Computational Statistics and Data Analysis 55 (2011) 1748–1759

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models

Hossein Baghishani, Mohsen Mohammadzadeh*

Department of Statistics, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 30 April 2009
Received in revised form 29 September 2010
Accepted 3 November 2010
Available online 21 November 2010

Keywords:
Data cloning
Generalized linear mixed models
MCMC algorithms
Spatial generalized linear mixed models

ABSTRACT

Non-Gaussian spatial data are common in many sciences such as environmental sciences, biology and epidemiology. Spatial generalized linear mixed models (SGLMMs) are flexible models for modeling these types of data. Maximum likelihood estimation in SGLMMs is usually made cumbersome due to the high-dimensional intractable integrals involved in the likelihood function and therefore the most commonly used approach for estimating SGLMMs is based on the Bayesian approach. This paper proposes a computationally efficient strategy to fit SGLMMs based on the data cloning (DC) method suggested by Lele et al. (2007). This method uses Markov chain Monte Carlo simulations from an artificially constructed distribution to calculate the maximum likelihood estimates and their standard errors. In this paper, the DC method is adapted and generalized to estimate SGLMMs and some of its asymptotic properties are explored. Performance of the method is illustrated by a set of simulated binary and Poisson count data and also data about car accidents in Mashhad, Iran. The focus is inference in SGLMMs for small and medium data sets.

1. Introduction

Conventional geostatistical methods, such as kriging, solve the problem of estimation and prediction for a random field which is Gaussian. These methods clearly are not applicable in many practical situations in which the data are discrete. Then generalized linear mixed models (GLMMs) with spatially correlated random effects come handy. As an extension of generalized linear models (GLMs) (McCullagh and Nelder, 1989), a GLMM assumes that the response variable follows a distribution from the exponential family and is conditionally independent given a latent variable, while the latent variable is modeled by a random effect that is typically Gaussian (Breslow and Clayton, 1993). For spatial data, the random effect is typically considered to be a spatial process which is modeled by a zero mean Gaussian random field (GRF). Diggle et al. (1998) introduced this type of GLMMs for spatial data and today many statisticians refer to them as spatial GLMMs (SGLMMs).

Fitting such models has been the subject of a great deal of research over the past decades. Both frequentist and Bayesian methods have been developed for inference in SGLMMs. Breslow and Clayton (1993) considered two approximate methods, penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) for inference in SGLMMs. Heagerty and Lele (1998) used a pairwise likelihood approach (Lindsay, 1988) to analyze binary spatial data in a spatial probit model. Zhang (2002) applied an EM algorithm aided by MCMC algorithms to obtain maximum likelihood estimates (MLE). Varin et al. (2005) used a pairwise likelihood approach aided by a new EM-type algorithm which utilizes numerical quadrature. There have also been contributions to Bayesian computation of SGLMMs in the literature. Diggle et al. (1998) used a Bayesian MCMC algorithm with priors on the unknown regression parameters and the covariance parameters of the GRF. A more efficient

* Corresponding author. Tel.: +98 21 82883483; fax: +98 21 82883483.
E-mail addresses: hbaghishani@modares.ac.ir (H. Baghishani), mohsen_m@modares.ac.ir (M. Mohammadzadeh).
URL: http://www.modares.ac.ir/mat/mohsen_m (M. Mohammadzadeh).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.11.004

Langevin–Hastings MCMC algorithm was given by Christensen and Waagepetersen (2002). A robust version of their method was also introduced by Christensen et al. (2006). Zhao et al. (2006) described a Bayesian approach in fitting a general design SGLMM.

Due to the advances in computation, the most commonly used approach for fitting SGLMMs is based on the Bayesian paradigm, especially MCMC algorithms. However, the most serious criticism of the Bayesian inferences is their dependence on the choice of the prior distributions.

In comparison with Bayesian analysis, likelihood based statistical inferences for SGLMMs can be extremely difficult. Maximum likelihood estimation in these models generally involves numerical integration of high dimensions, so the intractable integral is the main problem in these models. Therefore, methods for approximating the integral will be the main theme of this class of models. There are two general types of solutions. The first is to approximate the integral numerically, so that the marginal likelihood can be computed and maximized. The methods of this type include: adaptive Gauss–Hermite quadratures (AGHQ) (Pinheiro and Bates, 1995; Lesaffre and Speissens, 2001), Monte Carlo EM (MCEM), Monte Carlo Newton–Raphson (MCNR) (McCulloch, 1997; Booth and Hobert, 1999), and simulated maximum likelihood (SML) (Geyer and Thompson, 1992; McCulloch, 1997; Christensen, 2004). The second is to approximate the integrand, so that the integral of the approximation has closed form. The methods of this type also include: Laplace method (Breslow and Clayton, 1993; Breslow and Lin, 1995; Skaug, 2002), PQL and MQL (Breslow and Clayton, 1993).

The AGHQ is only useful for low-dimensional random effects since its computational loads increase exponentially with the dimension of random effects. MCEM and MCNR methods have good performance in fitting SGLMMs, but these methods can be time consuming and their convergence diagnosis is not trivial. Also the SML method, in general, has a weak performance. On the other hand, the approximate methods such as PQL and MQL are not difficult to implement within standard statistical softwares but they generally do not provide consistent estimators and may be seriously biased when applied to binary response data (Rodriguez and Goldman, 1995).

The main contribution of the present paper is to propose a computationally efficient strategy to fit SGLMMs based on the DC method introduced by Lele et al. (2007). This method, as a computational trick, uses MCMC simulations from an artificially constructed distribution to calculate MLE and their standard errors.

In the next section SGLMMs and the DC technique are described, while the adaptation of the DC method to estimate SGLMMs is carried out in Section 3. Section 4 states the asymptotic properties of the resulting estimators. In Section 5, the performance of the method is explored through simulation studies and an application to car accident data from Mashhad, Iran. Section 6 concludes with a discussion.

2. SGLMMs and the DC method

SGLMMs are flexible models for modeling spatially dependent and non-Gaussian random variables (Diggle et al., 1998). A SGLMM is a GLMM in which the underlying random effects are modeled by a random field. This random field is usually assumed to be Gaussian and like standard GLMM (Breslow and Clayton, 1993), given the random effects, the observations at the measurement locations are conditionally independent and follow a generalized linear model.

