Bayesian Analysis for Generalized Linear Mixed Models

Dongchu Sun
University of Missouri–Columbia

November 11, 2008
A General Linear Mixed Model
Consider a general linear mixed model

\[ v_i = x_{1i}^t \theta + x_{2i}^t u + e_i, \]  

(1)

where \( V = (v_1, \cdots, v_n)^t \) is an \( n \times 1 \) vector of data,
\( X_1 = (x_{11}, \cdots, x_{2n})^t : n \times p \) known;
\( X_2 = (x_{11}, \cdots, x_{2n})^t : n \times q \) known;
\( \theta : \) a \( p \times 1 \) vector of fixed effects,
u = (u_1', \cdots, u_r')' : r random vectors of random effects
and \( e = (e_1, \cdots, e_n)^t \) is an \( n \times 1 \) vector of random errors.

\[
\begin{aligned}
\left\{
\begin{array}{l}
u | (\delta_1, \cdots, \delta_r) \sim N_q(0, A^{-1}), \\
 e | \delta_0 \sim N_n(0, \delta_0 I),
\end{array}
\right.
\]

(2)

where \( u_i \) is \( q_i \times 1 \), \( \sum_{i=1}^r q_i = q \), \( A = diag(\delta_1^{-1} B_1, \cdots, \delta_r^{-1} B_r) \),
\[ B_i = (I_{q_i} - \rho_i C_i) \xi_i. \]  

(3)

Here \( C_i \) is known and symmetric, \( \xi_i \geq 0 \) is a known integer and \( \rho_i \in (\lambda_{i1}^{-1}, \lambda_{iq_i}^{-1}) \), \( \lambda_{i1} \) and \( \lambda_{iq_i} \) are smallest and the largest eigenvalues of \( C_i \).
Some Gaussian Random Effects $Z$

- **An AR(1) Model.** For $\rho \in (-1, 1)$,

  $$Z_i = \rho Z_{i-1} + \epsilon_i, \ i = 2, \ldots, k.$$  

  $\epsilon_i$'s iid $\sim N(0, \delta_1)$; $\rho$ is a correlation of $Z_i$ and $Z_{i+1}$. If $Z_1 \sim N(0, \delta_1/(1 - \rho^2))$, $Z_i$ is stationary.

- **A general AR(1) model (Ord, 1975).**

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

  where the $C_{ij}$ are fixed constants satisfying $C_{ii} = 0$, and $\epsilon_1, \ldots, \epsilon_k$ are iid $N(0, \delta_1)$. $\rho$ is a “correlation coefficient,” not correlation of $Z_i$ and $Z_{i+1}$.

- **A CAR(1) Model.** One popular CAR model takes ($d_i = \sum_j C_{ij}$),

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

  (5)
A general CAR model (Besag, 1974):

\[
f(Z_i | Z_{-i}) = \left( \frac{\alpha_i}{2\pi\delta_1} \right)^\frac{1}{2} \exp\left\{ -\frac{\alpha_i}{2\delta_1} \left( Z_i - \sum_{j \neq i}^k \beta_{ij} Z_j \right)^2 \right\},
\]

(6)

Let \( B = (b_{ij}), k \times k \) : \( b_{ii} = \alpha_i \) and \( b_{ij} = -\alpha_i \beta_{ij}, i \neq j \).

If \( B \) is symmetric and positive definite, the full conditional distributions lead to the joint pdf of \( Z \),

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

(7)

i.e. \( Z \sim \text{MVN}(0, \delta_1 B^{-1}) \). In our context, Let

\[
B = I - \rho C, \quad C = (C_{ij}).
\]

Let \( \lambda_1 (\lambda_k) \) be the smallest (largest) eigenvalue of \( C \).

If \( \lambda_1^{-1} < \rho < \lambda_k^{-1} \), \( I - \rho C \) is positive definite. Then

- \( \text{AR}(1): Z \sim \text{MVN}(0, \delta_1 (I - \rho C)^{-2}). \)
- \( \text{CAR}(1): Z \sim \text{MVN}(0, \delta_1 (I - \rho C)^{-1}). \)
Strongly Correlated Random Effects ($|B| = 0$)

For given full conditional distributions, 

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

Again let $B = (b_{ij})$, $b_{ii} = \alpha_i$ and $b_{ij} = -\alpha_i\beta_{ij}, i \neq j$.

