b_i/\delta_i). \quad (20)$$

Clearly, when $b_i > 0$, $n - p - k + 2a_0 > 0$ and $k > 2a_1$, condition (19) holds.

Remark 4.2 When the prior of ϕ is degenerate, i.e., a known constant as in the Poisson or binomial cases, condition (18) becomes

$$\int f_j(y_j | \eta_j, \phi) h'_j(\eta_j) d\eta_j < \infty, \text{ for } j \in \mathcal{J}_n,$$

which is equivalent to the condition,

$$\int \exp[A_j(\phi)^{-1} \{y_j \eta_j - B_j(\eta_j)\}] h'_j(\eta_j) d\eta_j < \infty, \text{ for } j \in \mathcal{J}_n. \quad (21)$$

A condition similar to (21) was required for all j in Ghosh *et al.* (1997) for propriety of the posterior distribution.

Example 4.1 Suppose $f_i(y_i | \eta_i, \phi)$ is Poisson with mean $\mu_i = m_i p_i$. This is a special case of (1) with $\phi = 1$ and $\eta_i = \log(m_i p_i)$. Let $h_i(\eta_i) = \eta_i - \log(m_i) = \log(p_i)$. Here $\phi = 1$ and $h_i(\eta_i)$ follows the linear structure (4). Then $f_i(y_i | \eta_i, \phi)$ is bounded for any $y_i \geq 0$, and

$$\int f_i(y_i | \eta_i, \phi = 1) h'_i(\eta_i) d\eta_i = \int_0^\infty \frac{e^{-\mu_i} \mu_i^{y_i-1}}{y_i!} d\mu_i,$$

which is finite for $y_i > 0$. Under assumptions (b)-(e) of Theorem 4.1, the joint posterior distribution of $(p_1, \dots, p_N, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ is proper.

Example 4.2 Suppose $f_i(y_i|\eta_i, \phi)$ is binomial with parameters m_i and p_i . This is a special case of (1) with $\phi = 1$ and $\eta_i = \log\{p_i/(1 - p_i)\}$. Assume $h_i(\eta_i) = \eta_i$ has structure (4). Then $f_i(y_i|\eta_i, \phi = 1)$ is bounded in η_i for any $0 \leq y_i \leq m_i$, and

$$\begin{aligned} \int_{-\infty}^{\infty} f_i(y_i|\eta_i, \phi = 1)h'_i(\eta_i)d\eta_i &= \int_{-\infty}^{\infty} \frac{e^{y_i\eta_i}}{(e^{\eta_i} + 1)^{m_i}}d\eta_i \\ &= \int_0^1 p_i^{y_i-1}(1 - p_i)^{m_i-y_i-1}dp_i, \end{aligned}$$

which is finite if and only if $0 < y_i < m_i$. Under assumptions (b)–(e) of Theorem 4.1, the joint posterior distribution of $(p_1, \dots, p_N, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ is proper.

Example 4.3 When $Y_i|(\mu_i, \sigma^2) \sim N(\mu_i, \sigma^2)$, we have $\eta_i = \mu_i$, $\phi = \sigma^2$, $A_i(\phi) = \phi$, $B_i(\eta_i) = \eta_i$ and $C_i(y_i, \phi) = -0.5 \log(\phi) - y_i^2/(2\phi)$. If $h_i(\eta_i) = \eta_i$, this is a typical example of a normal hierarchical model. It is easy to see that $M_i(\phi) = 1/\sqrt{2\pi\phi}$ and $\int f_i(y_i|\eta_i, \phi)d\eta_i = 1$. Condition (18) becomes

$$\int_0^{\infty} \phi^{-\frac{1}{2}(N-n)}F(d\phi) < \infty,$$

which always holds when $N = n$ and F is a proper prior for ϕ . In addition, assumptions (b)–(e) of Theorem 4.1 hold. Then the joint posterior distribution of $(\mu_1, \dots, \mu_N, \sigma^2, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ is proper.

