NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

On a Hybrid Data Cloning Method and Its Application in Generalized Linear Mixed Models

by

Hossein Baghishani, Håvard Rue and Mohsen Mohammadzadeh

PREPRINT
STATISTICS NO. 8/2010

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2010/S8-2010.pdf
Håvard Rue has homepage: http://www.math.ntnu.no/∼hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

On a Hybrid Data Cloning Method and Its Application in Generalized Linear Mixed Models

Hossein Baghishani[a,*], Håvard Rue[b], Mohsen Mohammadzadeh[a]

[a] *Department of Statistics, Tarbiat Modares University, Tehran, Iran*
[b] *The Norwegian University of Science and Technology, Trondheim, Norway*

Abstract

Data cloning method is a new computational tool for computing maximum likelihood estimates in complex statistical models such as mixed models. The data cloning method is synthesized with integrated nested Laplace approximation to compute maximum likelihood estimates efficiently via a fast implementation in generalized linear mixed models. Asymptotic normality of the hybrid data cloning based distribution is established aided by modification of Stein's Identity. The results are illustrated through a series of well known examples. It is shown that the proposed method as well as normal approximation perform very well and justify the theory.

Keywords: *Approximate Bayesian Inference, Asymptotic Normality, Data Cloning, Generalized Linear Mixed Models, Integrated Nested Laplace Approximation, Stein's Identity.*

---

1. Introduction

Non-Gaussian repeated measurements such as longitudinal and clustered data are common in many sciences such as biology, ecology, epidemiology and medicine. The Generalized Linear Mixed Models are flexible models for modeling these types of data. 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 latent variables, while the latent variables are modeled by random effects that are typically Gaussian (Breslow and Clayton, 1993).

*To whom correspondence should be addressed.
Email addresses: hbaghishani@modares.ac.ir, hossein.baghishani@math.ntnu.no (Hossein Baghishani), hrue@math.ntnu.no (Håvard Rue), mohsen_m@modares.ac.ir (Mohsen Mohammadzadeh)

April 30, 2010

Statistical inferences in such models has been the subject of a great deal of research over the past decade. Both frequentist and Bayesian methods have been developed for inference in GLMMs (McCulloch, 1997). Due to the advances in computation, the most commonly used approach for inference in these models is based on the Bayesian paradigm, especially Markov chain Monte Carlo (MCMC) algorithms. But Bayesian inferences depend on the choice of the prior distributions and the specification of prior distributions is not straightforward in particular for variance components (Fong *et al.*, 2009). Moreover, MCMC algorithms applied to these models come with a wide range of problems in terms of convergence and computational time.

A recent suitable alternative method to carry out likelihood based inference in GLMM can be the data cloning (DC) method, which was first introduced by Lele *et al.* (2007) in ecological studies. DC uses an MCMC algorithm from an artificial constructed distribution, named DC-based distribution, to compute maximum likelihood estimates (MLE) and their variance estimates. Although the distribution looks like a Bayesian posterior distribution, but 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 (Baghishani and Mohammadzadeh, 2009). The trick in the DC is generating samples from a DC-based distribution constructed by duplicating the original data set enough times, k say, such that the sample mean as well as the scaled sample variance converge to MLE and its variance estimate. Computation, however, is an issue since the usual implementation is via MCMC.

It is the main purpose of this paper to describe how DC method may be synthesized with integrated nested Laplace approximation (INLA) introduced by Rue and Martino (2007) and Rue *et al.* (2009), using Stein's Identity proposed by Weng and Tsai (2008), to compute MLE efficiently. It is also the aim to establish the asymptotic normality of the new hybrid DC-based distribution in a GLMM.

In the next section we describe the model and INLA methodology. Also, we represent Stein's Identity. Section 3 describes the new hybrid DC method. In Section 4, the performance of the method is explored through a series of well known examples. All technical details for establishing the asymptotic normality of the hybrid DC-based distribution are presented in Section 5. Finally, Section 6 concludes with a brief discussion.

2. Model and INLA

In this section we introduce the our basic model and INLA methodology. We also represent Stein's Identity to use for establishing the asymptotic normality of the hybrid DC-based distribution.

2.1. The Model

Generalized linear mixed models are flexible models for modeling non-Gaussian repeated measurements. On the basis of the GLM, the GLMM assumes that the responses are independent conditional on the random effects and are distributed according to a member of the exponential family.

We consider clustered data in which repeated measures of a response variable are taken on a random sample of m clusters. Consider the response vectors $\boldsymbol{y}_i = (y_{i1}, \dots, y_{in_i})^T$, $i = 1, \dots, m$. Let $n = \sum_{i=1}^{m} n_i$ be the total sample size. Conditional on $r \times 1$ vector of unobservable cluster-specific random effects $\boldsymbol{u}_i = (u_{i1}, \dots, u_{ir})^T$, 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 $i = 1, \dots, m; j = 1, \dots, n_i$, in which $\boldsymbol{x}_{ij}$ and $\boldsymbol{v}_{ij}$, are the corresponding p- and r-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 $g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{v}_{ij}^T\boldsymbol{u}_i$, where $g(\cdot)$ is a monotonic link function. Furthermore, assume $\boldsymbol{u}_i$ comes from a Gaussian distribution, $\boldsymbol{u}_i|Q \sim N(\boldsymbol{0}, Q^{-1})$, in which the precision matrix $Q = Q(\boldsymbol{\theta})$ depends on parameters $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}$ denote the $d \times 1$ vector of the variance components for which prior $\pi(\boldsymbol{\theta})$ is assigned. We further assume that $\boldsymbol{\beta}$ is assigned a normal prior distribution, $\pi(\boldsymbol{\beta})$, with known hyperparameters. Let also $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{u})$ denote the $q \times 1$ vector of parameters assigned Gaussian priors. Moreover, let $\boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_m)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, an open subset of $\Re^d$, and $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, an open subset of $\Re^q$. Now, the posterior density is defined by

$$\pi(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\beta})|Q(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{u}^T Q(\boldsymbol{\theta})\boldsymbol{u} + \sum_{i=1}^{m} \log f(\boldsymbol{y}_i|\psi)\right\}. \quad (1)$$

Then, we can write

$$\log \pi(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \log \pi(\boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta}|\boldsymbol{y}) = \ell_n(\boldsymbol{\psi}) + \ell_n(\boldsymbol{\theta}).$$

2.2. Integrated Nested Laplace Approximation

Because of the usefulness and easy implementation of the MCMC methods, the most commonly used approach for inference in the GLMMs is based on Bayesian methods and MCMC sampling. Considering (1) the main aim is to compute the posterior marginals $\pi(\psi_l|\boldsymbol{y})$, $l = 1, \ldots, q$ and $\pi(\theta_v|\boldsymbol{y})$, $v = 1, \ldots, d$. It is well known, however, that MCMC methods tend to exhibit poor performance when applied to such models (Rue *et al.*, 2009).

INLA is a new tool for Bayesian inference on latent Gaussian models introduced by Rue *et al.* (2009). The method combines Laplace approximations and numerical integration in a very efficient manner. INLA substitutes MCMC simulations with accurate, deterministic approximations to posterior marginal distributions. The quality of such approximations is high in most cases such that even very long MCMC runs could not detect any error in them.

We can write

$$\begin{aligned} \pi(\psi_l|\boldsymbol{y}) &= \int \pi(\psi_l|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}, \\ \pi(\theta_v|\boldsymbol{y}) &= \int \pi(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-v}, \end{aligned}$$

and the key feature of INLA is to use this form to construct nested approximations

$$\begin{aligned} \tilde{\pi}(\psi_l|\boldsymbol{y}) &= \int \tilde{\pi}(\psi_l|\boldsymbol{\theta}, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}, \\ \tilde{\pi}(\theta_v|\boldsymbol{y}) &= \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-v}, \end{aligned}$$

where Laplace approximation is applied to carry out the integrations required for evaluation of $\tilde{\pi}(\psi_l|\boldsymbol{\theta}, \boldsymbol{y})$. The approximate posterior marginals obtained from INLA can then be used to compute summary statistics of interest, such as posterior means, variances or quantiles.

2.3. Stein's Identity

This subsection represents Stein's Identity. We use it to establish the asymptotic normality of the hybrid DC-based distribution in Section 5. For a detailed account of Stein's Identity, we refer readers to Woodroofe (1989).

Let Γ be a finite signed measure of the form $d\Gamma = f d\Phi_r$ in which f is a real-valued function defined on $\Re^r$ satisfying $\int |f| d\Phi_r < \infty$. Write $\Phi_r h = \int h d\Phi_r$ for functions h for which the integral is finite and also write $\Gamma h = \int h d\Gamma$. 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

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

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

Lemma 1. *(Weng and Tsai, 2008). Let s be a nonnegative integer and let $d\Gamma = f d\Phi_r$, 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 $h \in H_s$.

3. A Hybrid Data Cloning Method

A recent suitable alternative method to carry out likelihood based inference in GLMMs is the DC method, which was first introduced by Lele *et al.* (2007). DC uses, as a computational trick, an MCMC algorithm to compute MLE and their variance estimates. The trick in DC is to apply an MCMC algorithm to a data set constructed by duplicating the original data set enough times, k say, that the resulting estimates converge to MLE.