If $B$ is not positive definite, there is $f(Z) \geq 0$ so that 

$$f(Z_i|Z_{-i}) = f(Z) \left/ \int_{-\infty}^{\infty} f(Z) dZ_i. \right.$$

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

In particular, we may choose 

$$f(Z) \propto (2\pi\delta_1)^\frac{3}{2} |B|_+^\frac{1}{2} \exp\left( -\frac{1}{2\delta_1} Z^t B Z \right),$$
where $|B|_+$ is the product of $r$ positive eigenvalues.

- Need specify either full conditional distributions or $B$. 
Special Cases when $|B| = 0$?

- *Pairwise Differences Priors* (Besag et al., 1995):

$$f(Z) \propto \exp\{-\frac{1}{\delta_1} \sum_{i,j} w_{ij} (Z_i - Z_j)^2\},$$

where $w_{ij} \geq 0$ are constants.

- *Markov Random Field Priors* (Besag et al., 1991):

If $B = \text{diag}(d_1, \cdots, d_k) - C$, where $C = (C_{ij}), C_{ij} = 1$ iff $i$ and $j$ are neighbors. Then

$$Z_i | Z_{-i} \sim N\left(\frac{1}{d_i} \sum_{j \neq i} Z_j, \frac{\delta_1}{d_i}\right)$$

(See, Ghosh et al., 1998).
• The 1st Order Difference Priors (Clayton, 1996):

\[
B = \begin{pmatrix}
1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\
-1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\
& & \ddots & & \ddots & & \ddots & \\
0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\
0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1
\end{pmatrix}_{k \times k}
\]
• The 2nd Order Difference Priors (Clayton, 1996):

\[
B = \begin{pmatrix}
1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
-2 & 5 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\
1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 \\
. & . & . & . & . & \cdots & . & . & . & . \\
0 & 0 & 0 & 0 & 0 & \cdots & 1 & -4 & 5 & -2 \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \\
\end{pmatrix}.
\]
Remark. If $|B| = 0$, $Z$ is partial informative normal.

Write $B = \Gamma \Lambda \Gamma^t$, where $\Gamma = (\Gamma_1, \Gamma_2)$ is orthogonal,

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}, \Lambda_1 = diag(\lambda_1, \ldots, \lambda_r), \text{ and } \lambda_1 > 0.$$ 

Then $B = \Gamma_1 \Lambda_1 \Gamma_1^t$.

Let $U_1 = (U_1, \ldots, U_r)^t$, where $U_i$ indep $\sim N(0, \delta_1 \lambda_i^{-1})$, independently, $U_2 = (U_{r+1}, \ldots, U_k)^t$ follows a constant density over a $k - r$ dimensional Euclidian space.

Now define $Z = \Gamma (U_1^t, U_2^t)^t$. The joint density of $Z$ still has the form

$$f(Z) \propto \delta_1^{-q/2} (\prod_{i=1}^r \lambda) \frac{1}{2} \exp \left( -\frac{1}{2\delta_1} Z^t B Z \right). \quad (8)$$

This density is improper if $B$ is not positive definite, but it is a partially informative normal distribution because it is proper on certain directions and noninformative on other directions.
Assume that
\[ p(\theta) \propto 1; \quad p(\delta_i) \propto \delta_i^{-a_i} \exp(-b_i/\delta_i). \]

Also assume arbitrary distribution \( F_i(\rho_i) \) on \((\lambda_i^{-1}, \lambda_{iq_i}^{-1})\).

**Theorem 1** 1. If \( t = q \) or if \( r = 1 \) the following conditions (a), (b), and (c1) are necessary and conditions (a), (b) and (c2) are sufficient for the propriety of the posterior distribution of \((\theta, u, \Delta, \rho, \delta_0)\), respectively.

(a) For each \( i = 1, \cdots, r \), either (a1) or (a2) holds:
\[
\text{(a1)} \quad a_i < b_i = 0, \\
\text{(a2)} \quad b_i > 0;
\]
(b) \( q_i + 2a_i > q - t \), for all \( i = 1, \cdots, r \);
(c1) \( n - p + 2a_0 + 2a_+ > 0 \), where \( a_+ = \sum_{i=1}^r a_i \);
(c2) \( n - p + 2a_0 + 2 \sum_{i=1}^r a_i^- > 0 \), where \( a_i^- = \min(0, a_i) \).

