NISS

Posterior Distribution of Hierarchical Models Using CAR(1) Distributions

Dongchu Sun, Robert K. Tsutakawa, and Paul Speckman

Technical Report Number 96
April, 1999

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

Posterior distribution of hierarchical models using CAR(1) distributions

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

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

e-mail: dsun@stat.missouri.edu, sutakawa@stat.missouri.edu

and speckman@stat.missouri.edu

Summary

We examine properties of the CAR(1) model, which is commonly used to represent regional effects in Bayesian analyses of mortality rates. We consider a Bayesian hierarchical linear mixed model where the fixed effects have a vague prior such as a constant prior and the random effect follows a class of CAR(1) models including those whose joint prior distribution of the regional effects is improper. We give sufficient conditions for the existence of the posterior distribution of the fixed and random effects and variance components. We then prove the necessity of the conditions and give a one-way analysis of variance example where the posterior may or may not exist. Finally, we extend the result to the generalised linear mixed model, which includes as a special case the Poisson log-linear model commonly used in disease mapping.

Some key words: Partially informative normal distribution; Spatial correlation; Gibbs sampling; Poisson distribution; Multivariate normal; Linear mixed model.

1. Introduction

This paper considers the propriety of the posterior distribution for the general mixed linear model and generalised mixed linear model when the random effects are represented by the conditional autoregressive model, or CAR(1), introduced by Besag (1974). CAR(1) is currently one of the most important and widely used models to represent spatial correlations in disease mapping (Clayton & Kaldor, 1987; Cressie & Chan, 1989; Marshall, 1991; Bernardinelli, Clayton & Montomoli, 1995; and Waller et al., 1997).

The recent popularity of CAR(1) is primarily due to the ease with which it may be implemented in the Gibbs sampler (Gelfand & Smith, 1990). However, such convenience may lead to overlooking the possibility that the posterior distribution may fail to exist when the joint distribution under CAR(1) is improper.

The use of the CAR(1) model to represent spatial effects may be illustrated by a log-linear model in mortality analysis. For a given target population of size m, let Y denote the frequency of the deaths due to some specific cause, such as lung cancer, during some fixed time period. Conditionally on a population parameter p, assume that Y has a Poisson distribution with mean mp, where p may be interpreted as the rate per individual. The target populations are typically cross-classified by demographic variables such as age, sex and geographic region. The dependence of p on such covariates can be represented by a log-linear model for p having the form

$$V = X_1\theta + X_2 Z + e, \tag{1}$$

where V is a vector of the set of $\log(p)$ associated with the target populations, θ is a vector of fixed effects, Z is a vector of random regional effects, and X_1 and X_2 are design matrices. The vector e represents unexplained random effects and is often omitted in the literature. Related models may be found in Tsutakawa (1988) and Marshall (1991).

In practice, two forms of the CAR(1) model are widely used to represent spatial effects. Let $Z = (Z_1, \cdots, Z_q)'$ denote the real-valued regional effects of q regions, Λ_i the set of regions that are geographically adjacent to region i and d_i the number of regions in Λ_i , $i = 1, \cdots, q$. In Model 1, the conditional distribution of Z_i given the other regional effects $Z_{-i} = (Z_1, \cdots, Z_{i-1}, Z_{i+1}, \cdots, Z_q)'$ is assumed to be $N(\sum_{j \in \Lambda_i} Z_j/d_i, \delta_1/d_i)$. In Model 2, the conditional distribution is assumed to be $N(\rho \sum_{j \in \Lambda_i} Z_j, \delta_1)$. Model 1 was proposed by Besag, York, & Mollié (1991) and used by Bernardinelli & Montomoli (1992), Bernardinelli et al. (1995), Waller et al. (1997) and Ghosh et al. (1998), among others. Model 2 was proposed by Clayton & Kaldor (1987) and used by N. J. McMillan and L. M. Berliner in a National Institute of Statistical Sciences technical report. Both models are designed to account for spatial correlations among neighbouring regions.

In §2 we examine the joint distribution of the CAR(1) model and demonstrate that, when the covariance matrix of Z is not positive definite, e.g. Model 1, the joint distribution may be decomposed into a component which is nonsingular normal and another which has a constant density over some Euclidian space, implying that the distribution is improper. We also examine properties of Models 1 and 2 and introduce a modification (Model 1A) of Model 1 which has a proper joint distribution.