Example 4.4 Suppose $Y_i|(\mu_i, \alpha) \sim \text{gamma}(\alpha, \alpha/\mu_i)$, with density

$$f_i(y_i|\mu_i, \alpha) = \frac{\alpha^\alpha y_i^{\alpha-1}}{(\alpha)\mu_i^\alpha} \exp\{-\alpha y_i/\mu_i\}.$$

Here α is the common shape parameter and μ_i is the mean of Y_i for given (μ_i, α) . This is a special case of (1) with $\phi = \alpha$, $\eta_i = 1/\mu_i$, $A_i(\phi) = -1/\phi$, $B_i(\eta_i) = \log(\eta_i)$ and $C_i(y_i, \phi) = \alpha \log(\alpha) + (\alpha - 1) \log(y_i) - \log\{(\alpha)\}$. Choose $h_i(\eta_i) = \log(\eta_i) = -\log(\mu_i)$. Then

$$M_i(\phi) = y_i^{-1}\alpha^\alpha e^{-\alpha}/(\alpha) \text{ and } \int_0^{\infty} f(y_i|\eta_i, \phi)\frac{1}{\eta_i}d\eta_i = y_i^{-1}.$$

If $N = n$ and ϕ has a proper prior, condition (18) holds. If assumptions (b)–(e) in Theorem 4.1 hold, the joint posterior distribution of $(\mu_1, \dots, \mu_N, \alpha, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ is proper.

When ϕ is unknown but Y_i has a continuous distribution, as in the normal and gamma examples, we often choose $n = N$ and $\mathcal{J}_n = \{1, \dots, N\}$, so that condition (18) becomes

$$\int \left\{ \prod_{j=1}^N \int f_j(y_j | \eta_j, \phi) h'_j(\eta_j) d\eta_j \right\} F(d\phi) < \infty.$$

Remark 4.3 Assumption (d) in Theorem 4.1 is crucial. Otherwise, the results may not hold. On the other hand, it is easy to see that the rank of the matrix $(\mathbf{X}_2^t \mathbf{R}_1 \mathbf{X}_2 + \mathbf{B})$ equals k if either the rank of $(\mathbf{X}_1, \mathbf{X}_2)$ is $p + k$ or the rank of $(\mathbf{B}) = k$. The following results can be proved similarly.

Theorem 4.2 Assume that the rank of $(\mathbf{X}_2^t \mathbf{R}_1 \mathbf{X}_2 + \mathbf{B}) < k$, where $\mathbf{R}_1 = \mathbf{I}_N - \mathbf{X}_1(\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t$, and \mathbf{X}_1 and \mathbf{X}_2 are design matrices based on the full data. Under assumption (d), for any proper prior of $(\delta_0, \delta_1, \phi)$, the posterior is improper.

Proof. Let G be defined as in the proof of Theorem 4.1, and replace n by N and $(\mathbf{X}_1^*, \mathbf{X}_2^*)$ by $(\mathbf{X}_1, \mathbf{X}_2)$. From Sun, Tsutakawa and Speckman (1997), we know that for any given (δ_0, δ_1) ,

$$\int_{\mathbb{R}^p} \int_{\mathbb{R}^k} G d\boldsymbol{\theta} d\mathbf{Z} = \infty.$$

The results follows. □

The following result can be proved using the same argument as that of Theorem 4.1.

Theorem 4.3 Given assumptions (a), (b) and (c) of Theorem 4.1, suppose that either condition (d1) or (d2) below holds:

(d1) the rank of $(\mathbf{X}_2^{*t} \mathbf{R}_1 \mathbf{X}_2^*)$ is k and (19) is replaced by

$$\mathbb{E} \{ \delta_0^{-\frac{1}{2}(n-p-k)} \delta_1^{-\frac{1}{2}k} \} < \infty. \quad (22)$$

(d2) the rank of \mathbf{B} is k and (19) is replaced by

$$\mathbb{E} \{ \delta_0^{-\frac{1}{2}(n-p)} \} < \infty. \quad (23)$$

Then the result of Theorem 4.1 still holds.

5 Bayesian Computation

Bayesian inference for hierarchical GLMMs can be implemented via Markov chain Monte Carlo methods such as Gibbs sampling and/or the Metropolis algorithm. We assume that the prior for the variance components δ_i follows an inverse gamma(a_i, b_i) distribution with density (20). The proof of the following fact is omitted.

Fact 5.1 The full conditional distributions are as follows.