Let $\pi^{(k)}(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta})\pi^{(k)}(\boldsymbol{\theta}|\boldsymbol{y})$ be the artificial constructed joint density, which we call it joint DC-based density, from k identical and independent clones of the data and prior distributions, $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\beta})$. According to Baghishani and Mohammadzadeh (2009),

$$\begin{aligned} \mathrm{E}^{(k)}(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) &\overset{k \to \infty}{\longrightarrow} (\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}), \\ \mathrm{Var}^{(k)}(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) &\overset{k \to \infty}{\longrightarrow} \sqrt{k} \times \mathrm{Var}((\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})). \end{aligned}$$

Following, we will combine the DC method with INLA. We first present the asymptotic normality of the DC-based distribution. Let $\boldsymbol{y}^{(k)} = (\boldsymbol{y}, \ldots, \boldsymbol{y})$ denote the k-repeated cloned vector of the data. Then,

$$\log \pi^{(k)}(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \log \pi^{(k)}(\boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta}) + \log \pi^{(k)}(\boldsymbol{\theta}|\boldsymbol{y}) = \ell_n^{(k)}(\boldsymbol{\psi}) + \ell_n^{(k)}(\boldsymbol{\theta}).$$

Now we define $F_{n,k}$, $G_{n,k}$, $\boldsymbol{z}_{n,k}$ and $\boldsymbol{w}_{n,k}$ as follow:

$$\begin{aligned} F_{n,k}^T F_{n,k} &= -\nabla^2 \ell_n^{(k)}(\hat{\boldsymbol{\psi}}_n), \quad \boldsymbol{z}_{n,k} = F_{n,k}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \\ G_{n,k}^T G_{n,k} &= -\nabla^2 \ell_n^{(k)}(\hat{\boldsymbol{\theta}}_n), \quad \boldsymbol{w}_{n,k} = G_{n,k}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \end{aligned}$$

Therefore,

$$\pi_n^{(k)}(\boldsymbol{z}_{n,k}, \boldsymbol{w}_{n,k}|\boldsymbol{y}) \propto \pi_n^{(k)}(\boldsymbol{\psi}(\boldsymbol{z}_{n,k}), \boldsymbol{\theta}(\boldsymbol{w}_{n,k})|\boldsymbol{y}) \propto e^{\ell_n^{(k)}(\boldsymbol{\psi}) - \ell_n^{(k)}(\hat{\boldsymbol{\psi}}_n)} e^{\ell_n^{(k)}(\boldsymbol{\theta}) - \ell_n^{(k)}(\hat{\boldsymbol{\theta}}_n)}.$$

Theorem 1 below shows the asymptotic distribution of the DC-based distribution is normal.

Theorem 1. *Suppose that h be any bounded measurable function as in Lemma 1 and $k \to \infty$. Moreover, suppose that the prior $\pi(\boldsymbol{\theta})$ satisfies (F1)-(F3), $\ell_n^{(k)}(\boldsymbol{\theta})$ and $\ell_n^{(k)}(\boldsymbol{\psi})$ satisfies appropriate conditions, similar to conditions (B1)-(B4) and (C1)-(C4) replacing n with nk. Then, $E_{nk}^c[h(\boldsymbol{z}_{n,k}, \boldsymbol{w}_{n,k})] \xrightarrow{p} \Phi h$.*

3.1. The Proposed Hybrid Method

As Rue *et al.* (2009) have mentioned, MCMC methods tend to exhibit poor performance when applied to GLMMs. Then the DC method can also acts worse. It would be expected that synthesizing of the DC method by INLA can reduces the computational efforts severely.

Combining obtained results of Theorem 1 with INLA methodology, we can establish the asymptotic normality of the new hybrid DC-based distribution. Let $\tilde{\ell}_n^{(k)}(\boldsymbol{\theta})$ and $\tilde{\ell}_n^{(k)}(\boldsymbol{\psi})$ be the corresponding approximates of $\ell_n^{(k)}(\boldsymbol{\theta})$ and $\ell_n^{(k)}(\boldsymbol{\psi})$ respectively obtained by INLA. Then,

$$\log \tilde{\pi}^{(k)}(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) = \tilde{\ell}_n^{(k)}(\boldsymbol{\theta}) + \tilde{\ell}_n^{(k)}(\boldsymbol{\psi}).$$

Let also

$$Q_{n,k}^T Q_{n,k} = -\nabla^2 \tilde{\ell}_n^{(k)}(\hat{\boldsymbol{\psi}}_n), \quad \boldsymbol{x}_{n,k} = Q_{n,k}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \tag{2}$$

$$V_{n,k}^T V_{n,k} = -\nabla^2 \tilde{\ell}_n^{(k)}(\hat{\boldsymbol{\theta}}_n), \quad \boldsymbol{s}_{n,k} = V_{n,k}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \tag{3}$$

Now we can state the following theorem.

Theorem 2. *Suppose that h be any bounded measurable function as in Lemma 1 and $k \to \infty$. Moreover, suppose that the prior $\pi(\boldsymbol{\theta})$ satisfies $(F1)$-$(F3)$, $\tilde{\ell}_n^{(k)}(\boldsymbol{\theta})$ and $\tilde{\ell}_n^{(k)}(\boldsymbol{\psi})$ satisfies appropriate conditions replacing n with nk. Then, $\tilde{E}^c_{nk}[h(\boldsymbol{x}_{n,k}, \boldsymbol{s}_{n,k})] \xrightarrow{p} \Phi h$.*

4. Examples

To illustrate the performance of the proposed hybrid method, we consider three examples by which both nested (Subsection 4.1) and crossed (Subsections 4.2 and 4.3) random effects are introduced. These examples have been considered previously by Breslow and Clayton (1993) and Fong *et al.* (2009). The computations are carried out by R INLA package (`www.r-inla.org`).

Fong *et al.* (2009) analyzed these examples by using INLA methodology and gave a number of prescriptions for prior specification especially for variance components. They also noticed that sometimes specification of a prior for variance components is not straightforward. But DC-based results are invariant to the choice of the priors.

4.1. Overdispersion

This example concerns data on the proportion of seeds that germinated on each of $m = 21$ plates arranged according to a 2×2 factorial design with respect to seed variety and type of root extract (Crowder, 1978). The sampling model is $Y_i|\boldsymbol{\beta}, p_i \sim Bin(n_i, p_i)$ where, for plate i, y_i is the number of germinating seeds and n_i is the total number of seeds for $i = 1, \ldots, m$. To account for the extraneous between plate variability, Breslow and Clayton (1993) introduced plate-level random effects and then fitted two main effects and interaction models:

$$\begin{aligned} logit(p_i) &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \qquad (4) \\ logit(p_i) &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + u_i, \end{aligned}$$

in which $u_i|\sigma^2 \overset{iid}{\sim} N(0, \sigma^2)$, and x_{1i}, x_{2i} represent the seed variety and type of root extract for plate i.

To show that the DC-based estimators are invariant to the choice of the priors, we used four different sets of prior distributions. Following Fong *et al.* (2009), the first set include $N(0, 10)$ for fixed effects and $Ga(0.5, 0.0164)$ for σ^{-2} in the main effects model and $N(1, 100)$ for fixed effects and $Ga(0.5, 0.0164)$ for σ^{-2} in the interaction effects model. The second, third and fourth sets for the main effects model include $N(1, 100)$ and $Ga(0.25, 0.1)$, $N(-2, 1)$, $N(-1, 1)$ and $Ga(0.75, 0.002)$,

Table 1: MLEs and DC-based estimates obtained by hybrid DC method with $k = 200$ for Seed data in the main effects model. [a] From Breslow and Clayton (1993). Standard errors are in brackets.

Par.	MLE[a]	DC_1	DC_2	DC_3	DC_4
Intercept	-0.389 (0.166)	-0.389 (0.165)	-0.389 (0.165)	-0.389 (0.165)	-0.389 (0.166)
Seed	-0.347 (0.215)	-0.346 (0.212)	-0.346 (0.212)	-0.346 (0.212)	-0.347 (0.212)
Extract	1.029 (0.205)	1.029 (0.204)	1.029 (0.204)	1.029 (0.204)	1.029 (0.204)
σ	0.295 (0.112)	0.293 (0.110)	0.293 (0.110)	0.293 (0.110)	0.294 (0.110)

and $N(0, 100)$ and $Ga(0.25, 0.3)$ respectively. The second, third and fourth sets for the interaction effects model also include $N(-1, 1)$, $N(-2, 1)$, $N(-1, 1)$ and $Ga(0.25, 0.1)$, $N(-2, 100)$, $N(2, 100)$, $N(1, 100)$ and $Ga(0.75, 0.002)$, and $N(2, 1000)$ and $Ga(0.25, 0.4)$ respectively.

To implement the method using R INLA package, we first prepared cloned data by duplicating the original data. Notice that the number of plates for cloned data, $plate.k$, is $m \times k = 21k$. Let $x.k_{1i}$ and $x.k_{2i}$ denote the new cloned covariates. Let also $r.k_i$ and $n.k_i$ denote the proportion and the total number of seeds for $i = 1, \dots, 21k$ respectively. Fitting the model (4), say, is done by calling the *inla()* function:

```
> formula = r.k ~ x1.k+x2.k+f(plate.k,model="iid",param=c(.5,.0164))
> result = inla(formula,data=clone.data,family="binomial",Ntrials=n.k)
```

Tables 1 and 2 present the results obtained by hybrid DC method with $k = 200$ for four different prior sets in the main and interaction effects models respectively. These results are compared with MLEs obtained by Breslow and Clayton (1993) using Gaussian quadrature. There is surprisingly very close correspondence between the MLE and obtained results from hybrid DC method for different priors. For other several prior sets the results, which are not reported, remained the same as well. Furthermore, marginal DC-based distributions converge to Normal distributions almost exactly. These findings are illustrated in Figure 1 for interaction effect and precision parameter of random effect in the interaction model. For other fixed effects the densities are indistinguishable for different priors.