Let $D \subseteq \Re^d$ be the region of interest and for the ith observation, let s_i be a particular location within D, $Y(s_i)$ be the response variable, and $\mathbf{x}(s_i)$ be a p-dimensional vector of covariates for the fixed effects, where $i = 1, \ldots, n$. We define the SGLMM as follows:

$$E(Y(s_i)|u(s_i)) = g^{-1}(\mathbf{x}'(s_i)\boldsymbol{\beta} + u(s_i)), \quad i = 1, \ldots, n;$$

where,

- $g(\cdot)$ is a real-valued differentiable and invertible link function and $\boldsymbol{\beta}$ is a vector of p regression parameters.
- $\{U(\mathbf{s}) : \mathbf{s} \in D\}$ is a second order stationary GRF with zero mean and spatial covariance function $\mathrm{cov}(U(\mathbf{s}), U(\mathbf{s}')) = C(\mathbf{s} - \mathbf{s}'; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of correlation parameters. $\boldsymbol{u} = (u(s_1), \ldots, u(s_n))'$ is a realization of $U(\cdot)$.
- Conditionally on $U(\cdot)$, $Y(\cdot)$ is an independent process, i.e. given $\boldsymbol{u}$, the observations $\boldsymbol{y} = (y(s_1), \ldots, y(s_n))'$ are mutually independent.
- The form of conditional density function of $y_i = y(s_i)$ given $u_i = u(s_i)$, $i = 1, \ldots, n$, is

$$f(y_i|u_i; \boldsymbol{\beta}) = \exp[a(\mu_i)y_i - b(\mu_i)]c(y_i),$$

where $\mu_i = E(Y_i|u_i)$ and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are specific functions.

Then, the marginal likelihood function of the SGLMM will be

$$L(\boldsymbol{\psi}; \boldsymbol{y}) = \int \prod_{i=1}^{n} f(y_i|u_i; \boldsymbol{\beta})\phi_n(\boldsymbol{u}; 0, \Sigma_{\boldsymbol{\theta}})\mathrm{d}\boldsymbol{u}, \tag{1}$$

where $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ are the parameters of the model and $\phi_n(\cdot; 0, \Sigma_{\boldsymbol{\theta}})$ is the n-variate normal density function with zero mean and covariance matrix $\Sigma_{\boldsymbol{\theta}} = (c_{ij}) = (C(s_i - s_j; \boldsymbol{\theta}))$, $i, j = 1, \ldots, n$. Here the calculation of the marginal likelihood function nearly always involves intractable integrals, which is the main impediment. Also the computational burden increases with the number of observations because the dimension of the random field is equal to the number of observations.

Now, let $\boldsymbol{y} = (y(s_1), \ldots, y(s_n))'$ be the observed data. Furthermore, let $\pi(\boldsymbol{\psi})$ denote the joint prior density of the parameters. Then the joint posterior density is defined by

$$\pi(\boldsymbol{\psi}, \boldsymbol{u}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\psi})\phi_n(\boldsymbol{u}; 0, \Sigma_{\boldsymbol{\theta}})\pi(\boldsymbol{\psi})}{m(\boldsymbol{y})},$$

where,

$$m(\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\psi})\phi_n(\boldsymbol{u}; 0, \Sigma_{\boldsymbol{\theta}})\pi(\boldsymbol{\psi})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{\psi},$$

which is not available in closed form because of the same intractable integrals that cause trouble in the likelihood function. A flat prior, $\pi(\boldsymbol{\psi}) \propto 1$, results in a posterior that is simply a constant multiple of the likelihood function (1). Therefore, if the resulting posterior is proper, MCMC algorithms can be used to study the likelihood function. But flat priors lead to improper posteriors for many of the GLMMs (Natarajan and McCulloch, 1995). Using a diffuse prior, which is a proper prior, leads to a proper posterior and one might hope that the resulting posterior is close to the likelihood function, but the use of it need not result in a posterior mode that is close to the MLE (Kass and Wasserman, 1996).

A recent suitable alternative method can be the DC method, which was first introduced by Lele et al. (2007) in ecological studies to calculate MLE for the parameters and their standard errors in a GLMM. This method uses an MCMC algorithm from an artificially constructed distribution to compute MLE and their variance estimates. Although the distribution looks like a Bayesian posterior distribution, it is constructed from two functions which are not in fact a prior distribution and a likelihood. However, considering them as prior and likelihood can be mimicked virtually. Hereafter, we call them prior and cloned likelihood, respectively.

Let $\pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y}) \propto \pi(\boldsymbol{\psi})[L(\boldsymbol{\psi}; \boldsymbol{y})]^k$ be the artificially constructed density, which we name as the DC-based density, constructed from k identical and independent clones of the data and $\pi(\boldsymbol{\psi})$. According to Theorem 1, we have

$$\mathrm{E}^{(k)}(\boldsymbol{\psi}|\boldsymbol{y}) \longrightarrow \hat{\boldsymbol{\psi}},$$
$$\mathrm{Var}^{(k)}(\boldsymbol{\psi}|\boldsymbol{y}) \longrightarrow k^{-1} \times \mathrm{Var}(\hat{\boldsymbol{\psi}}),$$

as $k \to \infty$, in which $\mathrm{E}^{(k)}(\boldsymbol{\psi}|\boldsymbol{y})$ and $\mathrm{Var}^{(k)}(\boldsymbol{\psi}|\boldsymbol{y})$ are the expectation and variance of the DC-based distribution, respectively. Then, the trick is to generate samples from a DC-based distribution, which is constructed by cloning the original data set, using a MCMC algorithm. One major advantage of this method is the invariancy of the results to the choice of priors. In the following section, we will first adapt a DC method to estimate a SGLMM and then present some asymptotic properties including the above mentioned results.

3. Adaptive DC for SGLMMs

The idea of the DC method for estimating a SGLMM is to derive the joint DC-based density of parameters as well as k realizations $\boldsymbol{u}_{(k)} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$ of GRF. Although we cannot have k independent replications of the same experiment yielding exactly the same data in reality, it can be mimicked using computers. Then, the cloned likelihood function for the k-repeated cloned data is the original likelihood raised to the kth power. Therefore, we can calculate the marginal DC-based density of the parameters, $\pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y})$, with the usual MCMC approach. So, according to Theorem 2, the marginal DC-based density will be concentrated around the MLE if k goes to infinity and the results are invariant to the choice of priors (for more details see Lele et al., 2007).

According to Theorem 1, stated later in Section 4, the mean of the resulting marginal DC-based distribution, in which the k realizations $\boldsymbol{u}_{(k)}$ have been integrated out from the joint DC-based density, equals the MLEs of the model parameters and k times the variance of the marginal DC-based distribution equals the asymptotic variance of the MLEs.

Let $\boldsymbol{y}^{(k)} = (\boldsymbol{y}, \ldots, \boldsymbol{y})$ denote the k-repeated cloned observed data. Because of the usefulness and easy implementation of the Metropolis–Hastings algorithm (Hastings, 1970), we apply it to sample from this marginal density.

A general data cloning algorithm can consist of the following steps:

Step 0: Choose initial values $(\boldsymbol{\psi}^{(0)}, \boldsymbol{u}^{(0)})$. Set $\boldsymbol{u}_{(k)}^{(0)}$ equal to k copies of $\boldsymbol{u}^{(0)}$ and $\ell = 0$.

Step 1: Generate a new proposed value $\boldsymbol{\theta}^* = (\sigma^{*2}, \phi^*)$ as follows:
1. σ^{*2} from a proposal density $q(\sigma^{*2}|\sigma^2)$.
2. ϕ^* from a proposal density $q(\phi^*|\phi)$.

