NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

Asymptotic Normality of Posterior Distributions for Generalized Linear Mixed Models

by

Hossein Baghishani and Mohsen Mohammadzadeh

PREPRINT
STATISTICS NO. 11/2010

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2010/S11-2010.pdf
Hossein Baghishani has homepage: http://www.math.ntnu.no/∼baghisha
E-mail: baghisha@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Asymptotic Normality of Posterior Distributions for Generalized Linear Mixed Models

Hossein Baghishani[,∗], Mohsen Mohammadzadeh

[a]*Department of Statistics, Tarbiat Modares University, Tehran, Iran*

Abstract

Generalized linear mixed models (GLMMs) are widely used in the analysis of non Gaussian repeated measurements such as clustered proportion and count data. The most commonly used approach for inference in these models is based on the Bayesian methods and Markov chain Monte Carlo (MCMC) sampling. However, the validity of Bayesian inferences can be seriously affected by the assumed model for the random effects. We establish the asymptotic normality of joint posterior distribution of the parameters and the random effects of a GLMM by the modification of Stein's Identity. We show that while incorrect assumptions on the random effects can lead to substantial bias in the estimates of the parameters, the assumed model for the random effects, under some regularity conditions, does not affect the asymptotic normality of joint posterior distribution. This motivates use of the approximate normal distributions for sensitivity analysis of the random effects distribution. We also illustrate the approximate normal distribution performs reasonable using both real and simulated data. This creates a primary alternative to MCMC sampling and avoids a wide range of problems for MCMC algorithms in terms of convergence and computational time.

Keywords: Asymptotic normality, clustered data, generalized linear mixed models, misspecification, posterior distribution, Stein's Identity.

1. Introduction

Discrete longitudinal and clustered data are common in many sciences such as biology, epidemiology and medicine. A popular and flexible approach to handle this type of data is the GLMM. This

[∗]To whom correspondence should be addressed.

Email addresses: hbaghishani@modares.ac.ir (Hossein Baghishani), mohsen_m@modares.ac.ir (Mohsen Mohammadzadeh)

Preprint submitted to Elsevier *July 6, 2010*

class of models are useful for modeling the dependence among observations inherent in longitudinal or clustered data (Breslow and Clayton, 1993). Statistical inferences in such models have been the subject of a great deal of research over two past decades. Both frequentist and Bayesian methods have been developed in GLMMs (McCulloch, 1997). In frequentist methods, in general, model fitting and inference are based on the marginal likelihood. Computing, and then maximizing, the marginal likelihood in these models generally involves numerical integration of high-dimensions and usually is prohibitive.

Due to the advances in computation, the most commonly used approach for inference in these models is based on the Bayesian methods, especially MCMC algorithms. However, the validity of Bayesian inferences can be greatly affected by the assumed model for the random effects. It is standard to assume that the random effects have a normal distribution. But, the normality assumption may be unrealistic in some applications. In general, erroneous distribution assumption for the random effects has unfavorable influence on the inferences, see e.g. Heagerty and Kurland (2001), Agresti *et al.* (2004) and Litiere *et al.* (2007).

Bayesian inferences, furthermore, comes with a wide range of problems in terms of convergence and computational time in GLMMs, especially in situations where the sample size is large. In such situations, benefit of the large sample theory and asymptotic posteriors which work well, can be useful. It is the main contribution of this paper to establish the asymptotic normality of joint posterior distribution of the parameters and the random effects of a GLMM by the modification of Stein's Identity (Weng and Tsai, 2008). It is also the purpose of the paper to show that the assumed model for the random effects, under some conditions, does not affect the asymptotic normality of joint posterior distribution. This motivates use of the approximate normal distributions for sensitivity analysis of the random effects distribution. However, we show that, through simulation study, the model misspecification can lead to substantial bias in the estimates of the parameters. We also illustrate the approximate normal distribution performs reasonable even for moderate sample sizes by simulated and real data examples. This can create a primary alternative to MCMC sampling and avoids above mentioned problems.

The asymptotic posterior normality generally is established by many authors under different conditions: e.g. LeCam (1953), Walker (1969), and Johnson (1970) for i.i.d. random variables and Heyde and Johnstone (1979), Chen (1985), Sweeting and Adekola (1987) and Sweeting (1992) for stochastic processes. Recently, Weng (2003) proposed an alternative method for posterior normality of stochastic processes in the one parameter cases. Weng and Tsai (2008) extended Weng's method to multiparameter problems. Their method uses of an adequate transformation Z, then for any bounded measurable function h, a version of Stein's Identity (Woodroofe, 1989) is employed to separate the remainder terms of the posterior expectations of $h(Z)$ so that the posterior normality becomes more lucid and can be easily established. We use the proposed method by Weng and Tsai (2008) to establish the asymptotic normality of joint posterior of the model parameters and the random effects in a GLMM.

The paper is structured as follows. In the next section, we give a brief description of GLMMs and some needed definitions. The main result for establishing the asymptotic posterior normality of GLMMs is provided in Section 3. Next, in Section 4, the theoretical results are illustrated on simulated spatial data with skew normal latent random effects and on Rongelap data of radionuclide counts. We conclude with a discussion in Section 5. Thereafter follows three technical appendices with introducing the modified Stein's Identity (A), regularity conditions needed for establishing the asymptotic posterior normality of GLMMs (B), and proof of the main theorem (C).

2. Preliminaries

2.1. The Generalized Linear Mixed Model

On the basis of the generalized linear models (GLMs), the GLMM assumes that the responses are independent conditional on the random effects and are distributed according to a member of the exponential family. Consider a clustered data set, in which repeated measures of a response variable are taken on a random sample of m clusters. Consider the response vectors $\boldsymbol{y}_i = (y_{i1}, \dots, y_{im_i})^T$, $i = 1, \dots, m$. Let $n = \sum_{i=1}^{m} m_i$ be the total sample size. Conditional on $q \times 1$ vector of unobservable cluster-specific random effects $\boldsymbol{u}_i = (u_{i1}, \dots, u_{iq})^T$, these data are distributed according to a member of the exponential family:

$$f(y_{ij}|\boldsymbol{u}_i, \boldsymbol{\beta}) = \exp\{y_{ij}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{v}_{ij}^T\boldsymbol{u}_i) - a(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{v}_{ij}^T\boldsymbol{u}_i) + c(y_{ij}))\},$$

for $i = 1, \dots, m; j = 1, \dots, m_i$, in which $\boldsymbol{x}_{ij}$ and $\boldsymbol{v}_{ij}$, are the corresponding p- and q-dimensional covariate vectors associated with the fixed effects and the random effects respectively, $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters, and $a(\cdot)$ and $c(\cdot)$ are specific functions.

Here $\tau_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{v}_{ij}^T\boldsymbol{u}_i$ is the canonical parameter. Let $\mu_{ij} = E[Y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i] = a^{'}(\tau_{ij})$ with $g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{v}_{ij}^T\boldsymbol{u}_i$, where $g(.)$ is a monotonic link function. Furthermore, assume $\boldsymbol{u}_i$ comes from a distribution $G(\cdot, \boldsymbol{\theta})$ in which $\boldsymbol{\theta}$ is usually the dependency parameter vector of the model.

Let $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ be the vector of model parameters where $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, an open subset of $\Re^d$. Then, the marginal likelihood function of the GLMM will be

$$L(\boldsymbol{\psi}; \boldsymbol{y}) = \prod_{i=1}^{m} L_i(\boldsymbol{\psi}; \boldsymbol{y}_i), \tag{1}$$

where, $L_i(\boldsymbol{\psi}; \boldsymbol{y}_i) = \displaystyle\int \ldots \int \prod_{j=1}^{m_i} f(y_{ij}|\boldsymbol{u}_i, \boldsymbol{\beta}) G(\boldsymbol{u}_i, \boldsymbol{\theta}) d\boldsymbol{u}_i$. Here the calculation of the marginal likelihood function (1) nearly always involves intractable integrals, which is the main problem for carrying out likelihood based statistical inferences.

Now let $\pi(\boldsymbol{\psi})$ denote the joint prior density of the parameters. Then the joint posterior density is defined by

$$\pi(\boldsymbol{\psi}, \boldsymbol{u}|\boldsymbol{y}) = \frac{\prod_{i=1}^{m} f(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\psi}) G(\boldsymbol{u}_i, \boldsymbol{\theta}) \pi(\boldsymbol{\psi})}{\int \prod_{i=1}^{m} f(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\psi}) G(\boldsymbol{u}_i, \boldsymbol{\theta}) \pi(\boldsymbol{\psi}) d\boldsymbol{u} d\boldsymbol{\psi}},$$

which is not available in closed form because of the same intractable integrals that cause trouble in the likelihood function. Because of the usefulness and easy implementation of the MCMC algorithms for sampling from this joint posterior density the most commonly used approach for inference in these models is based on Bayesian methods.