1. $\boldsymbol{\theta} | (\boldsymbol{\eta}, \phi, \mathbf{Z}, \delta_0, \delta_1) \sim \text{MVN}_p((\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t (\mathbf{V} - \mathbf{X}_2 \mathbf{Z}), \delta_0 (\mathbf{X}_1^t \mathbf{X}_1)^{-1})$.
2. $\mathbf{Z} | (\boldsymbol{\eta}, \phi, \boldsymbol{\theta}, \delta_0, \delta_1) \sim \text{MVN}_k(\mathbf{M}_1 \mathbf{X}_2^t (\mathbf{V} - \mathbf{X}_1 \boldsymbol{\theta}), \delta_0 \mathbf{M}_1)$, where $\mathbf{M}_1 = (\mathbf{X}_2^t \mathbf{X}_2 + \delta_0 \delta_1^{-1} \mathbf{B})^{-1}$.
3. $\delta_0 | (\boldsymbol{\eta}, \phi, \boldsymbol{\theta}, \mathbf{Z}, \delta_1) \sim \text{inverse gamma}(a_0 + \frac{n}{2}, b_0 + \frac{1}{2} (\mathbf{V} - \mathbf{X}_1 \boldsymbol{\theta} - \mathbf{X}_2 \mathbf{Z})^t (\mathbf{V} - \mathbf{X}_1 \boldsymbol{\theta} - \mathbf{X}_2 \mathbf{Z}))$.
4. $(\delta_1 | \boldsymbol{\eta}, \phi, \boldsymbol{\theta}, \mathbf{Z}, \delta_0) \sim \text{inverse gamma}(a_1 + \frac{k}{2}, b_1 + \frac{1}{2} \mathbf{Z}^t \mathbf{B} \mathbf{Z})$.
5. Given $(\phi, \mathbf{Z}, \delta_0, \delta_1)$, the η_j (or $v_j = h_j(\eta_j)$) are independent. In fact, since η_j and v_j are related by a one-to-one transformation, we can simulate from either η_j or v_j , depending on simplicity. The density of η_j given $(\phi, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ is

$$s_j(\eta_j) \propto \exp\left[\frac{y_j \eta_j - B_j(\eta_j)}{A_j(\phi)} - \frac{\{h_j(\eta_j) - \mathbf{x}_{1j}^t \boldsymbol{\theta} - \mathbf{x}_{2j}^t \mathbf{Z}\}^2}{2\delta_0}\right] h_j'(\eta_j),$$

and the density of v_j given $(\phi, \mathbf{Z}, \delta_0, \delta_1)$ is

$$\tilde{s}_j(v_j) \propto \exp\left[\frac{y_j h_j^{-1}(v_j) - B_j\{h_j^{-1}(v_j)\}}{A_j(\phi)} - \frac{\{v_j - \mathbf{x}_{1j}^t \boldsymbol{\theta} - \mathbf{x}_{2j}^t \mathbf{Z}\}^2}{2\delta_0}\right],$$

where h_j^{-1} is the inverse function of h_j .

6. If the prior for ϕ is degenerate, so is its posterior. If ϕ has the prior density $g(\phi)$, then its posterior density given $(\boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{Z}, \delta_0, \delta_1)$ is

$$g^*(\phi) \propto g(\phi) \prod_{i=1}^N \exp[A_i(\phi)^{-1} \{y_i \eta_i - B_i(\eta_i)\} + C_i(y_i; \phi)].$$

Sampling from a normal or inverse gamma distribution is very simple. In Part 5 of Fact 5.1, the conditional density of η_i or v_i is often log-concave. For sampling from a log-concave density, Gilks and Wild's (1992) adaptive method or Berger and Sun's (1993) direct method can be used. Here are Poisson and binomial examples.

Example 4.1 (continued). When $h_i(\eta_i) = \eta_i - \log(m_i) = \log(p_i)$,

$$s_i(\eta_i) \propto \exp\left[y_i\eta_i - e^{\eta_i} - \frac{1}{2\delta_0}\{\eta_i - \log(m_i) - (\mathbf{x}_{1i}^t\boldsymbol{\theta} + \mathbf{x}_{2i}^t\mathbf{Z})\}^2\right].$$

Therefore

$$\frac{\partial^2}{\partial\eta_i^2} \log\{s_i(\eta_i)\} = -e^{\eta_i} - \delta_0^{-1} < 0.$$

Consequently, the conditional density of η_i given $(\phi, \mathbf{Z}, \delta_0, \delta_1)$ is log-concave. Since v_j is a linear transformation of η_i , the conditional density of v_j is also log-concave.

Example 4.2 (continued). When $h_i(\eta_i) = \eta_i = \log\{p_i/(1-p_i)\}$, we have

$$s_i(\eta_i) \propto \exp\left[y_i\eta_i - m_i \log(1 + e^{\eta_i}) - \frac{\{\eta_i - \log(m_i) - (\mathbf{x}_{1i}^t\boldsymbol{\theta} + \mathbf{x}_{2i}^t\mathbf{Z})\}^2}{2\delta_0}\right].$$

We can show that

$$\frac{\partial^2}{\partial\eta_i^2} \log\{s_i(\eta_i)\} = -m_i e^{\eta_i} (1 + e^{\eta_i})^{-2} - \delta_0^{-1} < 0.$$

So the conditional density of $\eta_i = v_i$ given $(\phi, \mathbf{Z}, \delta_0, \delta_1)$ is log-concave.