The effect of such improper distributions on the posterior distribution for hierarchical models is examined in §3. We consider the linear mixed model where the fixed effects have a uniform prior and the random effects have an arbitrary Gaussian CAR(1) distribution. We give sufficient conditions for the linear effects and variance components to have a proper posterior distribution. We also show the necessity of one of these conditions and provide an illustration of a balanced one-way analysis of variance model where the posterior may or may not be proper. Finally, we prove the propriety of the posterior distribution for the generalised linear mixed model, which includes the mortality example as a special case. Our results are closely related to these of Ghosh et al. (1998), who have previously shown the existence of the posterior under Model 1, and to those of Hobert & Casella (1996), who considered the hierarchical model where the covariance matrix of the random effects is positive definite. The discussion of the CAR(1) model is also related to Hobert & Casella's (1998) functional compatibility.

2. Gaussian CAR(1) model

Let $Z = (Z_1, \cdots, Z_q)'$ be a random vector with full conditional densities

$$f(Z_i|Z_{-i}) = \Big(\frac{a_i}{2\pi\delta_1}\Big)^{\frac{1}{2}}\exp\Big\{-\frac{a_i}{2\delta_1}\Big(Z_i - \sum_{j\neq i}^{q}\beta_{ij}Z_j\Big)^2\Big\}, \tag{2}$$

$i = 1, \cdots, q$. Let B be the $q \times q$ matrix with diagonal elements a_i and ijth off-diagonal element $-a_i\beta_{ij}$. Besag (1974) proved that, if B is symmetric and positive definite, these conditional distributions lead to the joint probability density function of Z,

$$f(Z) = (2\pi\delta_1)^{-q/2}|B|^{1/2}\exp\Big\{-\frac{1}{2\delta_1}Z'BZ\Big\}. \tag{3}$$

In this case, Z is multivariate normal with mean 0_q and covariance matrix $\delta_1 B^{-1}$. When B is nonnegative definite but not positive definite, the relationship between

(2) and the joint distribution of Z is less clear. In this case we will call the joint distribution a partially informative normal distribution and represent the density by

$$f(Z) \propto \delta_1^{-q/2} \exp\Big\{-\frac{1}{2\delta_1} Z'BZ\Big\}. \tag{4}$$

We first note that Z cannot have a singular normal distribution that is consistent with (2). If Z has a singular normal distribution, Z is nonsingular normal over some hyperplane, so there exists at least one Z_i such that the conditional distribution of Z_i, given Z_{-i}, is degenerate, contradicting (2).

Suppose B is singular nonnegative definite with rank r. Then there exists an orthogonal matrix , such that $\Lambda = ,'B, = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)$, where $\lambda_i > 0$, $i = 1, \cdots, r < q$. Let $X = ,'Z$. Since $B = , \Lambda, '$, $Z'BZ = X'\Lambda X = \sum_{i=1}^r \lambda_i X_i^2$. Thus, if the density of Z is given by (4), the density of X is

$$f(X) \propto \delta_1^{-q/2} \exp\Big(-\frac{1}{2\delta_1} \sum_{i=1}^{r} \lambda_i X_i^2\Big),$$

i.e. $(X_1, \cdots, X_r)$ are independent normal variables with mean 0 and variance δ_1/λ_i, $i = 1, \cdots, r$, and $(X_{r+1}, \cdots, X_q)$ has density proportional to $\delta_1^{-(q-r)/2}$ over a $q - r$ dimensional Euclidian space. Thus the distribution of Z is improper when B is singular. Note that it may be more intuitive to use $\delta_1^{-r/2}$ rather than $\delta_1^{-q/2}$ in (4) so that the prior of $(X_1, \cdots, X_r)$ is constant independent of δ_1. We have chosen to follow Besag et al. (1995) and Ghosh et al. (1998) in using the exponent q in (4).

We can formally relate (4) to (2) by noting that

$$f(Z_i|Z_{-i}) = f(Z_1, \cdots, Z_q) \Big/ \int_{-\infty}^{\infty} f(Z_1, \cdots, Z_q) dZ_i.$$

Hobert & Casella (1998) have called this type of relationship 'functionally compatible,' in contrast to one which is 'compatible,' where the integral of $f(Z_1, \cdots, Z_q)$ with respect to $(Z_1, \cdots, Z_q)$ is finite. In the remainder of this section, we will examine compatibility conditions for some important CAR(1) models.