4.2. Longitudinal Data

Epilepsy data of Thall and Vail (1990) are a well known dataset that was analyzed several times by various authors. They presented data from a clinical trial of 59 epileptics who were randomized to a new drug (Trt=1) or a placebo (Trt=0). Baseline data available at entry into

Table 2: MLEs and DC-based estimates obtained by hybrid DC method with $k = 200$ for Seed data in the interaction effects model. [a] From Breslow and Clayton (1993). Standard errors are in brackets.

Par.	MLE[a]	DC_1	DC_2	DC_3	DC_4
Intercept	-0.548 (0.167)	-0.548 (0.166)	-0.548 (0.166)	-0.548 (0.166)	-0.548 (0.166)
Seed	0.097 (0.278)	0.097 (0.276)	0.097 (0.276)	0.097 (0.276)	0.096 (0.277)
Extract	1.337 (0.237)	1.337 (0.235)	1.337 (0.236)	1.337 (0.235)	1.337 (0.237)
Interaction	-0.811 (0.385)	-0.810 (0.383)	-0.810 (0.383)	-0.810 (0.383)	-0.811 (0.385)
σ	0.236 (0.110)	0.234 (0.108)	0.235 (0.108)	0.234 (0.109)	0.238 (0.106)

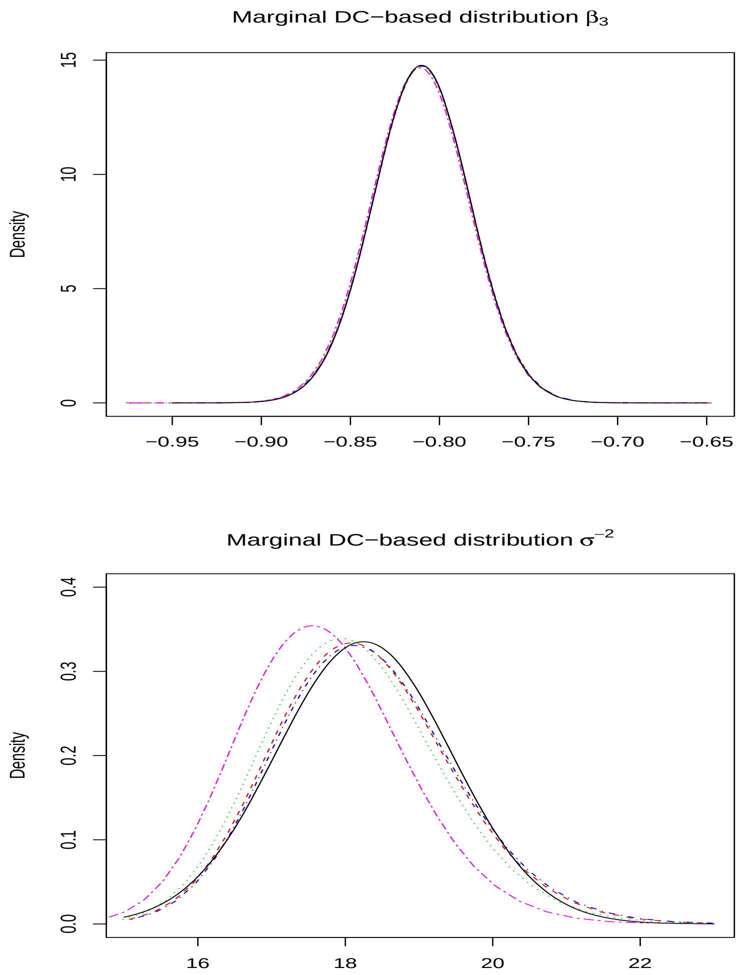

Figure 1: Marginal DC-based densities of interaction effect (top panel) and precision parameter of random effect (bottom panel) in the interaction model with $k = 200$; The graphs showing the densities for first prior set (dashes), second prior set (dots), third prior set (dot-small dash), fourth prior set (dot-dash), and approximate normal density (solid).

the trial included the number of epileptic seizures recorded in the preceding 8-weeks period and age in years. The logarithm of $\frac{1}{4}$ the number of baseline seizures (Base) and the logarithm of age (Age) were treated as covariables in the analysis. A multivariate response variable consisted of the counts of epileptic seizures, y_{ij}, for patient i during the 2-weeks before each of four clinic visits j (Visit, coded $-3, -1, +1, +3$), with $Y_{ij}|\boldsymbol{\beta}, u_i \stackrel{iid}{\sim} Po(\mu_{ij}), i = 1, \ldots, 59; j = 1, \ldots, 4$. An indicator of the fourth visit was also constructed to model to account its effect. Following Fong *et al.* (2009), we concentrate on the three random effects models fitted by Breslow and Clayton (1993):

$$\log(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + u_{1i}, \tag{5}$$

$$\log(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + u_{1i} + u_{0ij}, \tag{6}$$

$$\log(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + u_{1i} + u_{2i}V_j/10, \tag{7}$$

where $\boldsymbol{x}_{ij}$ is a 6×1 vector containing a 1 for intercept, the baseline by treatment interaction and above mentioned covariates and $\boldsymbol{\beta}$ is the associated fixed effect. All three models include patient-specific random effects $u_{1i} \sim N(0, \sigma_1^2)$, while in model (6) we introduce independent measurement errors $b_{0ij} \sim N(0, \sigma_0^2)$ and model (7) includes random effects on the slope associated with visit, u_{2i}, with

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim N(\mathbf{0}, Q^{-1}).$$

According to Fong *et al.* (2009) we assume $Q \sim Wishart(r, T)$ with

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}.$$

Here, similar to the previous subsection, we used three different sets of prior distributions. The first set for three models are priors considered by Fong *et al.* (2009). The second and third sets for the first model include $N(1, 100)$ for $\boldsymbol{\beta}$ and $Ga(1, 2.5)$ for σ_1^{-2} and $N(0, 10)$, $N(-1, 10)$, $N(-2, 100)$, $N(2, 100)$, $N(0, 10)$, $N(1, 100)$ for $\boldsymbol{\beta}$ and $Ga(1.5, 2)$ for σ_1^{-2} respectively. For the second model the second prior set include $N(1, 100)$ for $\boldsymbol{\beta}$ and $Ga(1, 2.5)$ and $Ga(2, 1.140)$ for σ_1^{-2} and σ_0^{-2} respectively. And the third set include $N(0, 10)$, $N(-1, 10)$, $N(-2, 100)$, $N(2, 100)$, $N(0, 10)$, $N(1, 100)$ for $\boldsymbol{\beta}$ and $Ga(1.5, 2)$ and $Ga(2, 1.140)$ for σ_1^{-2} and σ_0^{-2} respectively. Finally, for the third model the second and third prior sets include the same priors for fixed effects as second model but they include $r = 4$ and $T = diag(3, 4)$ and $r = 6$ and $T = diag(0.5, 0.5)$ respectively.

Table 3: DC estimates obtained by hybrid method with $k = 100$ for Epilepsy data in the model (5) compared with PQL and INLA estimates. [a] From Fong *et al.* (2009). Standard errors are in brackets.

Par.	PQL[a]	INLA[a]	DC_1	DC_2	DC_3
Base	0.87 (0.14)	0.88 (0.15)	0.88 (0.13)	0.88 (0.13)	0.88 (0.13)
Trt	-0.91 (0.41)	-0.94 (0.44)	-0.93 (0.40)	-0.93 (0.40)	-0.93 (0.40)
Base×Trt	0.33 (0.21)	0.34 (0.22)	0.34 (0.20)	0.34 (0.20)	0.34 (0.20)
Age	0.47 (0.36)	0.47 (0.38)	0.48 (0.35)	0.48 (0.35)	0.48 (0.35)
V_4 or $V/10$	-0.16 (0.05)	-0.16 (0.05)	-0.16 (0.05)	-0.16 (0.05)	-0.16 (0.05)
σ_1	0.53 (0.06)	0.56 (0.08)	0.50 (0.06)	0.50 (0.06)	0.50 (0.06)

Table 4: DC estimates obtained by hybrid method with $k = 100$ for Epilepsy data in the model (6) compared with PQL and INLA estimates. [a] From Fong *et al.* (2009). Standard errors are in brackets.

Par.	PQL[a]	INLA[a]	DC_1	DC_2	DC_3
Base	0.86 (0.13)	0.88 (0.15)	0.88 (0.13)	0.88 (0.13)	0.88 (0.13)
Trt	-0.93 (0.40)	-0.96 (0.44)	-0.95 (0.40)	-0.95 (0.40)	-0.95 (0.40)
Base×Trt	0.34 (0.21)	0.35 (0.23)	0.35 (0.20)	0.35 (0.20)	0.35 (0.20)
Age	0.47 (0.35)	0.48 (0.39)	0.49 (0.34)	0.49 (0.34)	0.49 (0.34)
V_4 or $V/10$	-0.10 (0.09)	-0.10 (0.09)	-0.10 (0.09)	-0.10 (0.09)	-0.10 (0.09)
σ_0	0.36 (0.04)	0.41 (0.04)	0.36 (0.04)	0.36 (0.04)	0.36 (0.04)
σ_1	0.48 (0.06)	0.53 (0.07)	0.46 (0.06)	0.46 (0.06)	0.46 (0.06)

Tables 3–5 present the results obtained by hybrid DC method with $k = 100$. The results are compared with PQL and INLA results of Fong *et al.* (2009). It is clear that the results are indistinguishable for different priors and are very close to results of Fong *et al.* (2009). Note that $\sigma_2^2 = T_{22}^{-1}$.