2.2. Definitions and Preliminary Results

Some notations and calculations are needed in the sequel. Considering (1) we can write

$$\pi(\boldsymbol{\psi}, \boldsymbol{u}|\boldsymbol{y}) = \pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\psi}) \pi(\boldsymbol{\psi}|\boldsymbol{y}) = \pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\psi}) L(\boldsymbol{\psi}; \boldsymbol{y}) \pi(\boldsymbol{\psi}).$$

Hence,

$$\log\left(\frac{\pi(\boldsymbol{\psi}, \boldsymbol{u}|\boldsymbol{y})}{\pi(\boldsymbol{\psi})}\right) = \log \pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\psi}) + \log L(\boldsymbol{\psi}; \boldsymbol{y}) = g_n(\boldsymbol{u}) + \ell_n(\boldsymbol{\psi}).$$

Assume that the functions $g_n(\boldsymbol{u})$ and $\ell_n(\boldsymbol{\psi})$ are twice continuously differentiable with respect to $\boldsymbol{u}$ and $\boldsymbol{\psi}$ respectively. Let $\nabla g_n(\boldsymbol{u})$ and $\nabla \ell_n(\boldsymbol{\psi})$ be the vectors of first-order partial derivatives, and $\nabla^2 g_n(\boldsymbol{u})$ and $\nabla^2 \ell_n(\boldsymbol{\psi})$ be the matrices of second-order partial derivatives with respect to $\boldsymbol{u}$ and $\boldsymbol{\psi}$

respectively. Here and subsequently, let $\hat{\boldsymbol{\psi}}_n$ be the MLE of $\boldsymbol{\psi}$ satisfying $\nabla \ell_n(\boldsymbol{\psi}) = 0$ and $\tilde{\boldsymbol{u}}_n$ be the mode of $g_n(\boldsymbol{u})$.

To facilitate asymptotic theory arguments, whenever the $\hat{\boldsymbol{\psi}}_n$ and $\tilde{\boldsymbol{u}}_n$ exist and $-\nabla^2 \ell_n(\boldsymbol{\psi})$ and $-\nabla^2 g_n(\boldsymbol{u})$ are positive definite, we define F_n, G_n, $\boldsymbol{z}_n$ and $\boldsymbol{w}_n$ as follow:

$$\begin{aligned} F_n^T F_n &= -\nabla^2 \ell_n(\hat{\boldsymbol{\psi}}_n), \;\; \boldsymbol{z}_n = F_n(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \quad (2) \\ G_n^T G_n &= -\nabla^2 g_n(\tilde{\boldsymbol{u}}_n), \;\; \boldsymbol{w}_n = G_n(\boldsymbol{u} - \tilde{\boldsymbol{u}}_n), \quad (3) \end{aligned}$$

otherwise define them arbitrarily, in a measurable way. Then the joint posterior density of $(\boldsymbol{z}_n, \boldsymbol{w}_n)$ is given by

$$\pi_n(\boldsymbol{z}_n, \boldsymbol{w}_n | \boldsymbol{y}) \propto \pi_n(\boldsymbol{\psi}(\boldsymbol{z}_n), \boldsymbol{u}(\boldsymbol{w}_n) | \boldsymbol{y}) \propto e^{\ell_n(\boldsymbol{\psi}) - \ell_n(\hat{\boldsymbol{\psi}}_n)} e^{g_n(\boldsymbol{u}) - g_n(\tilde{\boldsymbol{u}}_n)} \pi(\boldsymbol{\psi}). \quad (4)$$

Let $\boldsymbol{\psi}_0$ and $\boldsymbol{u}_0$ denote the true underlying parameter and the true realization of random effects respectively. Let also P_n^c and E_n^c denote the conditional probability and expectation given data $\boldsymbol{y}$. In what follows, unless indicated otherwise, all probability statements are with respect to the true underlying probability distribution. Then we want to show

$$P_n^c((\boldsymbol{z}_n^T, \boldsymbol{w}_n^T)^T \in B) \longrightarrow \Phi_{d+q}(B),$$

in probability as $n \to \infty$, where B is any Borel set in $\Re^{d+q}$ and Φ_{d+q} is the standard $d+q$-variate Gaussian distribution.

To conduct the posterior distribution in a form suitable for Stein's Identity (see Appendix A), we need following calculations. For converting $\ell_n(\boldsymbol{\psi})$ into a form close to normal, we first take a Taylor's expansion of $\ell_n(\boldsymbol{\psi})$ at $\hat{\boldsymbol{\psi}}_n$:

$$\ell_n(\boldsymbol{\psi}) = \ell_n(\hat{\boldsymbol{\psi}}_n) + \frac{1}{2}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n)^T \nabla^2 \ell_n(\boldsymbol{\psi}^*)(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n),$$

where $\boldsymbol{\psi}^*$ lies between $\boldsymbol{\psi}$ and $\hat{\boldsymbol{\psi}}_n$. Let

$$k_n(\boldsymbol{\psi}) = -\frac{1}{2}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n)^T [\nabla^2 \ell_n(\hat{\boldsymbol{\psi}}_n) - \nabla^2 \ell_n(\boldsymbol{\psi}^*)](\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n). \quad (5)$$

Thus,

$$\ell_n(\boldsymbol{\psi}) = \ell_n(\hat{\boldsymbol{\psi}}_n) - \frac{1}{2}\|\boldsymbol{z}_n\|^2 + k_n(\boldsymbol{\psi}). \quad (6)$$

With parallel arguments, we have

$$g_n(\boldsymbol{u}) = g_n(\tilde{\boldsymbol{u}}_n) - \frac{1}{2}\|\boldsymbol{w}_n\|^2 + l_n(\boldsymbol{u}), \quad (7)$$

where

$$l_n(\boldsymbol{u}) = -\frac{1}{2}(\boldsymbol{u}-\tilde{\boldsymbol{u}}_n)^T[\nabla^2 g_n(\tilde{\boldsymbol{u}}_n) - \nabla^2 g_n(\boldsymbol{u}^*)](\boldsymbol{u}-\tilde{\boldsymbol{u}}_n), \tag{8}$$

and $\boldsymbol{u}^*$ lies between $\boldsymbol{u}$ and $\tilde{\boldsymbol{u}}_n$. Therefor, we can rewrite the posterior (4) as

$$\pi_n(\boldsymbol{z}_n, \boldsymbol{w}_n|\boldsymbol{y}) \propto \phi_d(\boldsymbol{z}_n)\phi_q(\boldsymbol{w}_n)f_n(\boldsymbol{z}_n, \boldsymbol{w}_n), \tag{9}$$

where $f_n(\boldsymbol{z}_n, \boldsymbol{w}_n) = \exp\{k_n(\boldsymbol{\psi})+l_n(\boldsymbol{u})\}\pi(\boldsymbol{\psi}(\boldsymbol{z}_n))$ and $\phi_t(\cdot)$ displays the standard t-variate Gaussian density.

Note that from (6) and (7) we have

$$\begin{aligned}
\nabla k_n(\boldsymbol{\psi}) &= \nabla \ell_n(\boldsymbol{\psi}) - \nabla^2\ell_n(\hat{\boldsymbol{\psi}}_n)(\boldsymbol{\psi}-\hat{\boldsymbol{\psi}}_n),\\
\nabla l_n(\boldsymbol{u}) &= \nabla g_n(\boldsymbol{u}) - \nabla^2 g_n(\tilde{\boldsymbol{u}}_n)(\boldsymbol{u}-\tilde{\boldsymbol{u}}_n).
\end{aligned}$$

These imply that

$$\begin{aligned}
\nabla k_n(\boldsymbol{\psi}) &= [(\frac{\partial^2\ell_n}{\partial\psi_i\partial\psi_j}(\psi^{*ij})) - \nabla^2\ell_n(\hat{\boldsymbol{\psi}}_n)](\boldsymbol{\psi}-\hat{\boldsymbol{\psi}}_n),\\
\nabla l_n(\boldsymbol{u}) &= [(\frac{\partial^2 g_n}{\partial u_i\partial u_j}(u^{*ij})) - \nabla^2 g_n(\tilde{\boldsymbol{u}}_n)](\boldsymbol{u}-\tilde{\boldsymbol{u}}_n),
\end{aligned}$$

where $(\partial^2\ell_n/\partial\psi_i\partial\psi_j)(\eta^{ij})$ and $(\partial^2 g_n/\partial u_i\partial u_j)(\eta^{ij})$ denote the Hessian matrices of $\ell_n(\boldsymbol{\psi})$ and $g_n(\boldsymbol{u})$ with its (i,j)-component evaluated at η^{ij} respectively. Therefore,

$$(F_n^T)^{-1}\nabla k_n(\boldsymbol{\psi}) = \left\{I_d - (F_n^T)^{-1}[-(\frac{\partial^2\ell_n}{\partial\psi_i\partial\psi_j}(\psi^{*ij}))]F_n^{-1}\right\}\boldsymbol{z}_n, \tag{10}$$

$$(G_n^T)^{-1}\nabla l_n(\boldsymbol{u}) = \left\{I_q - (G_n^T)^{-1}[-(\frac{\partial^2 g_n}{\partial u_i\partial u_j}(u^{*ij}))]G_n^{-1}\right\}\boldsymbol{w}_n. \tag{11}$$