2. If \( t < q \) and \( r > 1 \), then conditions (a), (b)', and (c1) are necessary and conditions (a), (b)', and (c2) are sufficient for the propriety of the joint posterior, where
\[
\text{(b)'} \quad q_i + 2a_i > 0.
\]
Remark 1  There are two simple cases for the values of $b_0$. First, Since $\text{SSE} \geq 0$, when $b_0 > 0$, $2b_0 + \text{SSE} > 0$. Note that $\text{SSE} > 0$ with probability one. When $b_0 = 0$, we have $2b_0 + \text{SSE} > 0$ with probability one so that the results will be true for almost all observations. Furthermore, if all $a_i \leq 0, i = 1, \cdots, r$, then conditions (c1) and (c2) are identical.

Remark 2  Such a linear mixed model can also be used as the prior for a generalized linear mixed model.
Special Cases of $a_i$ and $b_i$.

Case 1. Hobert and Casella (1996) consider the case where $b_i \equiv 0$ and $\rho_i \equiv 0$ for all $i$, which is a special case of Theorem 2.

Case 2. Datta and Ghosh (1992) considered the case where $b_i > 0$ for all $i = 1, \cdots, r$. From Theorem 2, we know that when all $b_i > 0$, we need additional conditions (b) (or (b)’) and (c2) to get a proper posterior. Clearly, condition (b) is stronger than condition (b)’. The latter is necessary. There are examples showing that when the necessary condition (b)’ holds, the posterior distribution may or may not be proper.
Case 3. The constant prior has been used in the literature. For example, Yang and Chen (1995) used constant priors for a random effects model. In general, we could have the prior,

$$
\pi(\theta, u, \Delta, \rho, \delta_0) \propto 1. 
$$

(9)

This is a special case of the improper priors we considered by choosing $a_i = -1$ and $b_i = 0$. Note that under this constant prior, the Restricted MLE of $(\Delta, \rho, \delta_0)$ is the marginal posterior mode. Note that in the case conditions (c1) and (c2) are identical.

**Corollary 1** Assume that the prior (9) is used. In addition, if $q_i - 2 > 0$ and $n - p - 2(r + 1) > 0$, then the joint posterior distribution of $(\theta, u, \Delta, \rho, \delta_0)$ exists.
Case 4. Another interesting improper prior for $\delta_i$ is of the form

$$g_i(\delta_i) \propto \delta_i^{-1/2}.$$ \hspace{1cm} (10)

This is a special case by choosing $a_i = -1/2$ and $b_i = 0$.

**Corollary 2**  Under the assumptions of Theorem 1 and (10), if $q_i - 2 > 0$ and $n - p - 2(r + 1) > 0$, then the joint posterior of $(\theta, u, \Delta, \rho, \delta_0)$ exists.

Clearly the behavior of the prior (10) is similar to the scale-invariant prior $1/\delta_i$, while the latter will give improper posterior. In practice the prior (10) or the constant prior might be used.
The propriety of the posterior in terms of the hyperparameters \((a_i, b_i)\) \((i = 1, \cdots, r)\), may now be summarized according to the following table.

**Table 1: Propriety of the Posterior**

<table>
<thead>
<tr>
<th></th>
<th>(a_i &lt; 0)</th>
<th>(a_i = 0)</th>
<th>(a_i &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_i = 0)</td>
<td>proper</td>
<td>improper</td>
<td>improper</td>
</tr>
<tr>
<td>(b_i &gt; 0)</td>
<td>proper</td>
<td>proper or improper</td>
<td>proper</td>
</tr>
</tbody>
</table>
Assume $X_1$ is of full rank ($X'_1X_1$ is invertible).
Let $X = (X_1, X_2)$ and $\beta = (\theta', u')'$.
The usual LSE of $\beta$ is $\hat{\beta} = (X'X)^{-1}X'v$.

**Proposition 1** If the prior of $\rho_i$ is uniform on $(\lambda_{i1}, \lambda_{iq})$, the full conditional distributions are as follows:

1. $\theta|(u, \delta_0, \Delta, \rho, v) \sim MVN_p(\hat{\theta} + L_1, \delta_0(X'_1X_1)^{-1})$, where $L_1 = (X'_1X_1)^{-1}X'_1X_2(\delta_0^{-1}X_2u)$.
2. $u|(\theta, \delta_0, \Delta, \rho, v) \sim MVN_q(\delta_0^{-1}M_1X'_2(v - X_1\theta), M_1)$, where $M_1 = (\delta_0^{-1}X'_2X_2 + A^{-1})^{-1}$.
3. $\delta_0|(\theta, u, \Delta, \rho, v) \sim IG(a_0 + \frac{n}{2}, b_0 + \frac{1}{2}(v - X_1\theta - X_2u)'(v - X_1\theta - X_2u))$.
4. $\delta_i|(\theta, u, \Delta_{(-i)}, \rho, v) \sim IG(a_i + \frac{q_i}{2}, b_i + \frac{1}{2}u'_iB_iu_i)$, where $\Delta_{(-i)} = (\delta_j, j \neq i)$, for $i = 1, \ldots, r$.
5. let $\rho_{(-i)} = (\rho_j, j \neq i)$. The conditional density of $\rho_i$ given $(\theta, u, \Delta, \rho_{(-i)}, v)$ is of the form