Figures 2 and 3 show the marginal DC-based densities of fixed effects and precision parameters of random effects for third model (7) respectively. According to the figures, detection of any difference between curves is not possible. Approximate normal distributions are also illustrated in these figures. It is clear that the normal distributions give very good approximations to the hybrid DC-based distributions. The results remain the same for two other models as well.

4.3. Crossed Random Effects: The Salamander Data

McCullagh and Nelder (1989) described an interesting dataset on the success of matings between male and female salamander of two population types, roughbutts (RB) and whitesides (WS).

Table 5: DC estimates obtained by hybrid method with $k = 100$ for Epilepsy data in the model (7) compared with PQL and INLA estimates. [a] From Fong *et al.* (2009). Standard errors are in brackets.

Par.	PQL[a]	INLA[a]	DC_1	DC_2	DC_3
Base	0.87 (0.14)	0.88 (0.14)	0.89 (0.13)	0.88 (0.13)	0.89 (0.13)
Trt	-0.91 (0.41)	-0.94 (0.44)	-0.93 (0.40)	-0.93 (0.40)	-0.93 (0.40)
Base×Trt	0.33 (0.21)	0.34 (0.22)	0.34 (0.20)	0.34 (0.20)	0.34 (0.20)
Age	0.46 (0.36)	0.47 (0.38)	0.47 (0.35)	0.48 (0.35)	0.48 (0.35)
V_4 or $V/10$	-0.26 (0.16)	-0.27 (0.16)	-0.26 (0.17)	-0.27 (0.17)	-0.27 (0.16)
σ_1	0.52 (0.06)	0.56 (0.08)	0.50 (0.06)	0.50 (0.06)	0.50 (0.06)
σ_2	0.74 (0.16)	0.70 (0.14)	0.73 (0.15)	0.73 (0.15)	0.72 (0.15)

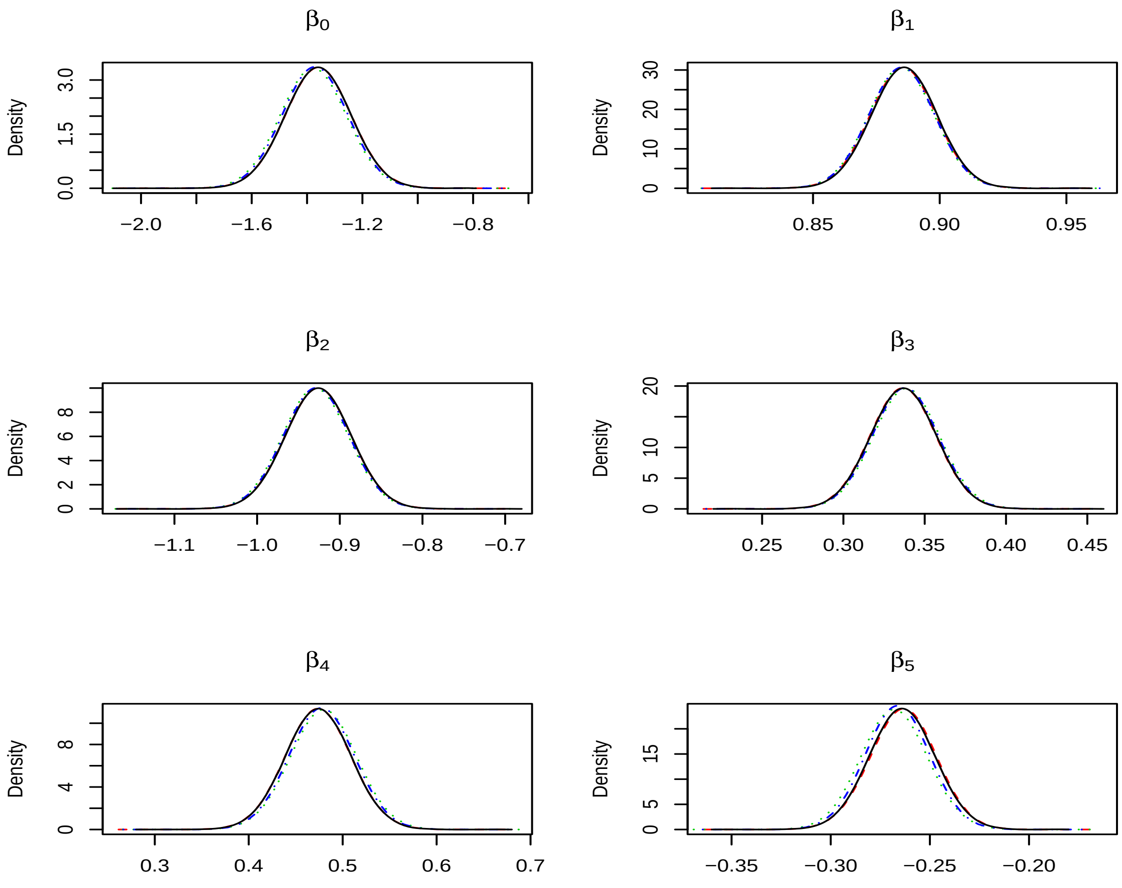

Figure 2: Marginal DC-based densities of fixed effects in the model (7) with $k = 100$; The graphs showing the densities for first prior set (dashes), second prior set (dots), third prior set (dot-dash), and approximate normal density (solid).

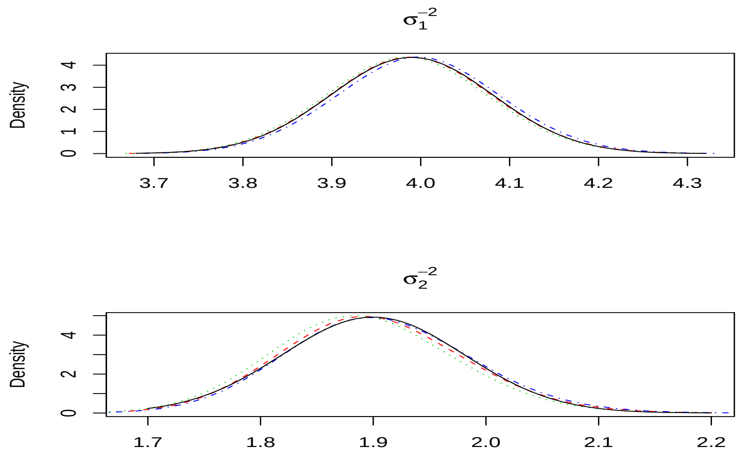

Figure 3: Marginal DC-based densities of precision parameters of random effects in the model (7) with $k = 100$; The graphs showing the densities for first prior set (dashes), second prior set (dots), third prior set (dot-dash), and approximate normal density (solid).

The experimental design involves three experiments having multiple pairings, with each salamander being involved in multiple matings, so that crossed random effects are required. The first experiment conducted during the summer of 1986 and the second and third conducted in the fall. Each experiment involved 30 matings of each of the four gender-population combinations. There are 360 binary responses in total. This complex data is reanalyzed by several authors such as karim and Zeger (1992), Breslow and Clayton (1993), Bellio and Varin (2005) and Fong *et al.* (2009).

Suppose y_{ijk} be the binary response for female i and male j in experiment k. Here, we focus on model that was considered by Fong *et al.* (2009):

$$logitPr(Y_{ijk} = 1|\boldsymbol{\beta}, u^f_{ik}, u^m_{jk}) = \boldsymbol{x}^T_{ijk}\boldsymbol{\beta}_k + u^f_{ik} + u^m_{jk},$$

where $\boldsymbol{x}_{ijk}$ is a 4×1 vector representing the intercept, an indicator WS_f of whiteside females, an indicator WS_m of whiteside males and their interaction and $\boldsymbol{\beta}_k$ is the corresponding fixed effect. As Fong *et al.* (2009) have mentioned this model allows the fixed effects to vary by experiment and the model contains six random effects

$$u^f_{ik} \overset{iid}{\sim} N(o, \sigma^2_{fk}), u^m_{ik} \overset{iid}{\sim} N(o, \sigma^2_{mk}), \quad k = 1, 2, 3$$

one for each of males and females, and in each experiment.

Table 6: DC estimates obtained by hybrid method with $k = 100$ for Salamander data for summer experiment compared with REML and INLA estimates. [a] From Fong *et al.* (2009). Standard errors are in brackets and $*$ denotes standard errors that were unavailable.

Par.	REML[a]	INLA[a]	DC_1	DC_2	DC_3
Intercept	1.34 (0.62)	1.48 (0.72)	1.33 (0.63)	1.34 (0.63)	1.34 (0.63)
WS_f	-2.94 (0.88)	-3.26 (1.01)	-2.93 (0.93)	-2.95 (0.93)	-2.94 (0.93)
WS_m	-0.42 (0.63)	-0.50 (0.73)	-0.42 (0.64)	-0.43 (0.64)	-0.43 (0.64)
$WS_f \times WS_m$	3.18 (0.94)	3.52 (1.03)	3.17 (0.99)	3.19 (0.99)	3.18 (0.99)
σ_{f1}	1.25 ($*$)	1.29 (0.46)	1.24 (0.40)	1.24 (0.41)	1.24 (0.40)
σ_{m1}	0.27 ($*$)	0.78 (0.29)	0.33 (0.41)	0.37 (0.40)	0.36 (0.39)

Table 7: DC estimates obtained by hybrid method with $k = 100$ for Salamander data for first fall experiment compared with REML and INLA estimates. [a] From Fong *et al.* (2009). Standard errors are in brackets and $*$ denotes standard errors that were unavailable.