Suppose also $\nabla_{\boldsymbol{z}_n} f(\boldsymbol{z}_n, \boldsymbol{w}_n)$ and $\nabla_{\boldsymbol{w}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)$ denote the partial derivatives of $f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)$ with respect to $\boldsymbol{z}_n$ and $\boldsymbol{w}_n$ respectively and $\nabla\pi(\boldsymbol{\psi})$ the derivative of prior distribution with respect to $\boldsymbol{\psi}$. Hence,

$$\frac{\nabla_{\boldsymbol{z}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}{f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)} = (F_n^T)^{-1}\left\{\frac{\nabla\pi(\boldsymbol{\psi})}{\pi(\boldsymbol{\psi})} + \nabla k_n(\boldsymbol{\psi})\right\}, \tag{12}$$

$$\frac{\nabla_{\boldsymbol{w}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}{f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)} = (G_n^T)^{-1}\nabla l_n(\boldsymbol{u}). \tag{13}$$

Let also $D = D_1 \cup D_2$, in which $D_1 = \{\nabla\ell_n(\hat{\boldsymbol{\psi}}_n) = 0, -\nabla^2\ell_n(\hat{\boldsymbol{\psi}}_n) > 0\}$ and $D_2 = \{\nabla g_n(\tilde{\boldsymbol{u}}_n) = 0, -\nabla^2 g_n(\tilde{\boldsymbol{u}}_n) > 0\}$. Here $A > 0$ means that the matrix A is positive definite. The following lemma and corollary are essential for establishing the asymptotic normality of posterior distribution of a GLMM.

Lemma 1. *Let s be a nonnegative integer. Suppose that $\pi(\boldsymbol{\psi})$ satisfies $(A1)$ and $(A2)$. Then for all $h : \Re^{d+q} \longrightarrow \Re$ where $h \in H_s$,*

$$E_n^c[h(\boldsymbol{z}_n, \boldsymbol{w}_n)] - \Phi h = E_n^c \left\{ (Uh(\boldsymbol{z}_n, \boldsymbol{w}_n))^T [\frac{\nabla_{\boldsymbol{z}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}{f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}, \frac{\nabla_{\boldsymbol{w}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}{f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}] \right\},$$

a.e. on D as $n \longrightarrow \infty$.

The proof of Lemma 1 is an extended form of the proof of Proposition 3.2 of Weng and Tsai (2008).

Corollary 1. *Partitioning $Uh(\boldsymbol{z}_n, \boldsymbol{w}_n)$ to $(Uh_1(\boldsymbol{z}_n), Uh_2(\boldsymbol{w}_n))$, we can say that the necessary and sufficient conditions to establish asymptotic normality of joint posterior distribution are*

$$\begin{aligned} E_n^c \left\{ (Uh_1(\boldsymbol{z}_n))^T [\frac{\nabla_{\boldsymbol{z}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}{f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}] \right\} &\longrightarrow 0, \\ E_n^c \left\{ (Uh_2(\boldsymbol{w}_n))^T [\frac{\nabla_{\boldsymbol{w}_n} f_n(\boldsymbol{z}_n, \boldsymbol{w}_N)}{f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)}] \right\} &\longrightarrow 0, \end{aligned}$$

a.e. on D as $n \longrightarrow \infty$.

3. Asymptotic Normality of Posterior Distribution

In this section we establish the main result and represent a result considering more general priors for parameters. First, let

$$S = \left\{ (\boldsymbol{z}_n, \boldsymbol{w}_n) : \boldsymbol{z}_n = F_n(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \boldsymbol{w}_n = G_n(\boldsymbol{u} - \tilde{\boldsymbol{u}}_n); \boldsymbol{\psi} \in \boldsymbol{\Psi}, \boldsymbol{u} \in \boldsymbol{\Lambda} \right\}, \tag{14}$$

where $\boldsymbol{\Lambda}$ is an open subset of $\Re^q$. Moreover, for any $k \times k$ matrix J, let $\|J\|^2 = \lambda_{\max}(J^T J)$ be the spectral norm of J and let $N(\boldsymbol{a}; r)$ denote a neighborhood of $\boldsymbol{a}$ with radius r.

The following theorem, reveals the asymptotic normality of joint posterior distribution of the parameters and the random effects of GLMMs. One attractive feature of the theorem is to select a flexible random effects distribution such that $g_n(\boldsymbol{u})$ just satisfies some conditions. Then, it presents that the assumed model for the random effects, subject to $g_n(\boldsymbol{u})$ satisfies $(C1)$-$(C4)$, does not affect the asymptotic normality of joint posterior distribution. This may encourage someone to use of the approximate normal distributions for detecting the model misspecification using sensitivity analysis of the random effects distribution.

Theorem 1. *Suppose that h be any bounded measurable function as in Lemma 2 in Appendix C. Moreover, suppose that the prior $\pi(\boldsymbol{\psi})$ satisfies $(A1)$-$(A3)$, $\ell_n(\boldsymbol{\psi})$ satisfies $(B1)$-$(B4)$ and $g_n(\boldsymbol{u})$ satisfies $(C1)$-$(C4)$. Then, $E_n^c[h(\boldsymbol{z}_n, \boldsymbol{w}_n)] \xrightarrow{p} \Phi h$.*

Proof. See Appendix C. □

Note that we have to consider distributions which justify the regularity conditions. Therefor, someone has to first check to hold the regularity conditions and then apply the asymptotic normal distributions as approximate distributions of posteriors.

The conditions $(A1)$ and $(A2)$ exclude priors that are not continuously differentiable and have not compact supports such as uniform. The following Corollary presents that the result of Theorem 1 holds for more general priors.

Corollary 2. *Following Weng and Tsai (2008, Theorem 4.2 and 4.3) and the proof of Theorem 1, asymptotic normality of joint posterior distribution of the parameter vector and the random effects for a GLMM, holds with priors that must be continuous but need not have compact supports, be bounded or be differentiable such as normal or Gamma(p, ν) with a shape parameter $p < 1$.*

4. Examples

To illustrate the obtained theoretical results, we have considered two examples on spatial discrete data. In the first example, we demonstrate the asymptotic normality of posterior densities of the parameters of a spatial GLMM (SGLMM) with count Poisson responses. In the second example, we simulate a spatially correlated binary data set on a 8×8 equally-spaced regular grid of locations with a skew normal distribution for random effects. To explore the effect of misspecification on inferences and asymptotic distribution of joint posterior using a sensitivity analysis, we reanalyze the second simulation example by considering normal distribution for spatial random effects while the true mixing distribution is skew normal.

4.1. Rongelap Data of Radionuclide Counts

Rongelap data is one of the data sets used in Diggle *et al.* (1998) and is provided in R package, *geoRglm* (Christensen and Ribeiro Jr, 2002). The observations were made at 157 registration sites, $\boldsymbol{s}_i; i = 1, \ldots, 157$, and the latent spatial random effect is modeled by a Gaussian distribution. The data contains radionuclide count for various time durations m_i. Also no explanatory variables have been used previously for this data and we just have a constant term, β_0.

Following Baghishani and Mohammadzadeh (2010), for the spatial latent random effect, we use an isotropic stationary exponential covariance function

$$C(\boldsymbol{h};\boldsymbol{\theta}) = \sigma^2 \exp\left(-\frac{\|\boldsymbol{h}\|}{\phi}\right), \quad \boldsymbol{h} \in \Re^2, \tag{15}$$

in which $\boldsymbol{\theta} = (\sigma^2, \phi)$ and $\|\cdot\|$ denotes the Euclidean norm and a log link function is used. The prior distributions are considered to be independent. We use a normal prior, $N(2, 9^2)$, for β_0, a flat prior for σ^2 and an exponential prior with mean 150 for ϕ. The posterior densities of the parameters obtain by using MCMC sampling. The MCMC sampler is a Langevin–Hastings scheme introduced by Christensen and Waagepetersen (2002). Proposal variance and truncation constant for implementation of the algorithm are considered 0.012 and 300 respectively. The approximate normal distributions are also obtained using Laplace approximation applied by Eidsvik *et al.* (2009). The normal approximation is computed at the mode of the full conditional density for spatial random effects, $\pi(\boldsymbol{u}|\boldsymbol{\psi}, \boldsymbol{y})$ using Newton-Raphson optimization and fitting the covariance matrix at this mode. Then, the approximate normal density for the model parameters is obtained by expression (11) of Eidsvik *et al.* (2009).

—

Figure 1 about here.

