Exact estimation procedures in a spatial mixed–effects probit model with binary outcomes

Rolando De la Cruz[*]   Susana Eyheramendy[†]   Cristian Meza[‡]   Felipe Osorio[§]

Abstract

Generalized linear mixed models form a general class of random effects models for discrete and continuous response in the exponential family. Spatial GLMM are an extension of such models that allows us to fit spatial-dependent data. A popular model in this class is the probit-normal model. In this study we develop a novel exact algorithm to estimate a probit spatial generalized linear mixed models (GLMM) that fits binary point-reference spatial data. The spatial dependence in this model is taking into account in the covariance matrix of the location-specific random effects. GLMM are generally hard to estimate due to the high-dimensional integrals involved. Popular methods such as PQL, Laplace, etc. overcome this problem by approximating these integrals from which biased estimators can be obtained. In this study we implement a stochastic version of the EM algorithm that allows to obtain the ML and REML estimates of the parameters in the model without incurring into any kind of approximations, therefore allowing us to obtain more reliable estimators.

Key Words: Binary outcomes, SAEM algorithm, Variance components, REML estimation.

1. Introduction

A popular class of GLMM is the probit–normal model for analyzing binary as well as ordinal data. Models for binary response variables are important in many fields of research, since subjects are often classified in two categories. Additionally, it is often the case that subjects are observed nested within clusters or are repeatedly assessed across time. For data that are clustered and/or longitudinal, mixed-effects regressions models have been developed primarily to model continuous but also dichotomous.

A convenient way to represent such model consists of discretizing a latent continuous distribution with a threshold. This has been extensively used in the biometric and econometric literature (see Ashford and Sowden, 1970; McFadden, 1989; Hausman and Wise, 1978).

Spatial GLMMs are commonly used for count or proportion data obtained over a continuous spatial domain (see for example Diggle et al., 1998; Zhang, 2002; Zhao et al., 2006). Non-Gaussian point-referenced spatial data are frequently modelled using GLMM with location-specific random effects. Then, the spatial dependence can be introduced in the covariance matrix of the random effects (see Gemperli and Vounatsou, 2003 for a logit spatial model).

Maximum likelihood estimation of parameters in GLMM is difficult as the exact likelihood function involves an intractable high–dimensional integration. Therefore, several approximations to

[*]Departamento de Salud Pública and Departamento de Estadística, P. Universidad Católica de Chile, CHILE

[†]Departamento de Estadística, P. Universidad Católica de de Chile, CHILE

[‡]Departamento de Estadística–CIMFAV, Universidad de Valparaíso, CHILE

[§]Departamento de Estadística–CIMFAV, Universidad de Valparaíso, CHILE

the likelihood function have been proposed in the literature (Schall, 1991; Breslow and Clayton, 1993; Wolfinger, 1992). Many of these approaches are reviewed in McCulloch (1997) and Rodríguez and Goldman (1995). The most frequently used methods are based on the first– or second– order Taylor expansion. Among them, the penalized quasi-likelihood (PQL) by Breslow and Clayton (1993) is one of the most popular. It approximates the high-dimensional integration using Laplace approximation. However as reported by several authors (Breslow and Lin, 1995; Rodríguez and Goldman, 1995; Raudenbush et al., 2000) this procedure usually produces estimates of variance components that are severely biased downward.

Numerical integration can also be used to perform the integration over the random–effects distribution. Gauss–Hermite quadrature can be used to approximate the above integral to any practical degree of accuracy.

Alternatively, Bayesian inference can be carried out with Markov Chain Monte Carlo (MCMC) and implemented via e.g. Gibbs Sampling (Albert and Chib, 1993).

The purpose of this article is to propose an exact estimation procedures to obtain ML and REML estimates in the spatial probit model for binary data using a stochastic version of EM, namely the SAEM algorithm (Delyon et al., 1999). This work is a natural extension of Meza et al. (2009). We introduce an alternative to the Monte Carlo EM algorithm (Natarajan and Kiefer, 2000 and Zhou and Liu, 2008) in the context of the spatial GLMM.

The structure of this article is as follows. Section 2 presents the spatial probit normal mixed model for binary outcomes and the SAEM algorithm is described in Section 3. A simulation study is presented in Section 4 to illustrate the properties of the REML estimates compared to ML in the spatial mixed–effects probit normal model.