Par.	REML[a]	INLA[a]	DC_1	DC_2	DC_3
Intercept	0.57 (0.67)	0.56 (0.71)	0.55 (0.65)	0.55 (0.65)	0.55 (0.65)
WS_f	-2.46 (0.93)	-2.51 (1.02)	-2.39 (0.95)	-2.39 (0.96)	-2.39 (0.95)
WS_m	-0.77 (0.72)	-0.75 (0.75)	-0.73 (0.70)	-0.73 (0.70)	-0.73 (0.70)
$WS_f \times WS_m$	3.71 (0.96)	3.74 (1.03)	3.58 (1.03)	3.59 (1.03)	3.58 (1.03)
σ_{f2}	1.35 ($*$)	1.38 (0.50)	1.30 (0.47)	1.31 (0.47)	1.30 (0.47)
σ_{m2}	0.96 ($*$)	1.00 (0.36)	0.91 (0.42)	0.92 (0.42)	0.91 (0.42)

Again similar to the previous subsections, we used three different sets of prior distributions and the first set presents priors considered by Fong *et al.* (2009). The second set include $N(0, 10)$ for $\boldsymbol{\beta}$ and $Ga(0.25, 1)$ for both σ_{fk}^{-2} and σ_{mk}^{-2}. The third set also include $N(0, 10)$ for $\boldsymbol{\beta}$ and $Ga(1, 1)$ for both σ_{fk}^{-2} and σ_{mk}^{-2}.

Tables 6–8 show the results obtained by hybrid DC method with $k = 100$. The results are compared with REML and INLA results of Fong *et al.* (2009). It is simple to see that the results are indistinguishable for different priors. The results are also very close to REML but there are some differences between them and INLA estimates, usually with slightly larger standard deviations under the latter.

Figures 4–6 show the marginal DC-based densities of precision parameters of random effects obtained for three experiments. According to the figures, there are a good matching between curves. Approximate normal distributions are also illustrated in these figures. It is clear that the

Table 8: DC estimates obtained by hybrid method with $k = 100$ for Salamander data for second fall experiment compared with REML and INLA estimates. [a] From Fong *et al.* (2009). Standard errors are in brackets and $*$ denotes standard errors that were unavailable.

Par.	REML[a]	INLA[a]	DC_1	DC_2	DC_3
Intercept	1.02 (0.65)	1.07 (0.73)	1.00 (0.64)	1.00 (0.64)	1.00 (0.64)
WS_f	-3.23 (0.83)	-3.39 (0.92)	-3.17 (0.85)	-3.18 (0.86)	-3.17 (0.86)
WS_m	-0.82 (0.86)	-0.85 (0.94)	-0.79 (0.85)	-0.79 (0.85)	-0.79 (0.85)
$WS_f \times WS_m$	3.82 (0.99)	4.03 (1.05)	3.74 (1.03)	3.75 (1.04)	3.75 (1.03)
σ_{f3}	0.59 ($*$)	0.80 (0.28)	0.54 (0.44)	0.57 (0.42)	0.55 (0.43)
σ_{m3}	1.36 ($*$)	1.46 (0.48)	1.33 (0.43)	1.34 (0.43)	1.34 (0.43)

normal distributions give very good approximations to the hybrid DC-based distributions and the approximation bias is negligible. Detection of any difference between hybrid DC-based densities of fixed effects and between their approximate normal densities is not possible.

5. Technical Details

In this section we give details for establishing the asymptotic normality of the hybrid DC-based distribution. For this purpose, we first establish the asymptotic normality of the approximate posterior distributions as well as the DC-based distribution. Then the asymptotic normality of the hybrid DC-based distribution follows by combining them together.

Some notations and calculations are needed in the sequel. We show the approximations of $\ell_n(\boldsymbol{\psi})$ and $\ell_n(\boldsymbol{\theta})$ by $\tilde{\ell}_n(\boldsymbol{\psi})$ and $\tilde{\ell}_n(\boldsymbol{\theta})$ which obtained by INLA. Therefore (1) is approximated by

$$\tilde{\pi}(\boldsymbol{\psi}, \boldsymbol{\theta}|\boldsymbol{y}) = \exp\{\tilde{\ell}_n(\boldsymbol{\psi}) + \tilde{\ell}_n(\boldsymbol{\theta})\}.$$

Assume that the functions $\ell_n(\boldsymbol{\theta})$ and $\tilde{\ell}_n(\boldsymbol{\theta})$ as well as $\ell_n(\boldsymbol{\psi})$ and $\tilde{\ell}_n(\boldsymbol{\psi})$ are twice continuously differentiable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ respectively. Let $\nabla\ell_n(\boldsymbol{\theta})$, $\nabla\tilde{\ell}_n(\boldsymbol{\theta})$, $\nabla\ell_n(\boldsymbol{\psi})$ and $\nabla\tilde{\ell}_n(\boldsymbol{\psi})$ be the vectors of first-order partial derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ respectively. Furthermore, let $\nabla^2\ell_n(\boldsymbol{\theta})$, $\nabla^2\tilde{\ell}_n(\boldsymbol{\theta})$, $\nabla^2\ell_n(\boldsymbol{\psi})$ and $\nabla^2\tilde{\ell}_n(\boldsymbol{\psi})$ be the matrices of second-order partial derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ respectively. Here and subsequently, let $\hat{\boldsymbol{\psi}}_n$ be the mode of $\ell_n(\boldsymbol{\psi})$, satisfying $\nabla\ell_n(\boldsymbol{\psi}) = 0$ and $\hat{\boldsymbol{\theta}}_n$ be the mode of $\ell_n(\boldsymbol{\theta})$, satisfying $\nabla\ell_n(\boldsymbol{\theta}) = 0$.

To facilitate asymptotic theory arguments, whenever the $\hat{\boldsymbol{\psi}}_n$ and $\hat{\boldsymbol{\theta}}_n$ exist and $-\nabla^2\ell_n(\boldsymbol{\psi})$, $-\nabla^2\tilde{\ell}_n(\boldsymbol{\psi})$, $-\nabla^2\ell_n(\boldsymbol{\theta})$ and $-\nabla^2\tilde{\ell}_n(\boldsymbol{\theta})$ are positive definite, we define F_n, G_n, Q_n, V_n, $\boldsymbol{z}_n$, $\boldsymbol{w}_n$, $\boldsymbol{x}_n$

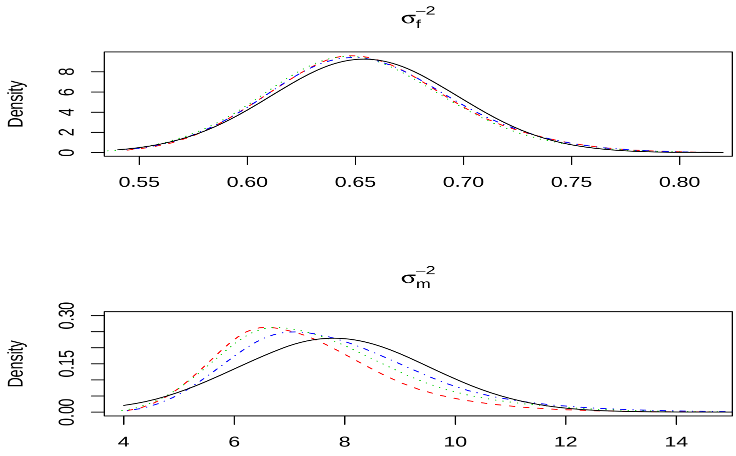

Figure 4: Marginal DC-based densities of precision parameters of random effects obtained for summer experiment with $k = 100$; The graphs showing the densities for first prior set (dashes), second prior set (dots), third prior set (dot-dash), and approximate normal density (solid).

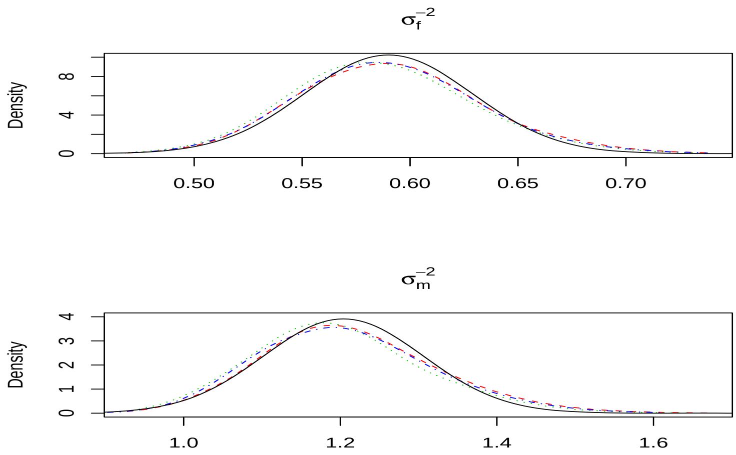

Figure 5: Marginal DC-based densities of precision parameters of random effects obtained for first fall experiment with $k = 100$; The graphs showing the densities for first prior set (dashes), second prior set (dots), third prior set (dot-dash), and approximate normal density (solid).