A simple nontrivial example is given for the case $q = 2$. Suppose that $f(Z_1|Z_2)$ is $N(\rho Z_1, 1)$ and $f(Z_2|Z_1)$ is $N(\rho Z_2, 1)$. In this case $B = \binom{1 \ \ -\rho}{-\rho \ \ 1}$, and if $|\rho| < 1$ then Z is bivariate normal with mean 0 and covariance matrix $B^{-1} = (1-\rho^2)^{-1}\binom{1 \ \ \rho}{\rho \ \ 1}$. If $\rho = 1$ (the case $\rho = -1$ is similar), B is singular and the joint distribution formally becomes $f(Z_1, Z_2) \propto \exp\{-\frac{1}{2}(Z_1 - Z_2)^2\}$, an improper distribution as in (4). Thus the conditional distributions are functionally compatible. We can generalise this example to the regional effects $(Z_1, \cdots, Z_q)'$ by the following model.

Model 1A. Let $a_i = d_i$ and $\beta_{ij} = \rho/d_i$, if $j \in \Lambda_i$, and 0, if $j \notin \Lambda_i$, where $|\rho| < 1$; the adjacency property is assumed to be symmetric so that $i \in \Lambda_j$ if and only if $j \in \Lambda_i$. Then $a_i\beta_{ij} = a_j\beta_{ji}$ and B is symmetric. Moreover,

$$B = D - \rho C, \tag{5}$$

where D is the diagonal matrix with diagonal elements $d_1, \cdots, d_q$ and C is the adjacency matrix, with element $c_{ij} = 1$, if i and j are adjacent, and 0, otherwise. The conditional distribution (2) can now be written

$$f(Z_i|Z_{-i}) = \Big(\frac{d_i}{2\pi\delta_1}\Big)^{\frac{1}{2}} \exp\Big\{-\frac{d_i}{2\delta_1}(Z_i - \rho\overline{z}_i)^2\Big\},$$

where $\overline{z}_i = d_i^{-1}\sum_{j\in\Lambda_i} Z_j$.

When $\rho = 1$ in (5), we have Model 1 proposed by Besag et al. (1991), and used by Carlin & Louis (1996), Ghosh et al. (1998) and others. In this case, B cannot be positive definite since the row and column sums of B will be 0_q. When $\rho = 0$ in (5), it reduces to the case where $Z_1, \cdots, Z_n$ are independent without spatial correlation. However, the variance of Z_i is still inversely proportional to the number of neighbours d_i, an unlikely situation under independence.

To show that B is positive definite under Model 1A, we note that, for any ρ for which $|\rho| < 1$, the ith diagonal element is larger than the sum of the absolute values of all off diagonal elements in the ith row. The following lemma, a corollary of diagonal dominance (c.f. Ortega, 1987, p. 225) stated here for completeness, implies that B is positive definite. According to Besag (1974), Z has a nonsingular normal distribution.

LEMMA 1. *Let $A = (a_{ij})$ be a $q \times q$ matrix. If $a_{ii} > \sum_{j \neq i} |a_{ij}|$ for all i, then A is positive definite.*

Model 2. Let $a_i = 1$ and $\beta_{ij} = \rho$, if $j \in \Lambda_i$, and 0, if $j \notin \Lambda_i$. Here Λ_i is again the set of regions adjacent to region i. Then $\beta_{ij} = \beta_{ji}$, B is symmetric, and

$$B = I_q - \rho C, \tag{6}$$

where I_q is the $q \times q$ identity matrix and C is the adjacency matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_q$ be the ordered eigenvalues of C. As seen in the following lemma, C is neither positive definite nor negative definite, so that $\lambda_1 < 0$ and $\lambda_q > 0$.