—

Figure 1 shows the marginal posterior densities estimates of $(\beta_0, \sigma^2, \phi)$ as well as 6 selected sites of the spatial random effects, $(9, 32, 51, 77, 98, 153)$, which are obtained from retaining 10000 samples every 10 iterations after an initial burn-in period 1000 iterations. Approximate normal densities are also drawn in the figure. It is clear that the univariate normal distributions give very good approximations to the marginal posterior distributions and the approximation bias is negligible. Figure 2 also shows the contour plots of three pairs of the parameters (β, σ^2), (β, ϕ) and (σ^2, ϕ) with superimposed MCMC samples to confirm goodness of the approximation.

—

Figure 2 about here.

—

4.2. A SGLMM with the Skew Normal Random Effects

In the class of GLMMs, most users are satisfied using a Gaussian distribution for the random effects, but it is unclear whether the Gaussian assumption holds. To show that the joint posterior normality, under conditions above, holds for other mixing distributions, we use a skew normal distribution for the spatial latent random effects in a SGLMM (Hosseini *et al.*, 2010). The skew normal distribution (Azzalini, 1985) is more flexible since it includes the normal with an extra parameter λ to simplify symmetry. For an n-dimensional random vector $\boldsymbol{y}$ a multivariate skew normal density is given by

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}) = 2\phi_n(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma)\Phi(\boldsymbol{\lambda}^T \Sigma^{-\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{\mu})), \tag{16}$$

in which $\boldsymbol{\lambda} \in \Re^n$ is an n-dimensional skewness parameter. If $\boldsymbol{\lambda} = \mathbf{0}$ the density (16) reduces to the multivariate normal distribution.

We simulate spatially binary data on a 8×8 equally-spaced regular grid of locations with the following model:

$$\begin{aligned} f(y_i|\boldsymbol{\beta}) &= \left(\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{y_i} \left(1 - \frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{1-y_i}, \\ g(\mu_i) &= \ln(\frac{p_i}{1-p_i}) = \eta_i = \beta_0 + \beta_1 d_{1i} + u_i, \;\; i = 1, \ldots, n, \end{aligned}$$

where $u_i = u(\boldsymbol{s}_i)$; $\;i = 1, \ldots, n$, is a realization of a zero mean multivariate skew normal distribution with covariance function (15) and d_{1i} is the first component of ith location, i.e. $\boldsymbol{s}_i = (d_{1i}, d_{2i})$. In this example, parameters are fixed at $(\beta_0, \beta_1, \sigma^2, \phi, \lambda_0) = (2, 1.5, 1.25, 3, 15)$ in which $\boldsymbol{\lambda} = \lambda_0 \mathbf{1}$. The prior distributions are considered to be independent. We use normal priors $N(0, 2^2)$, $N(0.5, 2^2)$ and $N(10, 3^2)$ for β_0, β_1 and λ_0 respectively, an inverse Gaussian prior $IG(3, 3)$ for σ^2, and a gamma prior $\Gamma(1, \frac{1}{2})$ for ϕ.

—

Figure 3 about here.

—

—

Figure 4 about here.

—

To approximate the posterior distributions of the parameter vector and random effects, we use a Metropolis-Hastings random walk sampler. Similar to previous example, the approximate

—

Figure 5 about here.

—

normal distributions are obtained in a manner similar to the Laplace approximation applied by Eidsvik *et al.* (2009) but based on the closed skew normal (CSN) (Dominguez-Molina *et al.*, 2003) approximation instead of the usual normal approximation, see Hosseini *et al.* (2010) for more details. In this case, the approximation of the full conditional distribution of spatial latent random effects is CSN, but when $n \longrightarrow \infty$ it converges to a normal distribution as well. To explore the effect of misspecification on inferences and asymptotic distribution of joint posterior, we also consider a normal distribution for spatial random effects while the true underlying mixing distribution is skew normal and compute joint posterior as well as approximate normal under this misspecified assumption with the same priors and initials considered for skew normal assumption.

Note that it is not our focus to check, rigourously, if the regularity conditions hold with this selection for the random effects distribution. But, by a heuristic argument, we can expect that the conditions hold utilizing good properties of CSN distribution similar to normal distribution.

Figure 3(a) presents the marginal posterior densities estimates for the parameters under two correct and misleading assumptions for mixing distribution along with corresponding approximate normal distributions. The samples are obtained, separately for both model, from retaining 1000 samples every 10 iterations after an initial burn-in period 1000 iterations. It is clear that the asymptotic normality under two assumptions perform good except for σ^2 with reasonable poor approaching. However, the misspecification leads to substantial bias in the estimates of the dependency parameters, σ^2 and ϕ. Especially, it seems that the range parameter ϕ does not estimate consistently under the misspecification. It is not the case for fixed effects and misspecifying the model for the random effects only results in a small amount of bias in estimates for the fixed effects.

Figure 4 also represents the marginal posterior densities estimates for the random effects of 6 selected sites $(1, 9, 25, 36, 42, 64)$, under correct and misleading assumptions. Their approximate normal densities are also drawn in the figure. The misspecification has mild effect on the random effects rather than dependency parameters. Figure 5 also shows the contour plots of three pairs of the parameters (β_0, β_1), (σ^2, ϕ) and (ϕ, λ) with superimposed MCMC samples. This figure confirms goodness of the normal approximations for posterior distributions as well.

5. Discussion

Bayesian inference methods are used extensively in the analysis of generalized linear mixed models, but it may be difficult to handle the posterior distributions analytically. Further, exploring of the asymptotic posterior distributions is not well studied in the literature as well. Yee *et al.* (2002) introduced a method to achieve asymptotic joint posterior normality in situations where full conditional distributions corresponding to two normalized blocks of variables, with one block consisting of the model parameters and the second block consisting of the random effects, have asymptotic normal distributions. Su and Johnson (2006) also generalized the work of Yee *et al.* (2002) for b blocks of variables under simplified conditions. There are limitations for using their method. In their method it is shown that if the limiting conditional distributions are compatible in the sense of Arnold and Press (1989) and the full conditional distributions have asymptotic normal distributions, then the joint posterior will be asymptotic normal. Furthermore, as they have hinted, the posterior approximations will be poor whenever the posterior mean of any component of a block is less than two standard deviations from a boundary of the parameter space. In this paper, we established the asymptotic normality of joint posterior distribution of the parameters and the random effects of a GLMM by extending the work of Weng and Tsai (2008). This approach covers more general priors with respect to some previous related works such as Weng (2003). In addition it bears conditions which in some situations justify the asymptotic normality of posterior distributions when introduced conditions of others like Sweeting (1992) are to be failed.

One topic that has received much attention lately is to detect random effects model misspecification in GLMMs (Huang, 2008). The problem in detecting misspecification for the random effects is mainly due to the fact that there is no data realization for the random effects. Agresti *et al.* (2004) suggested comparing results from both parametric and nonparametric methods, arguing that a substantial discrepancy between the two analyses indicates model misspecification. In this work, as many recent researches, we showed that, by simulation study, misspecifying the random effects distribution in GLMMs leads to inconsistent estimators especially for the covariance parameters. One advantage of our work is to use the approximate normal distributions that are obtained under different model assumption for the random effects to detect the misspecification using sensitivity analysis. For example, this method can be used in the sensitivity analysis of the model to the common assumption of normal distribution for the random effects. If there is not much difference in results, the normal appears sufficient. Then this approximation can be a fast alternative for using in diagnostic methods for random effects model misspecification in GLMMs.

Acknowledgment

The first author wants to thank the NTNU Department of Mathematical Sciences, Norway, for their support during his visit there. He also wishes to thank Håvard Rue for helpful comments and suggestions. Partial support of ordered and spatial data center of excellence of Ferdowsi University of Mashhad is also acknowledged.

Appendix A. Modified Stein's Identity

In this appendix we derive a version of Stein's Identity with extension of work of Weng and Tsai (2008) and use of it to establish the asymptotic posterior normality of the parameters and the random effects simultaneously in GLMMs in the Section 3.

Let Γ be a finite signed measure of the form $d\Gamma = fd\Phi_r$ in which f is a real-valued function defined on $\Re^r$ satisfying $\int |f|d\Phi_r < \infty$. Write $\Phi_r h = \int hd\Phi_r$ for functions h for which the integral is finite and also write $\Gamma h = \int hd\Gamma$. For $s \geq 0$, let H_s be the collection of all measurable functions $h : \Re^r \longrightarrow \Re$ for which $|h(\boldsymbol{a})| \leq c(1 + \|\boldsymbol{a}\|^s)$ for some $c > 0$, where $\boldsymbol{a} \in \Re^r$ and let $H = \cup_{s\geq o} H_s$. Given $h \in H_s$, let $h_0 = \Phi_r h$, $h_r = h$, and

$$\begin{aligned} h_j(b_1,\ldots,b_j) &= \int_{\Re^{r-j}} h(b_1,\ldots,b_j,\boldsymbol{e})\Phi_{r-j}(d\boldsymbol{e}), \\ g_j(b_1,\ldots,b_r) &= e^{\frac{1}{2}b_j^2}\int_{b_j}^{\infty}\{h_j(b_1,\ldots,b_{j-1},c) - h_{j-1}(b_1,\ldots,b_{j-1})\}e^{-\frac{1}{2}c^2}dc, \end{aligned}$$

for $-\infty < b_1,\ldots,b_r < \infty$ and $j = 1,\ldots,r$. Then let $Uh = (g_1,\ldots,g_r)^T$. Following lemma, modifies the Stein's Identity.