Step 2: Generate k values of $\boldsymbol{u}, \boldsymbol{u}_{(k)}^* = (\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_k^*)$, from zero mean GRF with correlation parameters $\boldsymbol{\theta}^*$.

Step 3: Generate a new proposed value $\boldsymbol{\beta}^*$ from a proposal density $q(\boldsymbol{\beta}^*|\boldsymbol{\beta})$.

Step 4: Accept new generated values $\boldsymbol{\psi}^* = (\boldsymbol{\beta}^*, \sigma^{*2}, \phi^*)$ with probability

$$\alpha(\boldsymbol{\psi}, \boldsymbol{\psi}^*) = \min\left\{1, \frac{\left\{\prod_{m=1}^{k} f(\boldsymbol{y}|\boldsymbol{u}_m^*, \boldsymbol{\beta}^*)\right\}\pi(\boldsymbol{\psi}^*)q(\boldsymbol{\beta}|\boldsymbol{\beta}^*)q(\sigma^2|\sigma^{*2})q(\phi|\phi^*)}{\left\{\prod_{m=1}^{k} f(\boldsymbol{y}|\boldsymbol{u}_m^*, \boldsymbol{\beta})\right\}\pi(\boldsymbol{\psi})q(\boldsymbol{\beta}^*|\boldsymbol{\beta})q(\sigma^{*2}|\sigma)q(\phi^*|\phi)}\right\},$$

and set $\boldsymbol{\psi}^{(\ell+1)} = \boldsymbol{\psi}^*, \boldsymbol{u}_{(k)}^{(\ell+1)} = \boldsymbol{u}_{(k)}^*$, otherwise $\boldsymbol{\psi}^{(\ell+1)} = \boldsymbol{\psi}^{(\ell)}, \boldsymbol{u}_{(k)}^{(\ell+1)} = \boldsymbol{u}_{(k)}^{(\ell)}$.

Step 5: Increase ℓ by 1 and repeat Steps 1 to 4 until the algorithm is judged to have reached its equilibrium distribution.

Note that we selected a Metropolis–Hastings MCMC algorithm to describe the data cloning algorithm for its simplicity. But it may be possible to use other MCMC algorithms, with good mixing behavior, such as adaptive Metropolis–Hastings to give faster convergence to the equilibrium distribution (Robert and Casella, 2005). Furthermore, note that a data cloning algorithm is a MCMC algorithm and all important practical problems such as tuning, monitoring and assessing convergency hold here. All strategies proposed to improve MCMC algorithms can also be applied for a DC algorithm. To know about what can be done when a DC algorithm convergence slows; see Robert and Casella (2005).

According to the Markov chain theory, the resulting sample, beginning after a burn-in period, has been generated from the marginal DC-based distribution $\pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y})$. In principle, these samples can be made arbitrarily precise by increasing the length of the simulation run. Note that the k sets of $\boldsymbol{u}_j,\ j = 1, \ldots, k$ generated in each step of the above described algorithm are just discarded. The MLE of $\boldsymbol{\psi}$ correspond to the sample mean values and the approximate variances of the MLE correspond to k times the sample variances.

According to Varin et al. (2005), choosing the initial values for $\boldsymbol{\psi}$, $\boldsymbol{u}$ and an appropriate covariance function is done as follows: First, the initial values for regression parameters, $\beta^{(0)}$, can be estimated under a fixed effects model neglecting the random effects. Second, the initial values for $\boldsymbol{u}$, $\boldsymbol{u}^{(0)}$, can be obtained by equating each y_i to its conditional expectation μ_i, and solving for u_i. And third, a plausible covariance function is fitted to the empirical variogram of $\boldsymbol{u}^{(0)}$ and the initial values for the covariance parameters, $\boldsymbol{\theta}^{(0)}$, are estimated by weighted least squares (Cressie, 1993). Some care is needed if the model contains count data with log link or binary data with logit link. In the case of count data, a solution is to add a small number to each observation. In the case of binary data, one may aggregate the observations over spatial subregions and use the mean frequencies of the aggregated data.

Ponciano et al. (2009) also described how to use the DC method to construct likelihood ratio (LR) tests to do hypothesis testing in hierarchical models in ecology as well as profile likelihood based confidence intervals. They introduced a straightforward algorithm to calculate LRs using DC. Therefore, we can use their algorithm to establish LR tests in SGLMMs.

4. Asymptotic properties

Walker (1969) proved that the posterior distribution asymptotically converges to a normal distribution centered at the MLE. By modifying Walker's theorems, Lele in an unpublished work proved that the DC-based distribution converges to a multivariate normal distribution. Also Lele et al. (2007) stated that the resulting estimators are invariant to the choice of priors as k increases.

Here, we present and prove these asymptotic results. The convergence of the DC-based distribution, under some regularity conditions, is treated in Theorem 1. Invariancy of inferences to the choice of priors is shown in Theorem 2. This theorem also shows that $\pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y})$ converges to a Dirac mass located at the MLE as k goes to infinity.

Suppose $\boldsymbol{\psi}$ denotes a q-dimensional vector in the parameter space $\Psi \in \Re^q$ and $\hat{\boldsymbol{\psi}}$ be an MLE. Let the log-likelihood function, $\log L_n(\boldsymbol{\psi}|\boldsymbol{y})$, which we shall denote by $\ell_n(\boldsymbol{\psi})$, leaving the argument $\boldsymbol{y}$ to be understood, possess a second derivative ℓ_n''. Clearly the subscript n in the log-likelihood function shows the number of observations. Also, let $\nu_n^2 = \{-\ell_n''(\hat{\boldsymbol{\psi}})\}^{-1}$ and $\ell_n^{(k)}(\boldsymbol{\psi}) = \log L_n^{(k)}(\boldsymbol{\psi}|\boldsymbol{y}) = \log[L_n(\boldsymbol{\psi}|\boldsymbol{y})]^k$.

Theorem 1. *Consider the q-dimensional version of regularity conditions of Walker (1969). Let $\boldsymbol{\psi}_0$, the true value of the parameter, denote an interior point of Ψ. Furthermore, let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be i.i.d. random variables from density function $f(y|\boldsymbol{\psi}_0)$ with respect to a σ-finite measure μ on the real line, with a joint density function $f_n(\boldsymbol{y}|\boldsymbol{\psi}_0)$ with respect to the product measure $\mu^{(n)} = \mu \times \cdots \times \mu$ in the Euclidean space $\Re^n$. Also, let $\mathbf{Y}^{(k)} = (\mathbf{Y}, \ldots, \mathbf{Y})$ denote the k-repeated cloned variables vector. If $\pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y})$ denotes the marginal DC-based density corresponding to the cloned log-likelihood $\ell_n^{(k)}(\boldsymbol{\psi})$, then*

$$\sqrt{k}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}})|\mathbf{y}^{(k)} \xrightarrow{D} N_q(0, \Gamma), \quad \text{as } k \longrightarrow \infty,$$

where $\Gamma = I^{-1}$ with

$$I = (\nu_{ij}) = \left(-\frac{\partial^2 \ell_n(\boldsymbol{\psi})}{\partial \psi_i \partial \psi_j}|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}} \right).$$

Proof. The proof is given in the Appendix. □

The next theorem shows that the DC-based estimators converge to the MLE as k increases.