LEMMA 2. *Let $A = (a_{ij})$ be an $q \times q$ nonzero symmetric matrix whose diagonal elements are all zero. Let λ_{min} and λ_{max} be the minimum and maximum eigenvalues of A. Then A is neither positive definite nor negative definite. That is to say,*

$$\lambda_{min} < 0 \quad and \quad \lambda_{max} > 0.$$

Proof. Since A is a nonzero matrix, there is a pair (i, j) such that $i < j$ and $a_{ij} \neq 0$. Let X be the q-dimensional vector whose ith and jth components are 1 and other components are 0. Clearly $X'AX = 2a_{ij} + a_{ii} + a_{jj} = 2a_{ij}$, since $a_{ii} = a_{jj} = 0$. If we let X be the q-dimensional vector whose ith component is 1, jth component is -1, and other components are 0, we get $X'AX = -2a_{ij}$. The result then follows. □

From Lemma 2, we know that $\lambda_1 < 0 < \lambda_q$.

THEOREM **1**. *If*

$$\lambda_1^{-1} < \rho < \lambda_q^{-1}, \tag{7}$$

then B is positive definite, and the conditional distribution (2) can be written

$$f(Z_i|Z_{-i}) = \Big(\frac{1}{2\pi\delta_1}\Big)^{\frac{1}{2}} \exp\Big\{-\frac{1}{2\delta_1}\Big(Z_i - \rho \sum_{i \in \Lambda_i} Z_j\Big)^2\Big\}. \tag{8}$$

Proof. Let , be an orthogonal matrix so that $C = \, , \, \mathrm{diag}\,(\lambda_1, \cdots, \lambda_q) \, ,'$. Then we have the decomposition

$$B = \, , \, \mathrm{diag}\,(1 - \rho\lambda_1, \cdots, 1 - \rho\lambda_q) \, ,'.$$

The eigenvalues of B are $(1 - \rho\lambda_1, \cdots, 1 - \rho\lambda_q)$. From (7), they are all positive. This proves the first part. The second part follows immediately. □

If $\rho = \lambda_1^{-1}$ or λ_q^{-1} then B will be singular and, by our previous result, the joint distribution of Z will be improper. Moreover, if $\rho < \lambda_1^{-1}$ or $\rho > \lambda_q^{-1}$, B is no longer nonnegative definite.

Note that (8) is the model used in Clayton & Kaldor (1987). Furthermore, if $\rho = 0$, then the model reduces to the case where $Z_1, \cdots, Z_n$ are independently and identically $N(0, \delta_1)$ distributed.

3. Existence of the Posteriors

3.1. *Propriety of the Posterior for a Linear Mixed Model*

Here and in §3.2 we examine whether or not the posterior distribution remains proper when both θ and Z have noninformative priors.

Consider a general linear mixed model

$$V_i = x'_{1i}\theta + x'_{2i}Z + e_i, \tag{9}$$

where x_{1i} and x_{2i} are vectors of fixed constants, θ is a vector of fixed effects, Z is a vector of random effects, and the e_i are independently and identically distributed normal errors with mean 0 and variance δ_0. For given $\delta_1 > 0, \delta_0 > 0$ and a nonnegative definite matrix B, we assume that θ, $Z|\delta_1$ and $e|\delta_0$ are independent with

$$w(\theta) \equiv 1, \tag{10}$$

and $Z|\delta_1$ has the partially informative normal density (4). Assume that the variance components δ_0 and δ_1 are a priori independent and that δ_i has an inverse gamma distribution (a_i, b_i) with density

$$g_i(\delta_i) \propto \delta_i^{-(a_i+1)} \exp(-b_i/\delta_i). \tag{11}$$

Let $V = (V_1, \cdots, V_n)'$ be the vector of n observations, and let $X_1 = (x_{11}, \cdots, x_{1n})'$ and $X_2 = (x_{21}, \cdots, x_{2n})'$ be the $n \times p$ and $n \times q$ design matrices. Denote the usual least squares estimator of (θ', Z') by $(\hat{\theta}', \hat{Z}')' = (X'X)^{-}X'V$, where $X = (X_1, X_2)$ and $(X'X)^{-}$ is a generalised inverse of $X'X$. Let the sum of squared errors be $SSE = V'\{I_n - X(X'X)^{-}X'\}V$, which is invariant for any choice of $(X'X)^{-}$.

Theorem 2. *Consider the linear mixed model (9) with prior distribution given by (4), (10), and (11), where B is nonnegative definite. Assume the following conditions.*

(a) rank $(X_1) = p$ *and rank* $(X_2' R_1 X_2 + B) = q$, *where* $R_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$;

(b) $a_1 > 0$ *and* $b_1 > 0$;

(c) $n - p - q + 2a_0 > 0$ *and* $SSE + 2b_0 > 0$.