Lemma 2. *(Weng and Tsai, 2008). Let s be a nonnegative integer and let $d\Gamma = fd\Phi_r$, where f is differentiable on $\Re^r$ such that*

$$\int_{\Re^r} |f|d\Phi_r + \int_{\Re^r}(1 + \|\boldsymbol{a}\|^s)\|\nabla f(\boldsymbol{a})\|\Phi_r(d\boldsymbol{a}) < \infty.$$

Then

$$\Gamma h - \Gamma 1 \cdot \Phi_r h = \int (Uh(\boldsymbol{a}))^T \nabla f(\boldsymbol{a})\Phi_r(d\boldsymbol{a}),$$

for all $h \in H_s$.

Appendix B. Regularity Conditions

Three conditions on prior distribution are needed.

(A1) $\pi(\boldsymbol{\psi})$ is continuously differentiable on $\Re^d$.

(A2) $\pi(\boldsymbol{\psi})$ has a compact support $\boldsymbol{\Psi} \subset \Re^d$

(A3) There exist ϵ_0 and δ_0 such that $\pi(\boldsymbol{\psi}) > \epsilon_0$ over $N(\boldsymbol{\psi}_0; \delta_0)$.

We also consider the regularity conditions of Weng and Tsai (2008) for $\ell_n(\boldsymbol{\psi})$ and similar modified versions of $g_n(\boldsymbol{u})$. The following conditions are required for $\ell_n(\boldsymbol{\psi})$.

(B1) $P(D_1^c) \longrightarrow 0$, $\|F_n^{-1}\| \xrightarrow{p} 0$, and $\hat{\boldsymbol{\psi}}_n \xrightarrow{p} \boldsymbol{\psi}_0$ as $n \longrightarrow \infty$, where $\xrightarrow{p}$ denotes convergence in probability.

(B2) There exists an increasing sequence of positive constants $\{b_{1n}\}$ that converges to ∞, such that

$$\sup_{\eta^{ij} \in \{\boldsymbol{\psi}: \|\boldsymbol{z}_n\| \leq b_{1n}\}} \|I_d + (F_n^T)^{-1}(\partial^2 \ell_n / \partial\psi_i \partial\psi_j(\eta^{ij}))F_n^{-1}\| \xrightarrow{p} 0.$$

(B3) Let b_{1n} be as in (B2). There exist constants $r_1 \geq 1$ and $c_1 \geq 0$ such that for all $\boldsymbol{\psi} \in \{\|\boldsymbol{z}_n\| > b_{1n}\} \cap \boldsymbol{\Psi}$, $\|(F_n^T)^{-1}\nabla k_n(\boldsymbol{\psi})\| \leq c_1 \|\boldsymbol{z}_n\|^{r_1}$.

(B4) There exist constant $r_1 \geq 1$ and a nonnegative function $v_1 : \Re^+ \times \Re^d \longrightarrow \Re$ for which, with probability tending to 1 and $\forall \boldsymbol{\psi} \in \boldsymbol{\Psi}$, $[\ell_n(\hat{\boldsymbol{\psi}}_n) - \ell_n(\boldsymbol{\psi})] \geq v_1(t, \boldsymbol{\psi})$, $e_{1n}(\boldsymbol{\psi}) = (detF_n)\|\boldsymbol{z}_n\|^{r_1} e^{-v_1(t,\boldsymbol{\psi})}$ are uniformly integrable in t and $\int_{\boldsymbol{\Psi}} e_{1n}(\boldsymbol{\psi}) d\boldsymbol{\psi}$ are uniformly bounded in t.

The following conditions are also required for $g_n(\boldsymbol{u})$.

(C1) $P(D_2^c) \longrightarrow 0$, $\|G_n^{-1}\| \xrightarrow{p} 0$, and $\tilde{\boldsymbol{u}}_n \xrightarrow{p} \boldsymbol{u}_0$ as $n \longrightarrow \infty$.

(C2) There exists an increasing sequence of positive constants $\{b_{2n}\}$ that converges to ∞, such that

$$\sup_{\zeta^{ij} \in \{\boldsymbol{u}: \|\boldsymbol{w}_n\| \leq b_{2n}\}} \|I_q + (G_n^T)^{-1}(\partial^2 g_n / \partial u_i \partial u_j(\zeta^{ij}))G_n^{-1}\| \xrightarrow{p} 0.$$

(C3) Let b_{2n} be as in (C2). There exist constants $r_2 \geq 1$ and $c_2 \geq 0$ such that for all $\boldsymbol{u} \in \{\|\boldsymbol{w}_n\| > b_{2n}\} \cap \boldsymbol{\Lambda}$, $\|(G_n^T)^{-1}\nabla l_n(\boldsymbol{u})\| \leq c_2 \|\boldsymbol{w}_n\|^{r_2}$.

(C4) There exist constant $r_2 \geq 1$ and a nonnegative function $v_2 : \Re^+ \times \Re^q \longrightarrow \Re$ for which, with probability tending to 1 and $\forall \boldsymbol{u} \in \boldsymbol{\Lambda}$, $[g_n(\tilde{\boldsymbol{u}}_n) - g_n(\boldsymbol{u})] \geq v_2(t, \boldsymbol{u})$, $e_{2n}(\boldsymbol{u}) = (detG_n)\|\boldsymbol{w}_n\|^{r_2} e^{-v_2(t,\boldsymbol{u})}$ are uniformly integrable in t and $\int_{\boldsymbol{\Lambda}} e_{2n}(\boldsymbol{u}) d\boldsymbol{u}$ are uniformly bounded in t.

As Weng and Tsai (2008) have represented, the first two conditions (B1)-(B2) for $\ell_n(\boldsymbol{\psi})$ and (C1)-(C2) for $g_n(\boldsymbol{u})$ regard information growth and continuity, respectively, and the last two conditions (B3)-(B4) for $\ell_n(\boldsymbol{\psi})$ and (C3)-(C4) for $g_n(\boldsymbol{u})$ concern some integrability properties of $\exp\{\ell_n(\hat{\boldsymbol{\psi}}_n) - \ell_n(\boldsymbol{\psi})\}$ and $\exp\{g_n(\tilde{\boldsymbol{u}}_n) - g_n(\boldsymbol{u})\}$ over S, respectively, which essentially involve the tail behavior of $\ell_n(\boldsymbol{\psi})$ and $g_n(\boldsymbol{u})$. Also note that the uniformly bounded conditions in (B4) and (C4) are guaranteed by the uniformly integrability, provided that Ψ and Λ are bounded. Furthermore, Weng and Tsai (2008) compared these conditions with conditions of Weng (2003) and Sweeting (1992). They showed that, by examples, their conditions are more easy to check nonlocal behavior of the posterior distribution. For more details see Weng and Tsai (2008, Section 5).

Appendix C. Proof of Theorem 1

For proving the Theorem 1, we need the following lemmas.

Lemma 3. *(Weng and Tsai, 2008). If h is a bounded measurable function, then $\|Uh(\boldsymbol{a})\| \leq c_0$ for some $c_0 > 0$ and for all $\boldsymbol{a} \in \Re^r$. Moreover if $h(\boldsymbol{a}) = \|\boldsymbol{a}\|^s$, $s \geq 1$, then for some $c_s > 0$*

$$\|Uh(\boldsymbol{a})\| \leq c_s(1 + \|\boldsymbol{a}\|^{s-1}).$$

Lemma 4. *(Weng and Tsai, 2008) Under the regularity conditions there exist $0 < K_3 < \infty$ such that, with probability tending to 1,*

$$E_n^c\{\|\nabla\pi(\boldsymbol{\psi})\|/\pi(\boldsymbol{\psi})\} \leq K_3.$$

Considering (5)-(8) and the regularity conditions above, two following lemmas are obtained so that their validates are established following results of Weng and Tsai (2008) but for both $\ell_n(\boldsymbol{\psi})$ and $g_n(\boldsymbol{u})$.