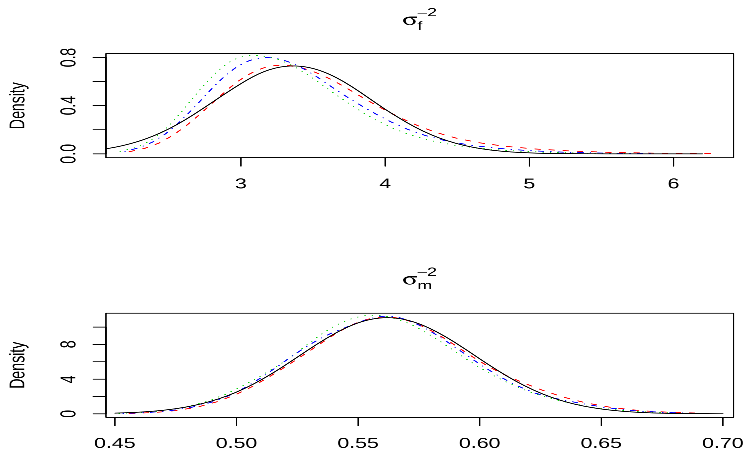

Figure 6: Marginal DC-based densities of precision parameters of random effects obtained for second fall experiment with $k = 100$; The graphs showing the densities for first prior set (dashes), second prior set (dots), third prior set (dot-dash), and approximate normal density (solid).

and $\boldsymbol{s}_n$ as follow:

$$\begin{aligned} F_n^T F_n &= -\nabla^2 \ell_n(\hat{\boldsymbol{\psi}}_n), \;\; \boldsymbol{z}_n = F_n(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \\ Q_n^T Q_n &= -\nabla^2 \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n), \;\; \boldsymbol{x}_n = Q_n(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \\ G_n^T G_n &= -\nabla^2 \ell_n(\hat{\boldsymbol{\theta}}_n), \;\; \boldsymbol{w}_n = G_n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \\ V_n^T V_n &= -\nabla^2 \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n), \;\; \boldsymbol{s}_n = V_n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \end{aligned}$$

otherwise define them arbitrarily, in a measurable way. Then the joint approximate posterior density of $(\boldsymbol{x}_n, \boldsymbol{s}_n)$ is

$$\tilde{\pi}_n(\boldsymbol{x}_n, \boldsymbol{s}_n|\boldsymbol{y}) \propto \tilde{\pi}_n(\boldsymbol{\psi}(\boldsymbol{x}_n), \boldsymbol{\theta}(\boldsymbol{s}_n)|\boldsymbol{y}) \propto e^{\tilde{\ell}_n(\boldsymbol{\psi}) - \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n)} e^{\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n)}. \tag{8}$$

Let $\boldsymbol{\theta}_0$ and $\boldsymbol{\psi}_0$ denote the true underlying parameters and the true realization of random effects respectively. Let also $\tilde{P}_n^c$ and $\tilde{E}_n^c$ denote the approximate conditional probability and expectation given data $\boldsymbol{y}$. In what follows, all probability statements are with respect to the true underlying probability distribution. Then we must show

$$\tilde{P}_n^c((\boldsymbol{x}_n^T, \boldsymbol{s}_n^T)^T \in B) \longrightarrow \Phi_{q+d}(B),$$

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

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

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

where $\boldsymbol{\psi}^*$ lies between $\boldsymbol{\psi}$ and $\hat{\boldsymbol{\psi}}_n$. According to Tierney and Kadane (1986), say for $\boldsymbol{\theta}$, $\ell_n(\boldsymbol{\theta}_1) = \tilde{\ell}_n(\boldsymbol{\theta}_1)(1+O_{\boldsymbol{\theta}_1}(n^{-1}))$, where $O_{\boldsymbol{\theta}_1}(n^{-1})$ is of order $O(n^{-1})$ but depends on $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$. Tierney and Kadane (1986) showed that this error term will be uniformly of order $O(n^{-1})$ for $\boldsymbol{\theta}_1$ in some fixed neighborhood of $\boldsymbol{\theta}_0$. They also showed the error in the approximation $\tilde{\ell}_n(\boldsymbol{\theta}_1)$ is of order $O(n^{-3/2})$ in $n^{-1/2}$ neighborhood of $\hat{\boldsymbol{\theta}}_{1n}$. Lemma 2 reveals their result.

Lemma 2. *Let $N_1 = N(\hat{\boldsymbol{\psi}}_n, n^{-1/2})$ and $N_2 = N(\hat{\boldsymbol{\theta}}_n, n^{-1/2})$ be $n^{-1/2}$ neighborhoods of $\hat{\boldsymbol{\psi}}_n$ and $\hat{\boldsymbol{\theta}}_n$ respectively. Then, for $\boldsymbol{\psi} \in N_1$ and $\boldsymbol{\theta} \in N_2$ we have,*

$$\begin{aligned} \ell_n(\boldsymbol{\psi}) &= \tilde{\ell}_n(\boldsymbol{\psi})(1 + O(n^{-3/2})), \\ \ell_n(\boldsymbol{\theta}) &= \tilde{\ell}_n(\boldsymbol{\theta})(1 + O(n^{-3/2})). \end{aligned}$$

Remark 1. *By Lemma 2, it is easy to see that*

$$\begin{aligned} \nabla\ell_n(\boldsymbol{\psi}) &= \nabla\tilde{\ell}_n(\boldsymbol{\psi})(1 + O(n^{-3/2})), \\ \nabla\ell_n(\boldsymbol{\theta}) &= \nabla\tilde{\ell}_n(\boldsymbol{\theta})(1 + O(n^{-3/2})). \end{aligned}$$

The same relations hold for $\nabla^2\tilde{\ell}_n(\boldsymbol{\psi})$ and $\nabla^2\tilde{\ell}_n(\boldsymbol{\theta})$.

Hereafter, whenever be required, we consider $\boldsymbol{\psi} \in N_1$ and $\boldsymbol{\theta} \in N_2$. Now by Remark 1 we have,

$$\tilde{\ell}_n(\boldsymbol{\psi}) = \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n) + \frac{1}{2}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n)^T \nabla^2 \tilde{\ell}_n(\boldsymbol{\psi}^*)(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n) + R_n',$$

in which

$$\begin{aligned} R_n' = R_n \quad &+ \quad [\ell_n(\hat{\boldsymbol{\psi}}_n) - \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n)] + [\tilde{\ell}_n(\boldsymbol{\psi}) - \ell_n(\boldsymbol{\psi})] \\ &+ \quad \frac{1}{2}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n)^T [\nabla^2 \ell_n(\boldsymbol{\psi}^*) - \nabla^2 \tilde{\ell}_n(\boldsymbol{\psi}^*)](\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n). \end{aligned}$$

Let

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

Thus,

$$\tilde{\ell}_n(\boldsymbol{\psi}) \approx \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n) - \frac{1}{2}\|\boldsymbol{x}_n\|^2 + k_n(\boldsymbol{\psi}).$$

With parallel arguments, we have

$$\begin{aligned}
\tilde{\ell}_n(\boldsymbol{\theta}) &\approx \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)^T \nabla^2 \tilde{\ell}_n(\boldsymbol{\theta}^*)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \\
l_n(\boldsymbol{\theta}) &= -\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)^T[\nabla^2 \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) - \nabla^2 \tilde{\ell}_n(\boldsymbol{\theta}^*)](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \\
\tilde{\ell}_n(\boldsymbol{\theta}) &\approx \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) - \frac{1}{2}\|\boldsymbol{s}_n\|^2 + l_n(\boldsymbol{\theta}),
\end{aligned}$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_n$. Therefor we can rewrite (8) as

$$\tilde{\pi}_n(\boldsymbol{x}_n, \boldsymbol{s}_n|\boldsymbol{y}) \propto \phi_q(\boldsymbol{x}_n)\phi_d(\boldsymbol{s}_n)f_n(\boldsymbol{x}_n, \boldsymbol{s}_n), \tag{9}$$

where $f_n(\boldsymbol{x}_n, \boldsymbol{s}_n) = \exp\{k_n(\boldsymbol{\psi}) + l_n(\boldsymbol{\theta})\}$ and $\phi_t(\cdot)$ display the standard t-variate Gaussian density.

Suppose $\nabla_{\boldsymbol{x}_n} f(\boldsymbol{x}_n, \boldsymbol{s}_n)$ and $\nabla_{\boldsymbol{s}_n} f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)$ denote the partial derivatives of $f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)$ with respect to $\boldsymbol{x}_n$ and $\boldsymbol{s}_n$ respectively. Hence,

$$\frac{\nabla_{\boldsymbol{x}_n} f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)}{f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)} = (Q_n^T)^{-1} \nabla k_n(\boldsymbol{\psi}), \tag{10}$$

$$\frac{\nabla_{\boldsymbol{s}_n} f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)}{f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)} = (V_n^T)^{-1} \nabla l_n(\boldsymbol{\theta}). \tag{11}$$

Let also $D = \{D_1 \cup D_2\}$, in which $D_1 = \{\nabla \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n) = 0, -\nabla^2 \tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n) > 0\}$ and $D_2 = \{\nabla \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) = 0, -\nabla^2 \tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) > 0\}$. Here $A > 0$ means that the matrix A is positive definite.