Theorem 2 (*Lele et al., 2007*)**.** *Let $\pi(\boldsymbol{\psi})$ be positive everywhere on Ψ and $k \to \infty$. Then $\pi^{(k)}(\boldsymbol{\psi}|\mathbf{y})$ converges to a Dirac mass located at the MLE which is obtained independent of $\pi(\boldsymbol{\psi})$.*

Table 1
DC-based estimates (and scaled standard errors) of β_0 and β_1 in the uncorrelated binary data model, using IWLS and the DC algorithm with $k = 200$ and three different sets of priors $U(-4, 4)$ and $U(-4, 4)$, $N(0, 10)$ and $U(-9, 9)$, $N(0, 10)$ and $N(0, 10)$ for DC_1, DC_2 and DC_3, respectively.

Param.	True val.	MLE	DC_1	DC_2	DC_3
β_0	-2	-1.963 (0.189)	-2.086 (0.167)	-2.086 (0.184)	-2.086 (0.175)
β_1	0.75	0.811 (0.307)	0.666 (0.266)	0.666 (0.293)	0.667 (0.280)

5. Examples

5.1. Simulation examples

To assess the performance of the DC algorithm, we carried out two simulation examples. In the first example, we simulated $n = 30 \times 30$ uncorrelated binary data on an equally spaced regular grid of locations. This example is a simple GLM from which the MLE of the parameters can be obtained easily. In the second example, we simulated spatially correlated Poisson count data sets on two small and medium equally spaced regular grids of locations, 10×10 and 25×25. The computations were carried out via R software.

5.1.1. Uncorrelated binary data

Here, we simulated $n = 30 \times 30$ uncorrelated binary data on an equally spaced regular grid with the following model:

$$f(y_i|\boldsymbol{\beta}) = \left(\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{y_i}\left(1-\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{1-y_i},$$
$$g(\mu_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \eta_i = \beta_0 + \beta_1 x_i, \quad i = 1, \ldots, n,$$

where $p_i = E(Y_i|\boldsymbol{\beta}) = P(Y_i = 1|\boldsymbol{\beta})$ and $x_i = i/900$ for $i = 1, \ldots, 900$. In this example, parameters were fixed at $(\beta_0, \beta_1) = (-2, 0.75)$. This example serves as an initial test for the DC algorithm because the likelihood function for the model can be approximated accurately.

As Lele et al. (2007) have mentioned, the DC method produces the global maximum as k theoretically tends to infinity. But in practice k is finite and the asymptotic conditions of the theory cannot always be approximated. To avoid the possibility that the method gets stuck in a local maximum, they suggested rerunning of the algorithm with several different starting priors, and with increasing values of k, until the results from the different starting priors are in agreement.

For this simulated data, we calculated the exact MLE using iterative weighted least squares (IWLS) algorithm (McCullagh and Nelder, 1989). To avoid reaching a local maximum, we prepared three different sets of priors with $k = 200$ clones of the data and for each set of priors we also computed the DC-based distribution. The MLE of the parameters are applied as initial values. For each set, 5000 values were generated from the DC-based distributions, after a burn-in period of 1000 MCMC steps. The results, which are summarized in Table 1, are mostly similar for the three different sets of priors and are nearly equal to the exact ML inferences.

5.1.2. Spatial Poisson count data

In this example, we simulated two sets of spatial Poisson count data on small and medium equally spaced regular grids of locations, 10×10 and 25×25. Here, the model is as follows:

$$f(y_i|\beta) = \exp(y_i \ln(\mu_i) - \mu_i - \ln(y_i!)),$$
$$\ln(\mu_i) = \eta_i = \beta d_{1i} + u_i, \quad i = 1, \ldots, n,$$

where d_{1i} is the first component of ith location, i.e. $s_i = (d_{1i}, d_{2i})$ and $u_i = u(\mathbf{s}_i)$; $i = 1, \ldots, n$, is a realization of a zero mean GRF with isotropic exponential covariogram

$$C(\mathbf{s}-\mathbf{s}';\boldsymbol{\theta}) = \sigma^2 \exp\left(-\frac{\|\mathbf{s}-\mathbf{s}'\|}{\phi}\right),$$

in which $\boldsymbol{\theta} = (\sigma^2, \phi)$ and $\|\mathbf{s}-\mathbf{s}'\|$ denotes the Euclidean distance between $\mathbf{s}$ and $\mathbf{s}'$. In this example to have large observed counts, the parameters were fixed at $(\beta, \sigma^2, \phi) = (0.5, 1.25, 3)$ so that the maximum was 1 470 453.

To show that the DC-based estimators are invariant to the choice of priors, we used three different sets of the prior distributions. The prior distributions are considered to be independent. The first set included $N(0, 9)$, $U(0.1, 4)$ and $U(0.1, 5)$ for β, σ^2 and ϕ, respectively. The second and third sets included $U(-6, 6)$, $LN(0, 3)$, $LN(0, 3)$ and $N(0, 20)$, $IG(1, 4)$, $G(2, 3)$ for β, σ^2 and ϕ, respectively.

To implement our DC algorithm, we used the Metropolis–Hastings random walk sampler with the proposal densities $q(\beta^*|\beta) = N(\beta^{(t-1)}, G)$, $q(\sigma^{*2}|\sigma^2) = IG(\sigma^{(t-1)2}, 2)$ and $q(\phi^*|\phi) = LN(\phi^{(t-1)}, 2)$. In the proposal density $q(\beta^*|\beta)$, G is the scaled variance of a simple GLM, but the linear predictor η_i includes $\frac{1}{k}\sum_{j=1}^{k} \boldsymbol{u}_i^{(j)}$; $i = 1, \ldots, n$, as offsets. $\boldsymbol{\psi}^{(t-1)}$ are the

Table 2
DC-based estimates and scaled standard errors (SE) of β, σ^2 and ϕ in the spatial Poisson count data model with: exponential covariogram function, equally spaced grid designs 10×10 and 25×25, and priors $N(0, 9)$, $U(0.1, 4)$ and $U(0.1, 5)$ for parameters, respectively.

k	Param.	Grid 10×10			Grid 25×25		
	True val.	β 0.5	σ^2 1.25	ϕ 3	β 0.5	σ^2 1.25	ϕ 3
100	Est.	0.397	1.575	3.710	0.432	1.555	3.735
	SE	0.027	0.192	0.191	0.002	0.180	0.180
200	Est.	0.400	1.597	3.738	0.434	1.562	3.781
	SE	0.038	0.274	0.266	0.004	0.260	0.258
400	Est.	0.399	1.559	3.694	0.429	1.579	3.793
	SE	0.052	0.389	0.389	0.005	0.375	0.332
800	Est.	0.400	1.563	3.723	0.429	1.546	3.728
	SE	0.067	0.477	0.460	0.005	0.461	0.454