Lemma 5. 1. *If (B2) and (C2) hold, there exist constants p_1, p_2, q_1 and q_2 such that, with probability tending to 1,*

$$\begin{aligned}
\sup_{\boldsymbol{\psi}:\|\boldsymbol{z}_n\| \leq p_1} \{\ell_n(\hat{\boldsymbol{\psi}}_n) - \ell_n(\boldsymbol{\psi})\} &\leq q_1, \\
\sup_{\boldsymbol{u}:\|\boldsymbol{w}_n\| \leq p_2} \{g_n(\tilde{\boldsymbol{u}}_n) - g_n(\boldsymbol{u})\} &\leq q_2.
\end{aligned}$$

2. *If $(B3)$ and $(C3)$ hold, then for some $0 < M < \infty$, with probability tending to 1,*

$$\int_S e^{\{\ell_n(\boldsymbol{\psi})-\ell_n(\hat{\boldsymbol{\psi}}_n)\}} e^{\{g_n(\boldsymbol{u})-g_n(\tilde{\boldsymbol{u}}_n)\}} d\boldsymbol{z}_n d\boldsymbol{w}_n < M,$$

$$\int_S \|\boldsymbol{z}_n\| \|\boldsymbol{w}_n\| e^{\{\ell_n(\boldsymbol{\psi})-\ell_n(\hat{\boldsymbol{\psi}}_n)\}} e^{\{g_n(\boldsymbol{u})-g_n(\tilde{\boldsymbol{u}}_n)\}} d\boldsymbol{z}_n d\boldsymbol{w}_n < M,$$

$$\int_{S'} \|\boldsymbol{z}_n\|^{r_1} \|\boldsymbol{w}_n\|^{r_2} e^{\{\ell_n(\boldsymbol{\psi})-\ell_n(\hat{\boldsymbol{\psi}}_n)\}} e^{\{g_n(\boldsymbol{u})-g_n(\tilde{\boldsymbol{u}}_n)\}} d\boldsymbol{z}_n d\boldsymbol{w}_n \xrightarrow{p} 0,$$

where, $S' = S \cap \{\|\boldsymbol{z}_n\| > b_{1n}\} \cap \{\|\boldsymbol{w}_n\| > b_{2n}\}$.

Lemma 6. *Let $f_n(\boldsymbol{z}_n, \boldsymbol{w}_n)$ and S be as in (9) and (14) respectively. Suppose that $\pi(\boldsymbol{\psi})$ satisfies $(A1)$-$(A3)$. Then,*

D1 *If $(B1)$, $(B2)$, $(C1)$ and $(C2)$ hold, then there exists $K_1 > 0$ such that, with probability tending to 1,*

$$\int_S \phi_d(\boldsymbol{z}_n)\phi_q(\boldsymbol{w}_n) f_n(\boldsymbol{z}_n, \boldsymbol{w}_n) d\boldsymbol{z}_n d\boldsymbol{w}_n > K_1.$$

D2 *If $(B4)$ and $(C4)$ hold, then there exists $K_2 > 0$ such that, with probability tending to 1,*

$$\int_S \phi_d(\boldsymbol{z}_n)\phi_q(\boldsymbol{w}_n) f_n(\boldsymbol{z}_n, \boldsymbol{w}_n) d\boldsymbol{z}_n d\boldsymbol{w}_n < K_2.$$

Proof of Theorem 1: Note that Uh and $\pi(\boldsymbol{\psi})$ are bounded by Lemma 3 and $(A1)$-$(A2)$. From (12), (13) and Lemma 1, for a.e. on D, we have

$$E_n^c[h(\boldsymbol{z}_n, \boldsymbol{w}_n)] - \Phi h = E_{n,\boldsymbol{z}_n}^c + E_{n,\boldsymbol{w}_n}^c,$$

as $n \longrightarrow \infty$, where

$$\begin{aligned} E_{n,\boldsymbol{z}_n}^c &= E_n^c \left\{ (Uh_1(\boldsymbol{z}_n))^T (F_n^T)^{-1} \frac{\nabla \pi(\boldsymbol{\psi})}{\pi(\boldsymbol{\psi})} \right\} \\ &+ E_n^c \left\{ (Uh_1(\boldsymbol{z}_n))^T (F_n^T)^{-1} \nabla k_n(\boldsymbol{\psi}) \right\} = I_{\boldsymbol{z}_n} + II_{\boldsymbol{z}_n} \end{aligned} \tag{17}$$

$$E_{n,\boldsymbol{w}_n}^c = E_n^c \left\{ (Uh_2(\boldsymbol{w}_n))^T (G_n^T)^{-1} \nabla l_n(\boldsymbol{u}) \right\} = I_{\boldsymbol{w}_n}. \tag{18}$$

Since $P(D_1^c) \longrightarrow 0$ by $(B1)$ and $P(D_2^c) \longrightarrow 0$ by $(C1)$, it suffices to show $I_{\boldsymbol{z}_n} + II_{\boldsymbol{z}_n} \xrightarrow{p} 0$ and $I_{\boldsymbol{w}_n} \xrightarrow{p} 0$. First component of right hand side of (17), $I_{\boldsymbol{z}_n}$, converges to zero by Lemma 4 and $\|F_n^{-1}\| \xrightarrow{p} 0$ under $(B1)$. Moreover, from (17) we have,

$$II_{\boldsymbol{z}_n} = \frac{\int_S (Uh_1(\boldsymbol{z}_n))^T (F_n^T)^{-1} \nabla k_n(\boldsymbol{\psi}) \phi_d(\boldsymbol{z}_n)\phi_q(\boldsymbol{w}_n) f_n(\boldsymbol{z}_n, \boldsymbol{w}_n) d\boldsymbol{z}_n d\boldsymbol{w}_n}{\int_S \phi_d(\boldsymbol{z}_n)\phi_q(\boldsymbol{w}_n) f_n(\boldsymbol{z}_n, \boldsymbol{w}_n) d\boldsymbol{z}_n d\boldsymbol{w}_n}. \tag{19}$$

The denominator of (19) is bounded below by some $K_1 > 0$ by Lemma 6(D1). Then we just need to show that the numerator converges to 0 in probability. First we decompose the numerator into two integrals over $\|\boldsymbol{z}_n\| \leq b_{1n}$ and $\|\boldsymbol{z}_n\| > b_{1n}$ and call the corresponding integrals as $II_{\boldsymbol{z}_n,1}$ and $II_{\boldsymbol{z}_n,2}$ respectively. With respect to (10), $(A1)$-$(A2)$ and Lemma 3, there exists a constant $C_1 > 0$ such that

$$\begin{aligned} |II_{\boldsymbol{z}_n,1}| &\leq \int_{\|\boldsymbol{z}_n\|\leq b_{1n}} |(Uh_1(\boldsymbol{z}_n))^T (F_n^T)^{-1} \nabla k_n(\boldsymbol{\psi})| \pi(\boldsymbol{\psi}) e^{\ell_n(\boldsymbol{\psi})-\ell_n(\hat{\boldsymbol{\psi}})} e^{g_n(\boldsymbol{u})-g_n(\tilde{\boldsymbol{u}})} d\boldsymbol{z}_n d\boldsymbol{w}_n \\ &\leq C_1 \sup_{\boldsymbol{\psi}:\|\boldsymbol{z}_n\|\leq b_{1n}} \|I_d - (F_n^T)^{-1}[-(\frac{\partial^2 \ell_n}{\partial\psi_i \partial\psi_j})(\psi^{*ij})]F_n^{-1}\| \\ &\times \int_{\|\boldsymbol{z}_n\|\leq b_{1n}} \|\boldsymbol{z}_n\| e^{\ell_n(\boldsymbol{\psi})-\ell_n(\hat{\boldsymbol{\psi}})} e^{g_n(\boldsymbol{u})-g_n(\tilde{\boldsymbol{u}})} d\boldsymbol{z}_n d\boldsymbol{w}_n, \end{aligned}$$

where the relations between $\boldsymbol{\psi}$ and $\boldsymbol{z}_n$ and $\boldsymbol{u}$ and $\boldsymbol{w}_n$ are given in (2) and (3). Using $(B2)$ and Lemma 5, part 2, we conclude that $II_{\boldsymbol{z}_n,1} \xrightarrow{p} 0$. Next, by $(B3)$, $(A1)$-$(A2)$ and Lemma 3, there exists a constant $C_2 > 0$ such that

$$|II_{\boldsymbol{z}_n,2}| \leq C_2 \int_{S\cap\{\|\boldsymbol{z}_n\|>b_{1n}\}} \|\boldsymbol{z}_n\|^{r_1} e^{\ell_n(\boldsymbol{\psi})-\ell_n(\hat{\boldsymbol{\psi}})} e^{g_n(\boldsymbol{u})-g_n(\tilde{\boldsymbol{u}})} d\boldsymbol{z}_n d\boldsymbol{w}_n,$$

which using Lemma 5, part 2, converges to 0 in probability. Hence, $II_{\boldsymbol{z}_n} \xrightarrow{p} 0$. Similarly, $I_{\boldsymbol{w}_n} \xrightarrow{p} 0$ follows from $(A1)$-$(A2)$, $(C2)$-$(C3)$, (11), Lemma 3 and Lemma 5, part 2. This completes the proof.

References

Agresti, A., Caffo, B. & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies, *Computational Statistics and Data Analysis,* **47**, 639-653.

Arnold, B. C. & Press, S. J. (1989). Compatible conditional distributions, *Journal of the American Statistical Association,* **84**, 152-156.