5.1. Asymptotic Normality of Approximate Posterior Distribution

We consider the regularity conditions of Weng and Tsai (2008) modifying for $\tilde{\ell}_n(\boldsymbol{\theta})$, (B1)-(B4), and $\tilde{\ell}_n(\boldsymbol{\psi})$, (C1)-(C4), instead of $\ell_n(\boldsymbol{\theta})$ and $\ell_n(\boldsymbol{\psi})$. We also suppose $\|J\|^2 = \lambda_{\max}(J^T J)$ be the spectral norm of J. The following conditions are required for $\tilde{\ell}_n(\boldsymbol{\theta})$:

(B1) $P(D_2^c) \longrightarrow 0$, $\|V_n^{-1}\| \xrightarrow{p} 0$, and $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ as $n \longrightarrow \infty$.

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

$$\sup_{\eta^{ij} \in \{\boldsymbol{\theta}: \|\boldsymbol{s}_n\| \leq b_{1n}\}} \|I_d + (V_n^T)^{-1}(\partial^2 \tilde{\ell}_n / \partial\theta_i \partial\theta_j(\eta^{ij}))V_n^{-1}\| \xrightarrow{p} 0.$$

(B3) Let b_{1n} be as in (B2). There exist constants $r_1 \geq 1$ and $c_1 \geq 0$ such that for all $\boldsymbol{\theta} \in \{\|\boldsymbol{s}_n\| > b_{1n}\} \cap \boldsymbol{\Theta}$, $\|(V_n^T)^{-1} \nabla l_n(\boldsymbol{\theta})\| \leq c_1 \|\boldsymbol{s}_n\|^{r_1}$.

(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 $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, $[\tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) - \tilde{\ell}_n(\boldsymbol{\theta})] \geq v_1(t, \boldsymbol{\theta})$, $e_{1n}(\boldsymbol{\theta}) = (detV_n)\|\boldsymbol{s}_n\|^{r_1} e^{-v_1(t,\boldsymbol{\theta})}$ are uniformly integrable in t and $\int_{\boldsymbol{\Theta}} e_{1n}(\boldsymbol{\theta}) d\boldsymbol{\theta}$ are uniformly bounded in t.

The following conditions are also required for $\tilde{\ell}_n(\boldsymbol{\psi})$:

(C1) $P(D_1^c) \longrightarrow 0$, $\|Q_n^{-1}\| \xrightarrow{p} 0$, and $\hat{\boldsymbol{\psi}}_n \xrightarrow{p} \boldsymbol{\psi}_0$ as $n \longrightarrow \infty$.

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

$$\sup_{\zeta^{ij} \in \{\boldsymbol{\psi} : \|\boldsymbol{x}_n\| \leq b_{2n}\}} \|I_q + (Q_n^T)^{-1} (\partial^2 \tilde{\ell}_n / \partial \psi_i \partial \psi_j (\zeta^{ij})) Q_n^{-1}\| \xrightarrow{p} 0.$$

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

(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 $\forall \boldsymbol{\psi} \in \boldsymbol{\Psi}$, $[\tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n) - \tilde{\ell}_n(\boldsymbol{\psi})] \geq v_2(t, \boldsymbol{\psi})$, $e_{2n}(\boldsymbol{\psi}) = (detQ_n)\|\boldsymbol{x}_n\|^{r_2} e^{-v_2(t,\boldsymbol{\psi})}$ are uniformly integrable in t and $\int_{\boldsymbol{\Psi}} e_{2n}(\boldsymbol{\psi}) d\boldsymbol{\psi}$ are uniformly bounded in t.

Furthermore, we consider the conditions (F1)-(F3) for $\pi(\boldsymbol{\theta})$:

(F1) $\pi(\boldsymbol{\theta})$ is continuously differentiable on $\Re^d$.

(F2) $\pi(\boldsymbol{\theta})$ has a compact support $\boldsymbol{\Theta} \subset \Re^d$

(F3) There exist ϵ_0 and δ_0 such that $\pi(\boldsymbol{\theta}) > \epsilon_0$ over $N(\boldsymbol{\theta}_0; \delta_0)$.

Let also

$$S = \left\{ (\boldsymbol{x}_n, \boldsymbol{s}_n) : \boldsymbol{x}_n = Q_n(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n), \boldsymbol{s}_n = V_n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n); \boldsymbol{\psi} \in N_1, \boldsymbol{\theta} \in N_2 \right\}. \tag{12}$$

The following theorem establishes the asymptotic normality of approximate posterior distribution.

Theorem 3. *Suppose that h be any bounded measurable function as in Lemma 1. Moreover, suppose that the prior $\pi(\boldsymbol{\theta})$ satisfies (F1)-(F3), $\tilde{\ell}_n(\boldsymbol{\theta})$ satisfies (B1)-(B4) and $\tilde{\ell}_n(\boldsymbol{\psi})$ satisfies (C1)-(C4). Then, $\tilde{E}_n^c[h(\boldsymbol{x}_n, \boldsymbol{s}_n)] \xrightarrow{p} \Phi h$.*

Before proving the theorem, we need the following lemmas. Their proofs are extended forms of similar lemmas of Baghishani and Mohammadzadeh (2010) such that replacing $\ell_n(\boldsymbol{\psi})$ and $\ell_n(\boldsymbol{\theta})$ with $\tilde{\ell}_n(\boldsymbol{\psi})$ and $\tilde{\ell}_n(\boldsymbol{\theta})$, they hold.

Lemma 3. *Let s be a nonnegative integer. Suppose that $\pi(\boldsymbol{\theta})$ satisfies $(F1)$ and $(F2)$. Then for all $h : \Re^{q+d} \longrightarrow \Re$ where $h \in H_s$,*

$$\tilde{E}_n^c[h(\boldsymbol{x}_n, \boldsymbol{s}_n)] - \Phi h = \tilde{E}_n^c \left\{ (Uh(\boldsymbol{x}_n, \boldsymbol{s}_n))^T [\frac{\nabla_{\boldsymbol{x}_n} f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)}{f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)}, \frac{\nabla_{\boldsymbol{s}_n} f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)}{f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)}] \right\},$$

a.e. on D.

Lemma 4. 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,*

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{s}_n\|\leq p_1} \{\tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) - \tilde{\ell}_n(\boldsymbol{\theta})\} \leq q_1,$$
$$\sup_{\boldsymbol{\psi}:\|\boldsymbol{x}_n\|\leq p_2} \{\tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n) - \tilde{\ell}_n(\boldsymbol{\psi})\} \leq q_2.$$

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

$$\int_S e^{\{\tilde{\ell}_n(\boldsymbol{\psi})-\tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n)\}} e^{\{\tilde{\ell}_n(\boldsymbol{\theta})-\tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n)\}} d\boldsymbol{x}_n d\boldsymbol{s}_n < M$$
$$\int_S \|\boldsymbol{x}_n\|\|\boldsymbol{s}_n\| e^{\{\tilde{\ell}_n(\boldsymbol{\psi})-\tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n)\}} e^{\{\tilde{\ell}_n(\boldsymbol{\theta})-\tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n)\}} d\boldsymbol{x}_n d\boldsymbol{s}_n < M$$
$$\int_{S'} \|\boldsymbol{s}_n\|^{r_1} \|\boldsymbol{x}_n\|^{r_2} e^{\{\tilde{\ell}_n(\boldsymbol{\psi})-\tilde{\ell}_n(\hat{\boldsymbol{\psi}}_n)\}} e^{\{\tilde{\ell}_n(\boldsymbol{\theta})-\tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n)\}} d\boldsymbol{x}_n d\boldsymbol{s}_n \xrightarrow{p} 0$$

where $S' = S \cap \{\|\boldsymbol{s}_n\| > b_{1n}\} \cap \{\|\boldsymbol{x}_n\| > b_{2n}\}$

Lemma 5. *Let $f_n(\boldsymbol{x}_n, \boldsymbol{s}_n)$ and S be as in (9) and (12) respectively. Suppose that $\pi(\boldsymbol{\theta})$ satisfies $(F1)$-$(F3)$. Then,*

D1 *If conditions $(B1)$, $(B2)$, $(C1)$ and $(C2)$ hold, then there exists $K_1 > 0$ such that, with probability tending to 1, $\int_S \phi_q(\boldsymbol{x}_n)\phi_d(\boldsymbol{s}_n) f_n(\boldsymbol{x}_n, \boldsymbol{s}_n) d\boldsymbol{x}_n d\boldsymbol{s}_n > K_1$.*

D2 *If $(B4)$ and $(C4)$ hold, then there exists $K_2 > 0$ such that, with probability tending to 1, $\int_S \phi_q(\boldsymbol{x}_n)\phi_d(\boldsymbol{s}_n) f_n(\boldsymbol{x}_n, \boldsymbol{s}_n) d\boldsymbol{x}_n d\boldsymbol{s}_n < K_2$.*

Proof. Note that Uh and $\pi(\boldsymbol{\theta})$ are bounded by Lemma 3.1 of Weng and Tsai (2008) and $(F1)$-$(F2)$. From (10), (11) and Lemma 3, for a.e. on D, we have

$$\tilde{E}_n^c[h(\boldsymbol{x}_n, \boldsymbol{s}_n)] - \Phi h = \tilde{E}_{n,\boldsymbol{x}_n}^c + \tilde{E}_{n,\boldsymbol{s}_n}^c,$$

where

$$\tilde{E}^c_{n,\boldsymbol{x}_n} = \tilde{E}^c_n \left\{ (Uh_1(\boldsymbol{x}_n))^T (Q_n^T)^{-1} \nabla k_n(\boldsymbol{\psi}) \right\} = I_{\boldsymbol{x}_n}, \tag{13}$$

$$\tilde{E}^c_{n,\boldsymbol{s}_n} = \tilde{E}^c_n \left\{ (Uh_2(\boldsymbol{s}_n))^T (V_n^T)^{-1} \nabla l_n(\boldsymbol{\theta}) \right\} = I_{\boldsymbol{s}_n}. \tag{14}$$

Since $P(D_2^c) \longrightarrow 0$ by $(B1)$ and $P(D_1^c) \longrightarrow 0$ by $(C1)$, it suffices to show $I_{\boldsymbol{s}_n} \xrightarrow{p} 0$ and $I_{\boldsymbol{x}_n} \xrightarrow{p} 0$.