Table 3
DC-based estimates and scaled standard errors (SE) of β, σ^2 and ϕ in the spatial Poisson count data model with: exponential covariogram function, equally spaced grid designs 10×10 and 25×25, and priors $U(-6, 6)$, $LN(0, 3)$ and $LN(0, 3)$ for parameters, respectively.

k	Param.	Grid 10×10			Grid 25×25		
	True val.	β 0.5	σ^2 1.25	ϕ 3	β 0.5	σ^2 1.25	ϕ 3
100	Est.	0.392	1.652	3.801	0.430	1.688	3.765
	SE	0.028	0.235	0.177	0.003	0.221	0.190
200	Est.	0.399	1.647	3.695	0.432	1.630	3.804
	SE	0.038	0.338	0.297	0.004	0.312	0.255
400	Est.	0.391	1.654	3.782	0.432	1.651	3.784
	SE	0.056	0.482	0.377	0.005	0.440	0.373
800	Est.	0.392	1.686	3.803	0.430	1.636	3.789
	SE	0.068	0.577	0.442	0.005	0.548	0.463

Table 4
DC-based estimates and scaled standard errors (SE) of β, σ^2 and ϕ in the spatial Poisson count data model with: exponential covariogram function, equally spaced grid designs 10×10 and 25×25, and priors $N(0, 20)$, $IG(1, 4)$ and $G(2, 3)$ for parameters, respectively.

k	Param.	Grid 10×10			Grid 25×25		
	True val.	β 0.5	σ^2 1.25	ϕ 3	β 0.5	σ^2 1.25	ϕ 3
100	Est.	0.394	1.670	3.795	0.430	1.688	3.764
	SE	0.027	0.234	0.179	0.003	0.221	0.190
200	Est.	0.391	1.672	3.749	0.432	1.629	3.803
	SE	0.041	0.333	0.270	0.004	0.322	0.256
400	Est.	0.394	1.706	3.778	0.433	1.650	3.784
	SE	0.060	0.488	0.385	0.005	0.447	0.376
800	Est.	0.400	1.642	3.816	0.425	1.639	3.788
	SE	0.067	0.578	0.415	0.005	0.540	0.453

previous values of parameters in the current step of the algorithm. We also chose initial values of $(\boldsymbol{\psi}^{(0)}, \boldsymbol{u}^{(0)})$ as described in Section 3.

Table 2 presents the results for the first set of the prior distributions. They were obtained from retaining 2000 samples, every 5 iterations after an initial burn-in of 1000 iterations for 4 different k. The analysis was repeated with two other prior sets. The results are reported in Tables 3 and 4. In all the three sets of the priors nearly the same results were obtained.

According to the results we can state/represent the following two points:

1. The results give point estimates which are approximately close to the true values for all the parameters in a medium data set. The similar estimates can be obtained for a small data set. It is immediate to see that in a medium data set, the algorithm extracts a better estimate for β. Furthermore, it is clear that the estimates do not fluctuate as k increases.
2. The accuracy of the DC-based estimators seems dependent on the sample size as well as k. The standard errors of the parameter estimates for a medium sample size converge quickly as k increases. This is not the case for a small sample size and the standard errors still tend to increase as k increases. Then, it seems that the standard errors tend to be underestimated when sample size is small. Therefore to obtain reliable estimates of standard errors, k must be selected large enough.

Table 5
Acceptance rates of the DC algorithm for the first set of priors in the spatial Poisson count data model for $k = 100, 200, 400, 800$ and $n = 100, 625$.

n	k			
	100	200	400	800
100	0.283	0.272	0.275	0.276
625	0.462	0.462	0.471	0.463

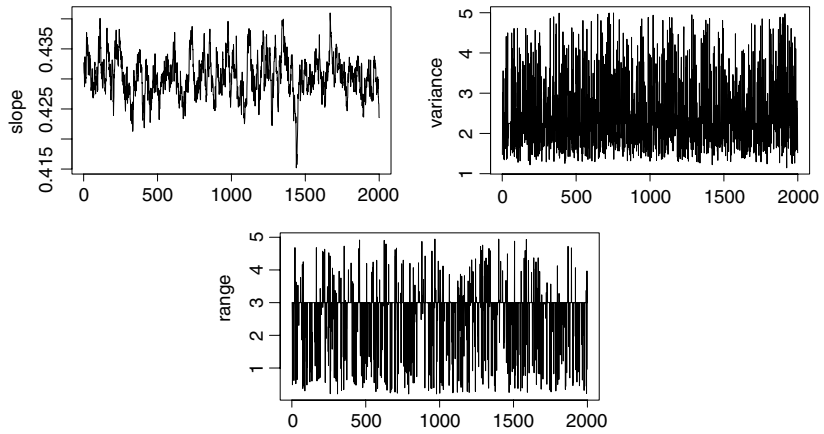

Fig. 1. Trace plots of the simulated draws of the parameters from the DC algorithm for the third set of priors in the spatial Poisson count data model with $n = 625$ and $k = 400$.

We found out that when n increases, acceptance rate of the DC algorithm increases as well. Furthermore, acceptance rates of the algorithm are approximately fixed for different k's. These findings are illustrated in Table 5. The table shows the acceptance rates of the DC algorithm for the first set of priors in the spatial Poisson count data model for $k = 100$, 200, 400, 800 and $n = 100$ and 625. For other prior sets the results remained the same. As we mentioned previously using other MCMC algorithms can be helpful to speed up the convergency of the algorithm.

To investigate the mixing behavior of the DC algorithm, we used the trace plots of the simulated draws of the parameters from the algorithm. Fig. 1 displays the trace plots for the third set of priors in the spatial Poisson count data model with $n = 625$ and $k = 400$. For other simulation sets, similar results were obtained. The figure confirms the fairly good performance of the algorithm. According to the DC algorithm described in Section 3, we used one block Metropolis–Hastings algorithm. As Browne et al. (2009) have noticed, the correlations between the random effects and the fixed effects will have more effect on the mixing of the chains. Then, using blocking MCMC algorithms can improve the mixing of the chain.

We also computed the effective sample size (ESS) of the simulated draws of the parameters (Kass et al., 1998) for different sets of our simulations. They also confirmed the good performance of the algorithm. Their values are not reported.

5.2. Car accident data

Iran is among those countries that suffer from very high yearly car accident casualty rates. Different human and environmental reasons contribute to this problem. Drivers' experience and their knowledge of traffic rules are two important human factors and the location type is a crucial environmental one.

The data set analyzed in this subsection consists of car accidents in Mashhad, the second biggest city in the north east of Iran, during 2006. 346 records were retained after data cleaning. The response is a binary variable recorded on its corresponding location either as 0 for only damages or as 1 for injury/fatality. In addition, for each location a number of explanatory variables such as driver's license type (DLT), a binary variable with levels class 1 and class 2, and location type such as street/intersection/highway/etc. are also recorded. It must be taken into account that in order to get a class 1 driver's license one needs a class 2 driver's license and thus people with class 1 are more experienced. Our analysis goal is to study the effects of these two important explanatory variables on the binary response. To study the effect of location type, indicator variables were used.