$$h_i(\rho_i) = |B_i|^{q_i/2} \exp \left\{ -\frac{1}{2\delta_i}u'_iB_iu_i \right\}.$$
Sampling from the first three conditional distributions is straightforward. Note that if $\xi_i = 0$, the parameter $\rho_i$ does not affect other parameters. We will consider only the case where $q_i > 1$ and $\xi_i \geq 1$. Here is a property of logconcavity for the conditional density of $\rho_i$. From this, the adaptive rejection sampling method by Gilks and Wild (1992) can be used.

**Lemma 1** If $\xi_i \geq 1$, the conditional density of $\rho_i$ is log-concave.
**A Generalized Linear Mixed Model**

Let $Y_1, \ldots, Y_N$ be indep. rv’s. $Y_i$ has the pdf

$$f_i(y_i| \eta_i, \phi) = \exp \left[ \frac{1}{A_i(\phi)} \{y_i \eta_i - B_i(\eta_i)\} + C_i(y_i; \phi) \right], \quad (11)$$

$A_i(\phi) = \phi w_i^{-1}$, for prespecified weights $w_i$.

Known $\phi$:

- $Y_i \sim \text{Poisson} \left( m_i p_i \right)$, where
  - $\phi = 1, A_i(\phi) = 1, \eta_i = \log[p_i/(1-p_i)],$
  - $B_i(\eta_i) = m_i \log\{1 + \exp(\eta_i)\},$ and
  - $C_i(y_i; \phi) = \log[m_i!\{y_i!(m_i - y_i)\}]$.

- $Y_i \sim \text{Binomial} \left( m_i, p_i \right)$, where
  - $\phi = 1, A_i(\phi) = 1, \eta_i = \log(m_ip_i), B_i(\eta_i) = \exp(\eta_i),$ and
  - $C_i(y_i; \phi) = -\log(y_i!)$.
— $Y_1, \ldots, Y_N$ are indep. rv’s. $Y_i$ has the pdf

$$f_i(y_i|\eta_i, \phi) = \exp\left[\frac{1}{A_i(\phi)}\{y_i\eta_i - B_i(\eta_i)\} + C_i(y_i; \phi)\right],$$

Unknown $\phi$:

- $Y_i|(\mu_i, \sigma^2) \sim N(\mu_i, \sigma^2)$, we have $\eta_i = \mu_i$, $\phi = \sigma^2$, $A_i(\phi) = \phi$, $B_i(\eta_i) = \eta_i$ and $C_i(y_i, \phi) = -0.5 \log(\phi) - y_i^2/(2\phi)$.

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

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

Here $\alpha$ is the common shape parameter and $\mu_i$ is the mean of $Y_i$ for given $(\mu_i, \alpha)$. Here $\phi = \alpha$, $\eta_i = 1/\mu_i$, $A_i(\phi) = -1/\phi$, $B_i(\eta_i) = \log(\eta_i)$, and $C_i(y_i, \phi) = \alpha \log(\alpha) + (\alpha - 1) \log(y_i) - \log[\Gamma(\alpha)]$. 
Generalized Linear Models.

To model the variability in $\eta_i$ to account for various fixed covariates.

$$h_i(\eta_i) = x_{1i}^t \theta,$$

where the $h_i$ are known monotone functions,

$X_1 = (x_{11}, \ldots, x_{1n})^t$ is an $N \times p$ design matrix and $\theta$ is the vector of fixed effects.

- Commonly referred to as a GLM with canonical parameter $\eta_i$, scale parameter $\phi$, and link function $h_i$. (cf. McCullagh and Nelder, 1989).

- Often $h_i = h$. 
Generalized Linear Mixed Models.