From (14) we have,

$$I_{\boldsymbol{s}_n} = \frac{\int_S (Uh_2(\boldsymbol{s}_n))^T (V_n^T)^{-1} \nabla l_n(\boldsymbol{\theta}) \phi_d(\boldsymbol{s}_n)\phi_q(\boldsymbol{x}_n) f_n(\boldsymbol{s}_n, \boldsymbol{x}_n) d\boldsymbol{s}_n d\boldsymbol{x}_n}{\int_S \phi_d(\boldsymbol{s}_n)\phi_q(\boldsymbol{x}_n) f_n(\boldsymbol{s}_n, \boldsymbol{x}_n) d\boldsymbol{s}_n d\boldsymbol{x}_n}.$$

The denominator is bounded below by some $K_1 > 0$ by Lemma 5(D1). Then we just need to show that the numerator converges to 0 in probability. First we decompose the numerator into two integrals over $\|\boldsymbol{s}_n\| \le b_{1n}$ and $\|\boldsymbol{s}_n\| > b_{1n}$ and call the corresponding integrals as $I_{\boldsymbol{s}_n,1}$ and $I_{\boldsymbol{s}_n,2}$ respectively. With respect to $(F1)$-$(F2)$, Lemma 3.1 of Weng and Tsai (2008), and

$$(V_n^T)^{-1} \nabla l_n(\boldsymbol{\theta}) = \left\{ I_d - (V_n^T)^{-1} [-(\frac{\partial^2 \tilde{\ell}_n}{\partial\theta_i \partial\theta_j}(\theta^{*ij}))] V_n^{-1} \right\} \boldsymbol{s}_n,$$

there exists a constant $C_1 > 0$ such that

$$\begin{aligned} |I_{\boldsymbol{s}_n,1}| &\le \int_{\|\boldsymbol{s}_n\| \le b_{1n}} |(Uh_2(\boldsymbol{s}_n))^T (V_n^T)^{-1} \nabla l_n(\boldsymbol{\theta})| e^{\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\hat{\boldsymbol{\theta}})} e^{\tilde{\ell}_n(\boldsymbol{\psi}) - \tilde{\ell}_n(\hat{\boldsymbol{\psi}})} d\boldsymbol{s}_n d\boldsymbol{x}_n \\ &\le C_1 \sup_{\boldsymbol{\theta}: \|\boldsymbol{s}_n\| \le b_{1n}} \| I_d - (V_n^T)^{-1} [-(\frac{\partial^2 \tilde{\ell}_n}{\partial\theta_i \partial\theta_j})(\theta^{*ij})] V_n^{-1} \| \\ &\times \int_{\|\boldsymbol{s}_n\| \le b_{1n}} \|\boldsymbol{s}_n\| e^{\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\hat{\boldsymbol{\theta}})} e^{\tilde{\ell}_n(\boldsymbol{\psi}) - \tilde{\ell}_n(\hat{\boldsymbol{\psi}})} d\boldsymbol{s}_n d\boldsymbol{x}_n. \end{aligned}$$

Using $(B2)$ and Lemma 4 part 2 we conclude that $I_{\boldsymbol{s}_n,1} \xrightarrow{p} 0$. Next, by $(B3)$, $(F1)$-$(F2)$ and Lemma 3.1 of Weng and Tsai (2008), there exists a constant $C_2 > 0$ such that

$$|I_{\boldsymbol{s}_n,2}| \le C_2 \int_{S \cap \{\|\boldsymbol{s}_n\| > b_{1n}\}} \|\boldsymbol{s}_n\|^{r_1} e^{\tilde{\ell}_n(\boldsymbol{\theta}) - \tilde{\ell}_n(\hat{\boldsymbol{\theta}})} e^{\tilde{\ell}_n(\boldsymbol{\psi}) - \tilde{\ell}_n(\hat{\boldsymbol{\psi}})} d\boldsymbol{s}_n d\boldsymbol{x}_n,$$

which using Lemma 4 part 2, converges to 0 in probability. Hence, $I_{\boldsymbol{s}_n} \xrightarrow{p} 0$.

Similarly, $I_{\boldsymbol{x}_n} \xrightarrow{p} 0$ follows from $(F1)$-$(F2)$, $(C2)$-$(C3)$, Lemma 4 part 2, Lemma 3.1 of Weng and Tsai (2008) and

$$(Q_n^T)^{-1} \nabla k_n(\boldsymbol{\psi}) = \left\{ I_q - (Q_n^T)^{-1} [-(\frac{\partial^2 \tilde{\ell}_n}{\partial\psi_i \partial\psi_j}(\psi^{*ij}))] Q_n^{-1} \right\} \boldsymbol{x}_n.$$

This completes the proof. □

Corollary 1. *Following Weng and Tsai (2008, Theorem 4.2 and 4.3) and Theorem 3, asymptotic normality of marginal approximate posterior distribution of the parameter vector for a GLMM, holds with priors that must be continuous and bounded but need not have compact supports or be differentiable.*

Now in the following, we can prove two Theorems 1 and 2.

Proof of Theorem 1

The proof of Theorem 1 is based on similar techniques of the proof of Theorem 3.

Proof of Theorem 2

The proof of Theorem 2 follows by combining (2) and (3) with Theorem 3.

Corollary 2. *Following Corollary 1 the asymptotic normality of marginal hybrid DC-based distribution of the parameter vector for a GLMM, holds with priors that must be continuous and bounded but need not have compact supports or be continuously differentiable.*

6. Discussion

Although Fong *et al.* (2009) gave a number of prescriptions for prior specification especially for variance components in a GLMM, but sometimes specification of a prior in these models is not straightforward. On the other side, the DC-based inferences are invariant to the choice of the priors. Computation and convergence, however, is an issue since the usual implementation of DC method is via MCMC. On the other hand, INLA provides precise estimates in seconds and minutes, even for models involving thousands of variables, in situations where any MCMC computation typically takes hours or even days. In this paper, we synthesized these two approach and introduced a new hybrid DC method so that its performance, according to the obtained results, is very good and inherits invariance property of DC method as well.

The benefits of our proposed method are the simplicity of implementation using R INLA package and to obtain MLE efficiently. The most available alternative methods to compute MLE in GLMMs, especially in models with crossed random effects, have disadvantages in the sense of consistent estimation, loss of efficiency, computational time required and convergence assessment, e.g. penalized quasi likelihood (Breslow and Clayton, 1993), composite likelihood (Bellio and Varin, 2005) and Monte Carlo expectation maximization (Booth *et al.*, 2001).

A disadvantage of the our work is that, according to INLA methodology, the prior distributions for fixed effects $\boldsymbol{\beta}$ must be Gaussian. However, we can use Gaussian priors with high variances to consider approximately flat priors and the results, theoretically, are invariant to the choice of the priors as well. Alongside good performance of DC method, Baghishani and Mohammadzadeh (2009) have mentioned some its limitations and their possible solutions. Although we did not discussed about selecting the number of clones, k, but, in general, selecting k under different circumstances such as sample size, random effects dimension and number of parameters needs further research.

In this paper, we assumed that the dimension of the random effects is fixed. But in some frameworks such as spatial models and spline smoothing models, the number of random effects (spline basis) increases with the sample size. Extending the hybrid DC method to such frameworks can be interesting.

Finally, as Ponciano *et al.* (2009) have noticed, nowadays, the choice between Bayesian and frequentist approaches in GLMMs is no longer a matter of feasibility but rather can be based on the philosophical views of researchers.

Acknowledgment

The first author wishes to thank the NTNU Department of Mathematical Sciences, Norway, for their support during his visit there.

References

Baghishani, H. & Mohammadzadeh, M. (2009). A data cloning algorithm for computing maximum likelihood in spatial generalized linear mixed models, *Submitted*.

Baghishani, H. & Mohammadzadeh, M. (2010). Asymptotic normality of posterior distributions for generalized linear mixed models, *Submitted*.

Bellio, R. & Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects, *Statistical Modelling,* **5**, 217-227.

Booth, J. G., Hobert, J. P. & Jank W. (2001). A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model, *Statistical Modelling* **1**, 333-349.

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

Crowder, M. J. (1978). Beta-binomial ANOVA for proportions, *Applied statistics,* **27**, 2437.

Fong, Y., Rue, H. & Wakefield, J. (2009). Bayesian inference for generalized linear mixed models, *Biostatistic,* doi:10.1093/biostatistics/kxp053.

Karim, M. R., & Zeger, S. L. (1992). Generalized linear models with random effects: Salamander mating revisited, *Biometrics,* **48**, 631644.

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.

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

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models,* Chapman and Hall, London.

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.

Rue, H. & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random fields models, *Journal of statistical planning and inference,* **137**, 3177-3192.

Rue, H., Martino, S., & Nicolas, Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the royal statistical society, Series B,* **71**, 1-35.

Thall, P. F., & Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion, *Biometrics,* **46**, 657671.

Tierney, L. & Kadane, J. B. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the American statistical association,* **81**, 82-86.

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.