Azzalini, A. (1985). A class of distributions which includes the normal one, *Scandinavian Journal of Statistics,* **12**, 171-178.

Baghishani, H. & Mohammadzadeh, M. (2010). A data cloning algorithm for computing maximum likelihood in spatial generalized linear mixed models, *Submitted*.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American statistical association,* **88**, 9-25.

Chen, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implication, *Journal of the Royal Statistical Society, Series B,* **47**, 540-546.

Christensen, O. F. & Ribeiro Jr, P. J. (2002). geoRglm - A package for generalized linear spatial models, *R News,* **2**, 26-28.

Christensen O. F. & Waagepetersen R. P. (2002). Bayesian prediction of spatial count data using generalized linear mixed models, *Biometrics,* **58**, 280-286.

Diggle, P., Tawn, J. A. & Moyeed, R. A. (1998). Model-based geostatistic, *Journal of the Royal Statistical Society, Series C. Applied Statistics,* **47**, 299-350.

Dominguez-Molina, J., Gonzalez-Farias, G. & Gupta, A. (2003). The multivariate closed skew normal distribution, *Technical Report 03-12, Department of mathematics and statistics, Bowling Green State University.*

Eidsvik, J., Martino, S. & Rue, H. (2009). Approximate Bayesian inference in spatial generalized linear mixed models, *Scandinavian journal of Statistics,* **36**, 1-22.

Heagerty, P. J. & Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models, *Biometrika,* **88**, 973-985.

Heyde, C. C. & Johnstone, I. M. (1979). On asymptotic posterior normality for stochastic processes, *Journal of the Royal Statistical Society, Series B,* **41**, 184-189.

Hosseini, F., Eidsvik, J. and Mohammadzadeh, M. (2010). Approximate Bayesian inference in spatial generalized linear mixed models with skew normal latent variables, *Submitted*.

Huang, X. (2008). Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response, *Biometrics,* DOI: 10.1111/j.1541-0420.2008.01103.x.

Johnson, R. (1970). Asymptotic expansions associated with posterior distributions, *The Annals of Mathematical Statistics,* **41**, 851-864.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, *University of California Publication in Statistics,* **1**, 277-330.

Litiere, S., Alonso, A. & Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models, *Biometrics,* **63**, 1038-1044.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models, *Journal of the American statistical association,* **92**, 162-170.

Su, C. L. & Johnson, W. O. (2006). Large-sample joint posterior approximations when full conditionals are approximately normal: application to generalized linear mixed models, *Journal of the American Statistical Association,* **101** 795-811.

Sweeting, T. J. & Adekola, A. O. (1987). Asymptotic posterior normality for stochastic processes revisited, *Journal of the Royal Statistical Society, Series B,* **49**, 215-222.

Sweeting, T. J. (1992). On asymptotic posterior normality in the multiparameter case, In: Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (Eds.), Bayesian Statistics. *Oxford University Press, Oxford,* 825-835.

Walker, A. M. (1969). On the asymptotic behavior of posterior distributions, *Journal of the Royal Statistical Society, Series B,* **31**, 80-88.

Weng, R. C. (2003). On Stein's identity for posterior normality, *Statistica Sinica,* **13**, 495-506.

Weng, R. C. & Tsai, W. C. (2008). Asymptotic posterior normality for multiparameter problems, *Journal of statistical planning and inference,* **138**, 4068-4080.

Woodroofe, M. (1989). Very weak expansions for sequentially designed experiments: linear models, *The annals of statistics,* **17**, 1087-1102.

Yee, J., Johnson, W. O. & Samaniego, F. J. (2002). Asymptotic approximations to posterior distributions via conditional moment equations, *Biometrika,* **89**, 755-767.

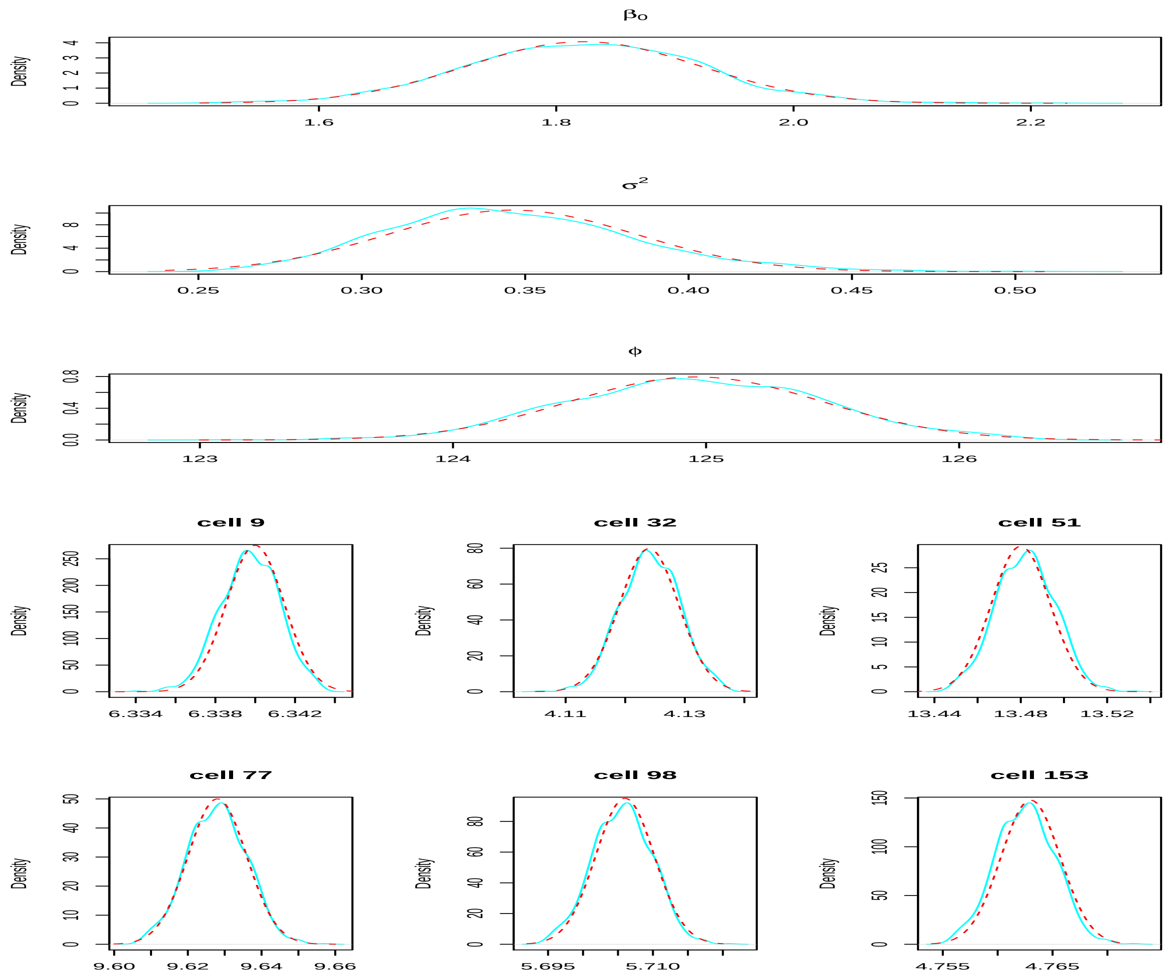

Figure 1: Rongelap data. (a): Posterior densities (solid) for parameters $(\beta_0, \sigma^2, \phi)$. (b): Posterior densities of the random effects (solid) for 6 selected sites $(9, 32, 51, 77, 98, 153)$. The dashed curves are approximate normal densities.

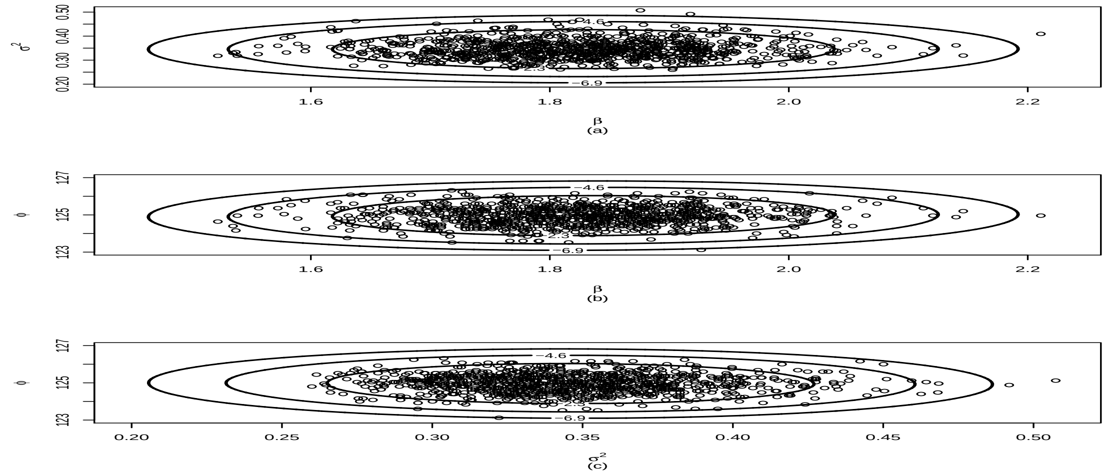

Figure 2: Rongelap data. Contour plots of the normal approximations of the joint posterior distribution of (a) (β, σ^2), (b) (β, ϕ) and (c) (σ^2, ϕ). The points represent obtained samples from MCMC sampling.