Example 4.3 (continued). When $h_i(\eta_i) = \eta_i$, we have

$$s_i(\eta_i) \propto \exp\left[-\frac{(y_i - \eta)^2}{2\sigma^2} - \frac{\{\eta_i - (\mathbf{x}_{1i}^t\boldsymbol{\theta} + \mathbf{x}_{2i}^t\mathbf{Z})\}^2}{2\delta_0}\right].$$

Clearly, the conditional distribution of η_i given others is normal with mean $\delta_i(\sigma^2 + \delta_i)^{-1}y_i + \sigma^2(\sigma^2 + \delta_i)^{-1}(\mathbf{x}_{1i}^t\boldsymbol{\theta} + \mathbf{x}_{2i}^t\mathbf{Z})$ and variance $\sigma^2\delta_i(\sigma^2 + \delta_i)^{-1}$.

Example 4.4 (continued). When $h_i(\eta_i) = \log(\eta_i)$, we have

$$s_i(\eta_i) \propto \exp\left[-\alpha y_i \eta_i + \alpha \log(\eta_i) - \frac{\{\log(\eta_i) - (\mathbf{x}_{1i}^t\boldsymbol{\theta} + \mathbf{x}_{2i}^t\mathbf{Z})\}^2}{2\delta_0}\right].$$

This conditional density is not necessary logconcave. However, its transformation $\xi_i = \log(\eta_i)$ has the conditional density

$$\tilde{s}_i(\xi_i) \propto \exp\left[-\alpha y_i e^{\xi_i} + (\alpha + 1)\xi_i - \frac{\{\xi_i - (\mathbf{x}_{1i}^t\boldsymbol{\theta} + \mathbf{x}_{2i}^t\mathbf{Z})\}^2}{2\delta_0}\right].$$

It is easy to verify that $\frac{\partial^2}{\partial\xi_i^2} \log\{\tilde{s}_i(\xi_i)\} = -\alpha y_i e^{\xi_i} - \delta_0^{-1}$, which is negative. Consequently, we can simply sample from the logconcave density of ξ_i , then make the transformation $\eta_i = e^{\xi_i}$.

- Gilks, W.R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337–348.
- Ghosh, M., Natarajan, K., Stroud, T.W.F., & Carlin, B.P. (1998). Generalized linear models for small area estimation. *Journal of the American Statistical Association*, 93, 273–282.
- Kindermann, R. and Snell, J.L. (1980). *Markov Random Fields and Their Applications*. Amer. Math. Soc., Providence. RI.
- Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320–338.
- He, Z., and Sun, D. (1998). Hierarchical Bayes estimation of hunting success rates with spatial correlations. Revised for *Biometrics*.
- Hobert, J.P. & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91, 1461–1473.
- Hobert, J.P. & Casella, G. (1998). Functional compatibility, Markov chains and Gibbs sampling with improper posteriors. *J. of Computational and Graphical Statistics*, 7, 42–66.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Chapman & Hall, London.
- Ord, K. (1975). Estimation methods for models with spatial interaction. *Journal of the American Statistical Association*, 70, 120–126.
- Ripley, B.D. (1981). *Spatial Statistics*. Wiley, New York.
- Ripley, B.D. (1988). *Statistical Inference for spatial Processes*. Cambridge University Press, Cambridge.
- Sun, D., Tsutakawa, R.K., Kim, H. & He, Z. (1997). Spatio-Time Interaction with Disease Mapping. Submitted.
- Sun, D., Tsutakawa, R.K., and He, Z. (1998). Propriety of posteriors with improper priors in hierarchical linear mixed models. Submitted.
- Sun, D., Tsutakawa, R.K. and Speckman, P.L. (1997). Bayesian inference for CAR (1) models with noninformative priors. *Biometrika*, in press.
- Tsutakawa, R.K. (1988). Mixed model for analysing geographic variability in mortality rates. *Journal of the American Statistical Association*, 83, 117–130.
- Waller, L.A., Carlin, B.P., Xia, H., & Gelfand, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92, 607–617.
- West, M. and Aguilar, O. (1997). Studies of Quality Monitor Time Series: The VA Hospital System, Discussion Paper 97-20a, ISDS, Duke University.