Fig. 2 shows the scatter of the data on map of Mashhad. According to the figure most of the accidents occurred in the center and the south. Furthermore, injuries or fatalities mostly happened in the streets and the highways in the downtown area. We used a model with Bernoulli data, logit link and five regression parameters, i.e., $\eta_i = \beta_0 + \beta_1 \mathrm{DLT}_i + \beta_2 \mathrm{str}_i + \beta_3 \mathrm{ints}_i + \beta_4 hw_i + u_i$. To obtain starting values, we used the procedure described in Section 3. The observations were fitted to a fixed effects model with initial values $\beta_0^{(0)} = -3.859$, $\beta_1^{(0)} = 1.058$, $\beta_2^{(0)} = 0.781$, $\beta_3^{(0)} = 0.387$ and $\beta_4^{(0)} = 0.453$. We also calculated an

Fig. 2. Car accident data. The figure shows the scatter of sample data on their related spatial locations. Light circles denote accidents with damage status and black stars denote accidents with injury/fatality status.

Table 6
DC-based inferences of the regression coefficients and GRF parameters for the car accidents in Mashhad.

Parameter	Estimate	SE	Est/SE
Intercept	−5.552	0.554	−10.022
DLT	1.534	0.265	5.787
Street	1.339	0.248	5.399
Intersection	0.775	0.270	2.870
Highway	0.999	0.323	3.093
σ^2	1.404	0.508	2.764
ϕ	0.790	0.291	2.715

empirical variogram of transformed residuals. The shape of the variogram suggested an exponential covariance function and a weighted least squares fit provided starting values $\sigma^{2(0)} = 0.649$ and $\phi^{(0)} = 0.266$ km.

To complete our setting and perform the DC method, similar to Section 5.1.2, the priors $N(0, 9)$ and $U(0.1, 4)$ were used for the regression coefficients and the correlation variables, respectively. Table 6 shows the results which were obtained from retaining 3000 samples every 10 iterations after an initial burn-in period of 2000 iterations for $k = 400$. According to the results, the effect of driver's experience on the response is clearly outstanding. This experience includes better driving skills and awareness of traffic rules. Drivers with less experience got involved in more accidents with injuries or fatalities. The results also show that accidents in the streets and the highways are more likely to cause injuries or fatalities. The estimated regression coefficients are interpretable based on the logarithmic scale of odds ratio.

6. Discussion

In this paper, we adapted the DC method to carry out statistical inference in SGLMMs. These models have commonly been fitted by Bayesian methods because frequentist inferences were computationally impractical. Now frequentist statistical inferences such as maximum likelihood parameter estimates, confidence intervals and hypotheses tests can be accomplished using the DC method in complex spatial models. And as Ponciano et al. (2009) have noticed, nowadays, the choice between Bayesian and frequentist approaches is no longer a matter of feasibility but rather can be based on the philosophical views of researchers. Here, we focused on small and medium data sets, which are obtained commonly, to show the applicability and reliability of the DC algorithm.

Although the DC method is a good alternative method for statistical inference in SGLMMs, but it has also some limitations. First, the method eventuates only Wald-type confidence intervals. Wald-type confidence intervals do not operate accurately for small samples, are not invariant to transformation of the parameters and often have less than nominal coverage rates (Ponciano et al., 2009). To overcome this limitation, Ponciano et al. (2009) proposed to use profile likelihood based confidence intervals which tend to have better statistical coverage rates than the Wald-type intervals and are invariant to transformation (Meeker and Escobar, 1995). An alternative method to surmount these disadvantages can be improving the standard errors via parametric bootstrapping, which needs further research.

Second, the method yields only MLE but not the actual maximized likelihood values. The maximized likelihood values are used to construct likelihood ratio tests, profile likelihood intervals and model selection criteria. As we informed in Section 3,

Ponciano et al. (2009) introduced a straightforward algorithm to calculate LR tests using the DC method as well as the differences in values of model selection criteria. Although they solved the problem of model selection by calculating the differences in values of information criteria, this does not assign a specific value like AIC or BIC to each model separately. Exploring to find a method to overcome this restriction can also be a topic of further research.

Third, k must be increased enough to increase the numerical accuracy in DC method and to stabilize the standard error estimates. This increases the computation time. According to Ponciano et al. (2009), constructing the profile likelihood using the DC method helps improve the numerical accuracy of the MLE and their standard errors.

Fourth, Lele et al. (2007) have noticed that when the data do not have information about the parameters in question then the likelihood based and Bayesian inferences both could be ill behaved. This is the case of complex spatial models with non-identifiable parameters. The DC method will not remedy ill-parameterized spatial models.

Fifth, in the most modern practical situations the special thing about SGLMMs is the high dimension of the random effects vector. Modern spatial data sets such as those collected by remote sensing can be very large. In such cases, MCMC approaches may be computationally very slow or even prohibitive. Then the DC method can also perform worse. Recently, a new approximate Bayesian inference method introduced by Rue and Martino (2007) called integrated nested Laplace approximation (INLA) substitutes MCMC simulations with accurate deterministic approximations. Eidsvik et al. (2009) and Hosseini et al. (2011) used it for inference and prediction in SGLMMs. They underlined that this type of approximate inference takes seconds of computation time and the quality of it is extremely high such that even very long MCMC runs could not detect any error in it. It would be expected that we can combine the DC algorithm by this new approximate method to reduce the computational efforts severely.

The DC method is also an adaptation of a computational ML approach developed by Robert (1993) called prior feedback. This method provides the MLE along with the calculation of standard errors, while prior feedback provides just MLE. The DC method also carries some similarity to the simulated annealing algorithm (Brooks and Morgan, 1995; Geyer and Thompson, 1995). But unlike the DC method, it requires calculation of the likelihood.

Theoretically, selection of k does not depend on n. But according to the simulation studies, as a guideline, for possessing the reliable estimates of standard errors, k must be chosen large enough when n is small. In general, selecting k under different circumstances such as sample size and number of parameters needs further research.

Acknowledgements

We wish to thank Professor Håvard Rue, NTNU, for his invaluable help and boundless patience and Rupali Akerkar for her help. Thanks are also due to two referees and the associate editor for their valuable comments and suggestions that greatly improved the previous version of this paper. Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is also acknowledged.

Appendix. Proof of Theorem 1

The proof uses the asymptotic results of Walker (1969). For simplicity, we prove Theorem 1 when $\boldsymbol{\psi}$ is a real parameter. The multivariate case is reasonably similar.

According to Walker (1969) the following results are easily extracted. To establish the results, we have only to note that $\boldsymbol{y}^{(k)}$, the k-repeated cloned data, is the response vector combined by k clones of the data and kn is considered instead of n in Walker (1969).