Then the joint posterior distribution of $(\theta, Z, \delta_0, \delta_1)$ *given* V *is proper.*

Proof. First, the joint posterior density of $(\theta, Z, \delta_0, \delta_1)$ is proportional to

$$G = \delta_0^{-\frac{1}{2}n}\delta_1^{-\frac{1}{2}q}\exp\Big\{-\frac{(V-X_1\theta-X_2Z)'(V-X_1\theta-X_2Z)}{2\delta_0} - \frac{Z'BZ}{2\delta_1}\Big\}\prod_{i=0}^{1} g_i(\delta_i). \quad (12)$$

Since $(V - X_1\theta - X_2Z)'(V - X_1\theta - X_2Z)$ equals $SSE + (\theta - \hat{\theta} - C_1)'X_1'X_1(\theta - \hat{\theta} - C_1)$ + $(Z - \hat{Z})'X_2'R_1X_2(Z - \hat{Z})$, where $C_1 = (X_1'X_1)^{-1}X_1'X_2(\hat{Z} - Z)$, we integrate G with respect to θ and get

$$\int G d\theta = \frac{(2\pi)^{\frac{p}{2}}|X_1'X_1|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p)}\delta_1^{\frac{1}{2}q}}\exp\Big\{-\frac{SSE}{2\delta_0} - \frac{(Z-\hat{Z})'X_2'R_1X_2(Z-\hat{Z})}{2\delta_0} - \frac{Z'BZ}{2\delta_1}\Big\}\prod_{i=0}^{1} g_i(\delta_i). \quad (13)$$

Define $R_2 = \delta_0^{-1}X_2'R_1X_2 + \delta_1^{-1}B$. For any fixed $\delta_0, \delta_1 > 0$, R_2^{-1} exists by Assumption (a). Let $C_2 = \delta_0^{-1}R_2^{-1}X_2'R_1X_2\hat{Z}$ and $R_3 = X_2'R_1X_2 - \delta_0^{-1}X_2'R_1X_2R_2^{-1}X_2'R_1X_2$. Then

$$\frac{(Z-\hat{Z})'X_2'R_1X_2(Z-\hat{Z})}{\delta_0} + \frac{Z'BZ}{\delta_1} = (Z - C_2)'R_2(Z - C_2) + \frac{\hat{Z}'R_3\hat{Z}}{\delta_0}.$$

Integrating G with respect to θ and Z, we get

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^p} G d\theta dZ = \frac{(2\pi)^{\frac{1}{2}(p+q)}|X_1'X_1|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p)}\delta_1^{\frac{1}{2}q}|R_2|^{\frac{1}{2}}}\exp\Big(-\frac{SSE + \hat{Z}'R_3\hat{Z}}{2\delta_0}\Big)\prod_{i=0}^{1} g_i(\delta_i). \quad (14)$$

Since R_3 is nonnegative definite, $\hat{Z}'R_3\hat{Z} \geq 0$. Note that

$$|R_2|^{-\frac{1}{2}} \leq \{\min(\delta_0^{-1}, \delta_1^{-1})^q|X_2'R_1X_2 + B|\}^{-\frac{1}{2}} < \Big(\delta_0^{\frac{1}{2}q} + \delta_1^{\frac{1}{2}q}\Big)\Big|X_2'R_1X_2 + B\Big|^{-\frac{1}{2}}.$$

We have

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^p} G d\theta dZ \leq (2\pi)^{\frac{1}{2}(p+q)}\Big|X_1'X_1\Big|^{-\frac{1}{2}}\Big|X_2'R_1X_2 + B\Big|^{-\frac{1}{2}}(J_1 + J_2), \tag{15}$$

where

$$\begin{aligned} J_1 &= \frac{1}{\delta_0^{\frac{1}{2}(n-p-q)+a_0+1}\delta_1^{\frac{1}{2}q+a_1+1}}\exp\Big(-\frac{2b_0 + SSE}{2\delta_0} - \frac{b_1}{\delta_1}\Big), \\ J_2 &= \frac{1}{\delta_0^{\frac{1}{2}(n-p)+a_0+1}\delta_1^{a_1+1}}\exp\Big(-\frac{2b_0 + SSE}{2\delta_0} - \frac{b_1}{\delta_1}\Big). \end{aligned}$$

From Assumption (c), $\frac{1}{2}q + a_0 > 0$. The Integral of J_1 with respect to (δ_0, δ_1) is finite. Also, Assumption (c) implies that $n - p + 2a_0 > 0$ and the integral of J_2 with respect to (δ_0, δ_1) is finite. The result then follows. □

Assumption (a) is equivalent to the assumption that the rank of $\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 + B \end{pmatrix}$ equals $p + q$. Also, Assumption (a) is satisfied by either of the following conditions:

(a1) the rank of X is $p + q$;

(a2) the rank of X_1 is p and rank of B is q.