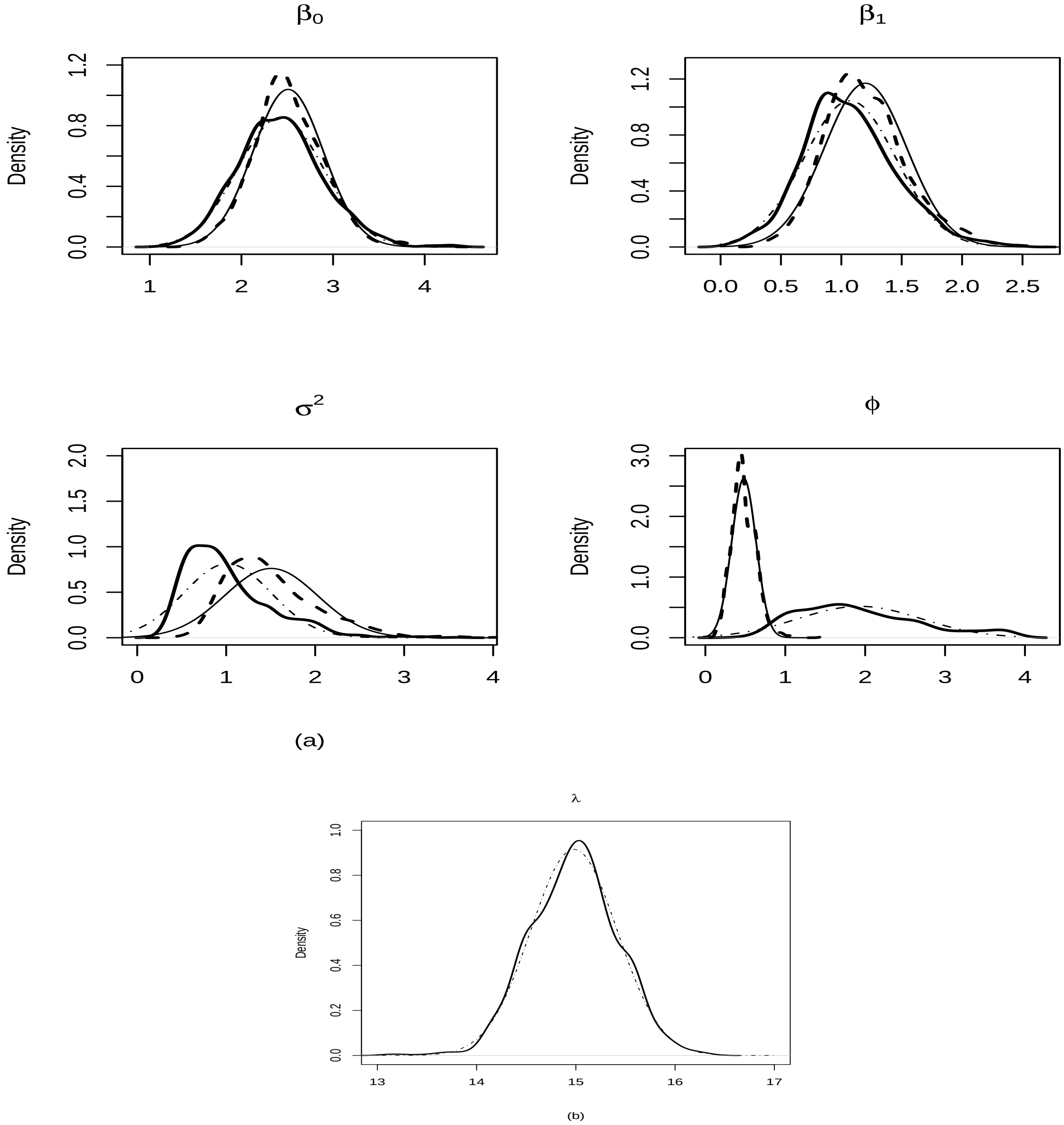

Figure 3: Simulation study. (a): Posterior densities estimates for the parameters of spatial GLMM, $(\beta_0, \beta_1, \sigma^2, \phi)$ obtained by two correct skew normal (thick solid) and misleading normal (thick dashes) assumptions for mixing distribution along side corresponding approximate normal densities (thin dot-dash and thin solid, respectively). (b): Posterior density estimate of λ (solid) along side its approximate normal density (dot-dashes).

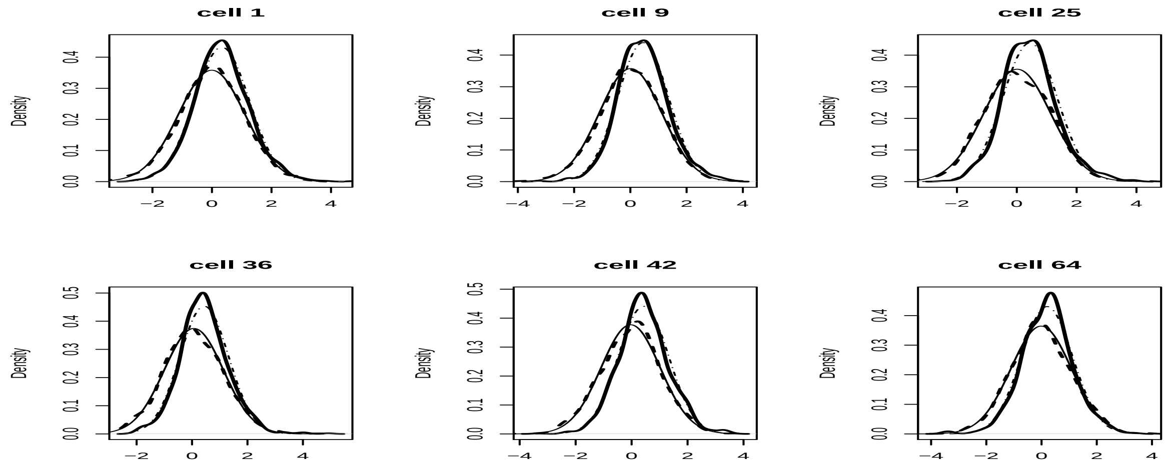

Figure 4: Simulation study. Posterior densities estimates for the random effects for 6 selected sites $(1, 9, 25, 36, 42, 64)$ computed by two correct skew normal (thick solid) and misleading normal (thick dashes) assumptions for mixing distribution along side corresponding approximate normal densities (thin dot-dash and thin solid, respectively).

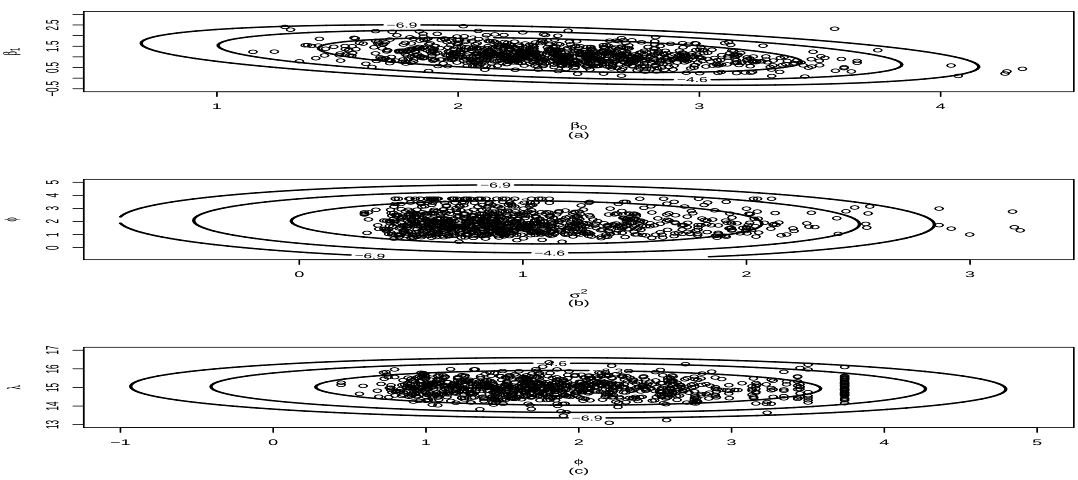

Figure 5: Simulation study. Contour plots of the normal approximations of the joint posterior distribution of (a) (β_0, β_1), (b) (σ^2, ϕ) and (c) (ϕ, λ). The points represent obtained samples from MCMC sampling.