Extend the model to include random effects. Let

\[ h_i(\eta_i) = x_{1i}^t \theta + x_{2i}^t Z, \]  

(13)

where \( h_i \) is a known monotone function,

\( X_1 = (x_{11}, \ldots, x_{1n})^t : N \times p \) design matrix;

\( X_2 = (x_{21}, \ldots, x_{2n})^t : N \times k \) design matrix;

\( \theta: p \times 1 \) fixed effects, and \( Z: k \times 1 \) random effects.

- Models given by \( f_i(y_i | \eta_i, \phi) \) and (13) are called generalized linear mixed models (GLMMs).

- Were widely used in disease mapping e.g., Breslow and Clayton (1993).
**GLMMs With Extra Residual Effects.**

Further extend it to add additional residual effects,

\[ h_i(\eta_i) = x_{1i}^t \theta + x_{2i}^t Z + e_i. \]  

(14)

Here \( e = (e_1, \ldots, e_N)^t \) are residual effects satisfying some restriction such as \( \mathbb{E}(e_i) = 0 \) or \( \mathbb{E}\exp(e_i) = 1 \). In addition, \( Z \) and \( e \) are assumed mutually independent.

- \( e_i \) accounts for the lack of fit due to extra variation, outliers, and other unexplained sources of variation.
- \( e_i \) is different from \( Z \):
  - \( Z \) oftens accounts for some special pattern such random geographical effects and spatial correlation.
  - \# of components of \( Z \) is smaller than \( N \), \# of \( e_i \).
- By a suitable choice of \( X_2 \), (14) may be encompassed under (13), but we do not do this to emphasize the separate roles of \( Z \) and \( e \).
- We will also call (14) a GLMM.
Possible Residual Effects

- *Normal Residual Effects:*
  \[ e_1, \ldots, e_N \text{ are iid } N(0, \sigma_0). \]

- *Normal Residual Effects:*
  \[ e_1, \ldots, e_N \text{ are iid } \Gamma(R, R), \text{ with pdf} \]
  \[ f(t) = \frac{R^R}{\Gamma(R)} w^{R-1} \exp(-Rw). \]

We will focus on the Normal residuals only.

**Remarks on Random Effects \( Z \).**

- Traditional treatment of random effects assumes independent and identical distributed.

- Recently, there is a large literature on correlated random effects, especially using CAR (conditional autoregressive, Besag, 1974, JRSSB) models to capture spatial correlations.
Overview

- With the advent of MCMC methods, especially Gibbs sampling, (generalized) linear mixed models with correlated random effects can be handled easily.

- Commonly used CAR models, especially pairwise difference priors are typically improper since it is constant when all $Z_i$ are equal. Hobert and Casella (1996, JASA) pointed out the importance of having a proper posterior in running Gibbs sampling.

- If $B$ is not p.d., one could put constraints on the random effects, but the full condi. dist’s will be changed.

- Will examine the propriety of the posterior in using CAR models.

- Bayesian computation via MCMC.
Hierarchical GLMMs

Let $H_i$ be the inverse function of $B'_i$. Define

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

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

**Th 1.** For $f_i(\cdot|\eta_i, \phi)$ and GLMMs assume that

a) There is a $\mathcal{J}_n = (i_1, \ldots, i_n) \subset \{1, \ldots, N\}$, such that

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

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

b) Rank of $X_1^* = (x_{1,i_1}, \ldots, x_{1,i_n})^t$ is $p$, and rank of $X_2^* = (x_{2,i_1}, \ldots, x_{2,i_n})^t$ equals to rank of $X_2 = (x_{2,1}, \ldots, x_{2,N})^t$;

c) $p(\theta) \propto 1$ and $p(Z) \propto (2\pi\delta_1)^{\frac{n}{2}} |B|^{\frac{1}{2}} \exp\left(-\frac{1}{2\delta_1} Z^t B Z\right)$,

d) rank $(X_2^{*t} R_1 X_2^* + B) = k$, $R_1 = I_n - X_1^* (X_1^{*t} X_1^*)^{-1} X_1^*$;
e) The prior of $(\delta_0, \delta_1)$ satisfies the condition,

$$\int \int \{ \delta_0^{-\frac{1}{2}(n-p-k)} \delta_1^{-\frac{1}{2}k} + \delta_0^{-\frac{1}{2}(n-p)} \} p(\delta_0, \delta_1) d\delta_0 d\delta_1 < \infty.$$ 

The posterior of $(\eta, \phi, \theta, Z, \delta_0, \delta_1)$ is proper.