Corollary 1. *Consider the linear mixed model (9), whose prior distribution is given by (10) and suppose δ_i follows the prior (11).*

(a) If Condition (a1) holds, the posterior distribution of $(\theta, Z, \delta_0, \delta_1)$ given V exists for any $q \times q$ nonnegative definite matrix B.

(b) If Condition (a2) holds, the posterior distribution of $(\theta, Z, \delta_0, \delta_1)$ given V exists for any $n \times q$ design matrix X_2.

Proof. From the same argument as the proof of Theorem 2, (14) holds. For Part (a), we know that $|R_2| \geq \delta_0^{-q}|X_2'R_1X_2|$. Then

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^p} G d\theta dZ \leq \frac{(2\pi)^{\frac{1}{2}(p+q)}|X_1'X_1|^{-\frac{1}{2}}|X_2'R_1X_2|^{-\frac{1}{2}}}{\delta_0^{\frac{1}{2}(n-p-q)+a_0+1}\delta_1^{\frac{1}{2}q+a_1+1}}\exp\Big(-\frac{2b_0 + SSE}{2\delta_0} - \frac{b_1}{\delta_1}\Big).$$

The result is immediate. Part (b) follows from the fact that $|R_2| \geq \delta_1^{-q}|B|$. $\square$

COROLLARY 2. *(a) Assume that* $(X_1'X_1)^{-1}$ *exists. Then the posterior distribution of* $(\theta, Z, \delta_0, \delta_1)$ *given* V *is proper under Model 1A.*
(b) Assume that $(X'X)^{-1}$ *exists. Then the posterior distribution of* $(\theta, Z, \delta_0, \delta_1)$ *given* V *exists under Model 1 or Model 2 when* $\rho = 1/\lambda_1$ *or* $\rho = 1/\lambda_q$.

This corollary follows from Corollary 1. A related result has appeared in Hobert & Casella (1996) for the special case where rank $(X_1) = p$ and B is a diagonal matrix with unknown elements. In this case, rank $(R_2) = q$. If, however, B is not positive definite, the posterior distribution may not be proper, as shown in the following theorem.

THEOREM **3**. *Assume that rank* $(X_1) = p$ *and rank* $(X_2'R_1X_2 + B) < q$. *For any priors on* δ_0 *and* δ_1, *the posterior distribution of* $(\theta, Z, \delta_0, \delta_1)$ *given* V *does not exist.*

Proof. Since rank $(X_1) = p$, (13) still holds. Since rank $(X_2'R_1X_2 + B) < q$, for any fixed δ_0 and δ_1,

$$\int_{\mathbb{R}^q} \exp\Big\{-\frac{(Z-\hat{Z})'X_2'R_1X_2(Z-\hat{Z})}{2\delta_0} - \frac{Z'BZ}{2\delta_1}\Big\} dZ = \infty.$$

This proves the result. $\square$

We note that if B is positive definite our model reduces to that of Hobert & Casella (1996), who provide necessary and sufficient conditions for the posterior distribution to be proper. Our result is an interesting extension to the situation where B is not positive definite, as is often the case with the CAR(1) model.

One implication of our results is that, among the assumptions of Theorem 2, rank $(X_2'R_1X_2 + B) = q$ is both necessary and sufficient for the posterior distribution of $(\theta, Z, \delta_0, \delta_1)$ given V to be proper. We illustrate the point by a simple example for which the posterior distribution may or may not be proper.

Example 1. Consider a balanced one-way analysis of variance,

$$Y_{ij} = \theta + Z_i + e_{ij}, \quad i = 1, \cdots, q; \quad j = 1, \cdots, m.$$

Here θ is the fixed effect, $Z = (Z_1, \cdots, Z_q)'$ is the random effect, and the e_{ij} are independently and identically $N(0, \delta_0)$ distributed. This is a special case of (9) with $X_1 = 1_{qm}$ (a vector of ones) and $X_2 = I_q \otimes 1_m$. Assume the prior (10). Clearly, $X_2'R_1X_2 = r(I_q - \frac{1}{q}1_q1_q')$.