(i) Let $N_0(\delta) = \{\psi : |\psi - \psi_0| < \delta\}$ be a neighborhood of ψ_0 contained in Ψ. Then there exist a positive number $g(\delta)$, depending on δ, such that

$$\lim_{k\to\infty} P\left[\sup_{\psi\in\Psi-N_0(\delta)} k^{-1}\left\{\ell_n^{(k)}(\psi) - \ell_n^{(k)}(\psi_0)\right\} < -g(\delta)\right] = 1. \tag{A.1}$$

(ii) Defining $J(\psi_0) = \int (\frac{\partial \ell_n(\psi_0)}{\partial \psi_0})^2 f(\mathbf{y}|\psi_0)\mathrm{d}\mu^{(n)}$, expected Fisher information matrix, where $0 < J(\psi_0) < \infty$, then

$$k^{-1}\ell_n^{''(k)}(\hat{\psi}) \xrightarrow{P} -J(\psi_0). \tag{A.2}$$

(iii)

$$\ell_n^{(k)}(\psi_0) - \ell_n^{(k)}(\hat{\psi}) = O_p(1). \tag{A.3}$$

We write

$$\begin{aligned} L_n^{(k)}(\psi;\boldsymbol{y}) &= L_n^{(k)}(\hat{\psi};\boldsymbol{y}) \exp\left\{\ell_n^{(k)}(\psi) - \ell_n^{(k)}(\hat{\psi})\right\} \\ &= L_n^{(k)}(\hat{\psi};\boldsymbol{y}) \exp\left\{k\left[-\frac{(\psi-\hat{\psi})^2}{2v_n^2}\right](1+R_n)\right\} \end{aligned} \tag{A.4}$$

where, with respect to the extension $\ell_n^{(k)}(\psi) = \ell_n^{(k)}(\hat{\psi}) + \frac{1}{2}(\psi - \hat{\psi})^2 \ell_n^{''(k)}(\psi^*)$ with $\psi^* \in (\psi, \hat{\psi})$, $R_n = \nu_n^2(\ell_n''(\psi^*) - \ell_n''(\hat{\psi}))$ whenever ψ and $\hat{\psi}$ belong to the neighborhood of ψ_0.

Now we first calculate $m_n^{(k)}(\boldsymbol{y}) = \int \pi^{(k)}(\psi|\boldsymbol{y})\mathrm{d}\psi$ over two sets $S_1 = \Psi - N_0(\delta)$ and $S_2 = N_0(\delta)$. We have from (A.4),

$$B_1 = L_n^{(k)}(\hat{\psi}; \boldsymbol{y}) \exp\left\{\ell_n^{(k)}(\psi_0) - \ell_n^{(k)}(\hat{\psi})\right\} \int_{S_1} \pi(\psi) \exp\left\{\ell_n^{(k)}(\psi) - \ell_n^{(k)}(\psi_0)\right\} \mathrm{d}\psi. \tag{A.5}$$

From (A.1), the integral in (A.5) is less than

$$\exp\{-kg(\delta)\} \int_{S_1} \pi(\psi)\mathrm{d}\psi \leq \exp\{-kg(\delta)\}$$

in probability. From (A.3),

$$\exp\left\{\ell_n^{(k)}(\psi_0) - \ell_n^{(k)}(\hat{\psi})\right\} = O_p(1),$$

and from (A.2),

$$\sqrt{k}\nu_n^{-1} \exp\{-kg(\delta)\} \xrightarrow{P} \{J(\psi_0)\}^{\frac{1}{2}} \sqrt{k} \exp\{-kg(\delta)\} \longrightarrow 0.$$

Therefore,

$$\left[L_n^{(k)}(\hat{\psi}; \boldsymbol{y})\sqrt{k}\nu_n\right]^{-1} B_1 \xrightarrow{P} 0. \tag{A.6}$$

For S_2 we have,

$$B_2 = L_n^{(k)}(\hat{\psi}; \boldsymbol{y}) \int_{S_2} \pi(\psi) \exp\left\{k\left[\frac{(\psi - \hat{\psi})^2}{2\nu_n^2}\right](1 + R_n)\right\} \mathrm{d}\psi.$$

With respect to continuity and positivity of the prior density at $\psi = \psi_0$, given any $\epsilon > 0$, there exists a δ such that if $\psi \in N_0(\delta)$, then

$$|\pi(\psi) - \pi(\psi_0)| < \epsilon\pi(\psi_0). \tag{A.7}$$

Hence,

$$(1 - \epsilon)B_3 < \{\pi(\psi_0)L_n^{(k)}(\hat{\psi}; \boldsymbol{y})\}^{-1}B_2 < (1 + \epsilon)B_3, \tag{A.8}$$

where,

$$B_3 = \int_{S_2} \exp\left\{k\left[-\frac{(\psi - \hat{\psi})^2}{2\nu_n^2}\right](1 + R_n)\right\} \mathrm{d}\psi. \tag{A.9}$$

Furthermore,

$$\begin{aligned}
\left|(k\nu_n^2)^{-1}R_n\right| &= \left|k^{-1}\left\{\ell_n''(\psi^*) - \ell_n''(\hat{\psi})\right\}\right| \\
&\leq \left|k^{-1}\left\{\ell_n''(\psi^*) - \ell_n''(\psi_0)\right\}\right| + \left|k^{-1}\left\{\ell_n''(\hat{\psi}) - \ell_n''(\psi_0)\right\}\right| \\
&< 2k^{-1}\sum_{j=1}^{k} M_\delta(\boldsymbol{y}^{(j)}|\psi_0) \quad \text{if } |\hat{\psi} - \psi_0| < \delta,
\end{aligned}$$

where $\boldsymbol{y}^{(j)}$ is the jth cloned observations of $\boldsymbol{y}^{(k)}$. Then if δ is chosen such that $E\{M_\delta(\boldsymbol{y}^{(j)}|\psi_0)\} < \frac{\epsilon/4}{J(\psi_0)}$,

$$\lim_{k\to\infty} P\{\sup_{\psi\in S_2} |R_n| < \epsilon\} = 1. \tag{A.10}$$

Hence, when $\epsilon < 1$,

$$\begin{aligned}
\int_{S_2} \exp\left\{k\left[-\frac{(\psi - \hat{\psi})^2}{2\nu_n^2}\right](1 \pm \epsilon)\right\} \mathrm{d}\psi = \sqrt{2\pi}\frac{\nu_n}{\sqrt{(1 \pm \epsilon)k}}\Bigg[\Phi\left\{\frac{\sqrt{k(1 \pm \epsilon)}}{\nu_n}(\psi_0 + \delta - \hat{\psi})\right\} \\
- \Phi\left\{\frac{\sqrt{k(1 \pm \epsilon)}}{\nu_n}(\psi_0 - \delta - \hat{\psi})\right\}\Bigg],
\end{aligned} \tag{A.11}$$

where Φ denotes the standard normal distribution function. The expression in square brackets in the above formula converges to 1 in probability. Hence from (A.9) and (A.10),

$$\lim_{k\to\infty} P\left[\sqrt{2\pi}\frac{\nu_n}{\sqrt{k}}(1+\epsilon)^{-1/2} < B_3 < \sqrt{2\pi}\frac{\nu_n}{\sqrt{k}}(1-\epsilon)^{-1/2}\right] = 1. \tag{A.12}$$