**Remark 3** A common prior for the variance components $\delta_i$ is inverse gamma$(a_i, b_i)$, whose density is

$$g_i(\delta_i) \propto \frac{1}{\delta_i^{a_i+1}} \exp(-b_i/\delta_i). \quad (15)$$

Clearly, when $b_i > 0$, $n - p - k + 2a_0 > 0$ and $k > 2a_1$, Condition (e) holds.
Remark 4  When the prior of $\phi$ is degenerate, as in the Poisson or binomial cases, Condition (a) 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,

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

- For binomial case, there are $n$ $y'_i$s, $0 < y_i < m_j$.
- For Poisson case, it requires $0 < y_i, j = 1, \ldots, n$.
- For normal case, it is equivalent to

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

which holds if $N = n$ and $\phi$ has a proper prior.
• For gamma case, it is equivalent to

\[
\int_0^\infty \left\{ \frac{\alpha^\alpha}{e^{\alpha} \Gamma(\alpha)} \right\}^{N-n} F(d\alpha) < \infty,
\]

which holds if \( N = n \) and \( \phi \) has a proper prior.

**Th 2.** Assume that the rank of \((X_2^t R_1 X_2 + B) < k\), where \( R_1 = I_N - X_1 (X_1^t X_1)^{-1} X_1^t \), and \( X_1 \) and \( X_2 \) are design matrices based on the full data. Under Assumption (d), for any proper prior of \((\delta_0, \delta_1, \phi)\), the joint posterior of \((\eta, \phi, \theta, Z, \delta_0, \delta_1)\) is improper.

**Remark 5** When \( B = D - C \), if 1, a vector of 1’s, is in the column space of \((X_1, X_2)\), then the posterior is always improper.
Bayesian Computation—via MCMC

Assume that the prior for \( \delta_i \) is inverse gamma \((a_i, b_i)\).

**Fact 1** The full conditional distributions are as follows.

1. \( \theta | (\eta, \phi, Z, \delta_0, \delta_1) \sim \mathcal{N}_p((X_1^t X_1)^{-1} X_1^t (V - X_2 Z), \delta_0 (X_1^t X_1)^{-1}). \)
2. \( Z | (\eta, \phi, \theta, \delta_0, \delta_1) \sim \mathcal{N}_k(M_1 X_2^t (V - X_1 \theta), \delta_0 M_1), \) where \( M_1 = (X_2^t X_2 + \delta_0 \delta_1^{-1} B)^{-1}. \)
3. \( \delta_0 | (\eta, \phi, \theta, Z, \delta_1) \sim \text{inverse gamma}(a_0 + \frac{n}{2}, \)
   \( b_0 + \frac{1}{2} (V - X_1 \theta - X_2 Z)^t (V - X_1 \theta - X_2 Z). \)
4. \( (\delta_1 | \eta, \phi, \theta, Z, \delta_0) \sim \text{inverse gamma}(a_1 + \frac{k}{2}, b_1 + \frac{1}{2} Z^t B Z). \)
5. Given \((\phi, Z, \delta_0, \delta_1)\), the \( \eta_j \) (or \( v_j = h_j(\eta_j) \)) are independent. The pdf of \( v_j \) is often log-concave.
6. If the prior for \( \phi \) is degenerate, so is its posterior; if \( \phi \) has the prior density \( g(\phi) \), then its posterior density given \((\eta, \theta, Z, \delta_0, \delta_1)\) is

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

- Gaussian CAR model gives a specific form to model spatial correlation. Besag et al. (1991) assumed that $B = D - C$. It does not tell the degree of correlation.

- Sun, Tsutakawa and He (1997) considered noninformative priors for GLMMs, when $Z \sim \text{MVN}(0, \delta_1 (I - \rho C)^{-k})$. CAR(1) and AR models are special cases when $k = 1$ and 2, respectively. They model independent ($\rho = 0$) or correlated random effects ($\rho \neq 0$).

- One can argue that it is possible to assume $B = D - \rho C$, $\rho \in (-1, 1)$. In this case even if $\rho = 0$, $Z_i$ are independent normal with mean 0 but unequal variances. Furthermore, Like the case of $B = D - C$, it often requires updating each component in running Gibbs sampling unless one wants to compute an inverse of $q \times q$ matrix in each Gibbs cycle.

- The CAR and AR models ($B = (I - \rho C)^k, k = 1, 2$) allow us to use a very efficient algorithm for simulating all regional parameters as a
block and result in an MCMC sampler that converges in far fewer iterations than would be required for a similar model employing a pairwise difference prior and simulating each spatial parameter individually (cf. He and Sun (1997), Sun et al. (1998)).