Case 1. $B = I_q$. For a proper prior on δ_i given in (11), the posterior of $(\theta, Z, \delta_0, \delta_1)$ is proper.

Case 2. $B = D - C$, a special case of Model 1A. Since $1_q'B1_q = 0$, we know that the rank of $X_2'R_1X_2 + B$ is $q - 1$. From Theorem 3, we know that for any proper prior on δ_i, the joint posterior of $(\theta, Z, \delta_0, \delta_1)$ does not exist.

Case 3. $B = I_q - \rho C$, with the limiting case $\rho = 1/\lambda_1$ or $\rho = 1/\lambda_q$. This case is quite complicated. We study a simple case with $q = 3$. Assume that $C = 1_31_3' - I_3$, whose eigenvalues are $-1, -1, 2$. If $\rho = 1/\lambda_1 = -1$, then the rank of $X_2'R_1X_2 + B$ is $q = 3$. Then the posterior distribution might be proper. On the other hand, if $\rho = 1/\lambda_3 = 0.5$ then $1_q'B1_q = 0$, and the posterior distribution does not exist. We now change C to

$$C = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

whose eigenvalues are $-\sqrt{2}, 0, \sqrt{2}$. By numerical calculation, when either $\rho = -1/\sqrt{2}$ or $1/\sqrt{2}$, the rank of $X_2' R_1 X_2 + B$ is $q = 3$. Consequently, the posterior distribution would be proper for any proper prior of δ_i.

3.2. *Propriety of the posterior for a generalised linear mixed model*

Consider the hierarchical model where $Y_1, Y_2, \cdots, Y_N$ are conditionally independent given parameters $V = (V_1, V_2, \cdots, V_N)$, and V_i follows the hierarchical prior defined by (9)–(11). Let $f_i(Y_i|V_i)$ be the distribution of Y_i given V_i.

THEOREM 4. *Suppose there exist* $Y_{i_1}, \cdots, Y_{i_n}$ $(1 \le i_1 < \cdots < i_n \le N; p + q \le n \le N)$ *such that*

$$\int f_j(Y_j|V_j)dV_j < \infty,\ j \in \{i_1, \cdots, i_n\} \quad \textit{and} \quad f_j(Y_j|V_j) \le M,\ j \notin \{i_1, \cdots, i_n\}, \quad (16)$$

for some constant M*, the corresponding design matrix* $X_1^* = (x_{1,i_1}, \cdots, x_{1,i_n})'$ *has full rank, and* $X_2^* = (x_{2,i_1}, \cdots, x_{2,i_n})'$ *has the same rank as the matrix* $(x_{2,1}, \cdots, x_{2,N})'$. *For any proper prior on* δ_i*, the posterior distribution of* $(V, \theta, Z, \delta_0, \delta_1)$ *given* $Y = (Y_1, \cdots, Y_N)$ *exists.*

Proof. Without loss of generality, assume that $i_j = j$, $j = 1, \cdots, n$. Let $V^* = (V_1, \cdots, V_n)$. The posterior density of $(V, \theta, Z, \delta_0, \delta_1)$ given Y is

$$p(V, \theta, Z, \delta_0, \delta_1|Y) \propto \prod_{i=1}^{N} f_i(Y_i|V_i)\delta_0^{-\frac{1}{2}(N-n)} \prod_{i=n+1}^{N} \exp\Big[-\frac{1}{2\delta_0}(V_i - x_{1i}'\theta - x_{2i}'Z)^2\Big]\, G^*.$$

Here G^* is defined by (12) with X_1 and X_2 being replaced by X_1^* and X_2^*, respectively. Using the second inequality in (16) and integrating with respect to $V_{n+1}, \cdots, V_N$, we obtain

$$p(V^*, \theta, Z, \delta_0, \delta_1 | Y) \propto \prod_{i=1}^{n} f_i(Y_i|V_i)\, G^*.$$