Since δ can be chosen such that (A.8) and (A.12) hold for arbitrarily small ϵ, we deduce that

$$\lim_{k\to\infty} P\left[\frac{\sqrt{2\pi}\nu_n(1-\xi)}{\sqrt{k}} < \frac{B_2}{\{\pi(\psi_0)L_n^{(k)}(\hat{\psi};\boldsymbol{y})\}} < \frac{\sqrt{2\pi}\nu_n(1+\xi)}{\sqrt{k}}\right] = 1,$$

for arbitrarily small positive ξ. In conjunction with (A.6) this shows that

$$\frac{m_n^{(k)}(\boldsymbol{y})}{L_n^{(k)}(\hat{\psi};\boldsymbol{y})k^{-1/2}\nu_n} = \sqrt{2\pi}\pi(\psi_0).$$

Now, let

$$B_4 = \int_{\hat{\psi}+a\frac{\nu_n}{\sqrt{k}}}^{\hat{\psi}+b\frac{\nu_n}{\sqrt{k}}} \pi(\psi)L_n^{(k)}(\psi;\boldsymbol{y})\mathrm{d}\psi,$$

which is equal to B_2 with S_2 replaced by $(\hat{\psi}+a\frac{\nu_n}{\sqrt{k}}, \hat{\psi}+b\frac{\nu_n}{\sqrt{k}})$. Since clearly

$$\lim_{k\to\infty} P\left[\left(\hat{\psi}+a\frac{\nu_n}{\sqrt{k}}, \hat{\psi}+b\frac{\nu_n}{\sqrt{k}}\right) \subset N_0(\delta)\right] = 1,$$

this replacement can be made in some of the above expressions, the only difference is that now (A.7) and (A.8) hold in probability. Instead of (A.11) we then get

$$\sqrt{2\pi}\frac{\sqrt{k(1\pm\epsilon)}}{\nu_n}[\Phi\{b\sqrt{1\pm\epsilon}\} - \Phi\{a\sqrt{1\pm\epsilon}\}].$$

We thus obtain,

$$\frac{B_4}{L_n^{(k)}(\hat{\psi};\boldsymbol{y})\frac{\nu_n}{\sqrt{k}}} \xrightarrow{P} \sqrt{2\pi}\{\Phi(b)-\Phi(a)\}\pi(\psi_0).$$

Hence,

$$\int_{\hat{\psi}+a\frac{\nu_n}{\sqrt{k}}}^{\hat{\psi}+b\frac{\nu_n}{\sqrt{k}}} \pi^{(k)}(\psi|\boldsymbol{y})\mathrm{d}\psi = \frac{B_4}{m_n^{(k)}(\boldsymbol{y})} \xrightarrow{P} \Phi(b)-\Phi(a).$$

References

Booth, J.G., Hobert, J.P., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society, Series B 61, 265–285.

Breslow, N., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, 9–25.

Breslow, N.E., Lin, X., 1995. Bias correction in generalised linear mixed models with a single component of dispersion. Biometrika 8, 81–91.

Brooks, S.P., Morgan, B.J.T., 1995. Optimization using simulated annealing. Statistician 44, 241–257.

Browne, W.J., Steele, F., Golalizadeh, M., Green, M.J., 2009. The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. Journal of the Royal Statistical Society, Series A 172, 579–598.

Christensen, O.F., 2004. Monte Carlo maximum likelihood in model-based geostatistics. Journal of Computational and Graphical Statistics 13, 702–718.

Christensen, O.F., Roberts, G.O., Skold, M., 2006. Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. Journal of Computational and Graphical Statistics 15, 1–17.

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

Cressie, N., 1993. Statistics for Spatial Data, second ed. Wiley, New York.

Diggle, P., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistic. Applied Statistics 47, 299–350.

Eidsvik, J., Martino, S., Rue, H., 2009. Approximate Bayesian inference in spatial generalized linear mixed models. Scandinavian Journal of Statistics 36, 1–22.

Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo maximum likelihood for dependent data (with discussion). Journal of the Royal Statistical Society, Series B 54, 657–699.

Geyer, C.J., Thompson, E.A., 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. Journal of the American Statistical Association 90, 909–920.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Heagerty, P., Lele, S.R., 1998. A composite likelihood approach to binary data in space. Journal of the American Statistical Association 93, 1099–1111.

Hosseini, F., Eidsvik, J., Mohammadzadeh, M., 2011. Approximate Bayesian inference in spatial GLMM with skew normal latent variables. Computational Statistics and Data Analysis 55, 1791–1806.

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R., 1998. Markov chain Monte Carlo in practice: a roundtable discussion. American Statistician 52, 93–100.

Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. Journal of the American Statistical Association 91, 1343–1370.

Lele, S.R., Dennis, B., Lutscher, F., 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecology Letters 10, 551–563.
Lesaffre, E., Speissens, B., 2001. On the effect of the number of quadrature points in a logistic random-effects model: an example. Applied Statistics 50, 325–335.
Lindsay, B., 1988. Composite likelihood methods. Contemporary Mathematics 80, 221–239.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, London.
McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association 92, 162–170.
Meeker, W., Escobar, L.A., 1995. Teaching about approximate confidence regions based on maximum likelihood estimation. American Statistician 49, 48–53.
Natarajan, R., McCulloch, C.E., 1995. A note on the existence of the posterior distribution for a class of mixed models for binomial responses. Biometrika 82, 639–643.
Pinheiro, J.C., Bates, D.M., 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics 4, 12–35.
Ponciano, J.M., Taper, M.L., Dennis, B., Lele, S.R., 2009. Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. Ecology 90, 356–362.
Robert, C.P., 1993. Prior feedback: Bayesian tools for maximum likelihood estimation. Journal of Computational Statistics 8, 279–294.
Robert, C.P., Casella, G., 2005. Monte Carlo Statistical Methods, second ed. Springer-Verlag.
Rodriguez, G., Goldman, N., 1995. An assessment of estimation procedures for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A 158, 73–89.
Rue, H., Martino, S., 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. Journal of Statistical Planning and Inference 137, 3177–3192.
Skaug, H.J., 2002. Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. Journal of Computational and Graphical Statistics 11, 458–470.
Varin, C., Gudmund, H., Skare, O., 2005. Pairwise likelihood inference in spatial generalized linear mixed models. Computational Statistics and Data Analysis 49, 1173–1191.
Walker, A.M., 1969. On the asymptotic behaviour of posterior distributions. Journal of the Royal Statistical Society, Series B 31, 80–88.
Zhang, H., 2002. On estimation and prediction for spatial generalized linear mixed models. Biometrics 58, 129–136.
Zhao, Y., Staudenmayer, J., Coull, B.A., Wand, M.P., 2006. General design Bayesian generalized linear mixed models with applications to spatial statistics. Statistical Science 21, 35–51.