From the same argument as in the proof of Theorem 2, we obtain

$$p(V^*, \delta_0, \delta_1 | Y) \propto \prod_{i=1}^{n} f_i(Y_i|V_i) \frac{\delta_0^{\frac{1}{2}q} + \delta_1^{\frac{1}{2}q}}{\delta_0^{\frac{1}{2}(n-p)+a_0+1}\, \delta_1^{\frac{1}{2}q+a_1+1}} \exp\Big\{-\frac{2b_0}{2\delta_0} - \frac{b_1}{\delta_1}\Big\}.$$

Clearly, $p(V^*|Y) \propto \prod_{i=1}^{n} f_i(Y_i|V_i)$, which is proper by the first inequality in (16). □

In the original example given in §1, $f_i(Y_i|V_i)$ is Poisson with mean $\mu_i = m_i e^{V_i}$, where V_i has the linear structure (9). Then, $f_i(Y_i|V_i)$ is bounded for any $Y_i \geq 0$ and

$$\int f_i(Y_i|V_i) dV_i = \int_0^\infty \frac{e^{-\mu_i} \mu_i^{Y_i - 1}}{Y_i!} d\mu_i,$$

which is finite for $Y_i > 0$. For another example, let $Y_i|V_i$ be binomial with parameter m_i and $p_i = e^{V_i}/(e^{V_i}+1)$, where V_i has prior (9). Then $f_i(Y_i|V_i)$ is bounded in V_i for any $0 \leq Y_i \leq m_i$, and

$$\int_{-\infty}^{\infty} f_i(Y_i|V_i) dV_i \;=\; \int_{-\infty}^{\infty} \frac{e^{V_i Y_i}}{(e^{V_i}+1)^{m_i}} dV_i \;=\; \int_0^1 p_i^{Y_i - 1} (1-p_i)^{m_i - Y_i - 1} dp_i,$$

which is finite if and only if $0 < Y_i < m_i$.

Acknowledgement

Sun's research was supported by a grant from the National Security Agency. Tsutakawa's research was supported by a Research Board award from the University of Missouri-System. Speckman's research was partially supported by a grant from the National Science Foundation. Research was also supported in part by the US Federal Highway Administration, under subcontract E77690017-3Y from Los Alamos National Laboratory to the National Institute of Statistical Sciences. The authors gratefully acknowledge the very constructive comments of the editor, an associate editor and two referees.

References

Bernardinelli, L. & Montomoli, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statist. in Med.* **11**, 983-1007.

Bernardinelli, L., Clayton, D. & Montomoli, C. (1995). Bayesian estimates of disease maps: how important are priors? *Statist. in Med.* **14**, 2411-31.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc.* B **36**, 192-236.

Besag, J., Green, P., Higden, D., & Mengersen, K. (1995). Bayesian computation and Stochastic Systems (with Discussion). *Statist. Sciences* **10**, 3-66.

Besag, J., York, J. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with Discussion). *Ann. Inst. Statist. Math.* **43**, 1-59.

Carlin, B.P. & Louis, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis.* London: Chapman and Hall.

Clayton, D. & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671-81.

Cressie, N & Chan, N.H. (1989). Spatial modeling of regional variables. *J. Am.*

Statist. Assoc. **84**, 393-401.

GELFAND, A.E. & SMITH, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398-409.

GHOSH, M., NATARAJAN, K., STROUD, T.W.F. & CARLIN, B.P. (1998). Generalized linear models for small area estimation. *J. Am. Statist. Assoc.* **93**, 273-82.

GILKS, W.R. & WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* 41, 337-48.

HOBERT, J.P. & CASELLA, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Am. Statist. Assoc.* **91**, 1461-73.

HOBERT, J.P. & CASELLA, G. (1998). Functional compatibility, Markov chains and Gibbs sampling with improper posteriors. *J. Comp. and Graph. Statist.*, **7**, 42–66.

MARSHALL, R.J. (1991). A review of methods for the statistical analysis of spatial patterns of disease. *J. R. Statist. Soc.* A **154**, 421-41.

ORTEGA, J. M. (1987). *Matrix Theory.* New York: Plenum Press.

TSUTAKAWA, R.K. (1988). Mixed model for analysing geographic variability in mortality rates. *J. Am. Statist. Assoc.* **83**, 117-30.

WALLER, L.A., CARLIN, B.P., XIA, H. & GELFAND, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates. *J. Am. Statist. Assoc.* **92**, 607-17.