2. Spatial mixed–effects probit model

Let y_{ij} be the outcome of a dichotomous variable corresponding, for example, to the mortality risk of child j at site s_i, $i = 1, \cdots, n$ taking value 1 if the child survived the first year of life and 0 otherwise. The probability p_{ij} of a positif event, i.e. $y_{ij} = 1$, can be expressed in terms of the standard normal cumulative distribution function (probit model).

We study the spatial GLMM with the probit link:

$$
\begin{aligned}
y_{ij} &\sim \mathrm{Ber}(p_{ij}), \quad \text{with} \\
p_{ij} &= P(y_{ij} = 1) \\
&= \Phi(X'_{ij}\boldsymbol{\beta} + \phi_i).
\end{aligned}
\tag{1}
$$

Then

$$
\Phi^{-1}(p_{ij}) = X'_{ij}\boldsymbol{\beta} + \phi_i, \quad 1 \le i \le n, \quad 1 \le j \le n_i
$$

where $\Phi(\cdot)$ denotes the normal cumulative distribution function and X_{ij} is a p known vector. The spatial variation is represented by the unobserved random vector $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_n) \in \mathbb{R}^n$ which follows a Gaussian distribution $\mathcal{N}_n(0, \boldsymbol{\Gamma})$, where $\Gamma_{ii^\star}$ is a parametric function of the distance $d_{ii^\star}$ between locations s_i and $s_{i^\star}$, with $i, i^\star = 1, \cdots, n$. This model can also be represented with a

latent variable $\boldsymbol{\omega}$ as follows:

$$y_{ij} = sign(\omega_{ij}) = \begin{cases} 1 \text{ if } \omega_{ij} > 0 \\ 0 \text{ si } \omega_{ij} \le 0 \end{cases}, \text{ with} \tag{2}$$

$$\omega_{ij} = X'_{ij}\boldsymbol{\beta} + \phi_i + \varepsilon_{ij},$$

where $\varepsilon_{ij} \sim \mathcal{N}(0,1)$ and ϕ_i and (ε_{ij}) are assumed to be mutually independent.
A commonly used parametrization for the covariance $\boldsymbol{\Gamma}$ is

$$\Gamma_{ii^\star} = \tau^2 \delta_{ii^\star} + \sigma^2 \exp(-\zeta d_{ii^\star}) \tag{3}$$

where $\delta_{ii^\star}$ is the Kronecker delta which takes the value of one if $i = i^\star$ and zero otherwise.
It is possible to separate in (2) a set of location–specific random effects to account for unexplained non–spatial variation as follows: let $\boldsymbol{\omega}_{i_{(n_i \times 1)}} = (\omega_{i1}, \cdots, \omega_{in})'$, $\boldsymbol{\eta}_{(n \times 1)} = (\eta_1, \cdots, \eta_n)'$ and $\boldsymbol{W}_{(n \times 1)} = (W_1, \cdots, W_n)'$

$$\begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{W} \end{pmatrix} \sim \mathcal{N}_{2n} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 R(\zeta) & 0 \\ 0 & \tau^2 I \end{pmatrix} \right) \tag{4}$$

where $R(\zeta) = (\rho_{ij})$ with $\rho_{ij} = \rho_{ij}(\zeta)$ (in (3), $\rho_{ij} = \exp(-\zeta d_{ij})$).

Then, using Equation (4), in the model (2) we have $\boldsymbol{b} = (\boldsymbol{\eta}', \boldsymbol{W}')'$ and $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2})$ with $Z_{ik} = (0 \cdots \mathbf{1}_{n_i} \cdots 0) \in \mathbb{R}^{n_i \times n}$:

$$\boldsymbol{\omega} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{5}$$

where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1', \cdots, \boldsymbol{\omega}_n')'$, $\quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_n \end{pmatrix}.$

Our purpose is then to compute the maximum likelihood (ML) and the restricted maximum likelihood (REML) estimates of the unknown parameter vectors $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\Gamma}(\theta)) = (\boldsymbol{\beta}, \tau^2, \sigma^2, \zeta)$.

3. Estimation procedure

To obtain both ML and REML estimates, we will use a stochastic approximation version of EM algorithm, the SAEM algorithm proposed by Delyon et al. (1999). This stochastic version of EM replaces the standard E–step of EM by a simulation step of the missing data under the conditional distribution given the observations and a stochastic aproximation step. When the simulation step cannot be directly performed, as in GLMMs, Kuhn and Lavielle (2004) propose to combine the SAEM algorithm with a Markov Chain Monte Carlo (MCMC) procedure.

Let $\boldsymbol{q} = (\boldsymbol{\omega}, \boldsymbol{b}) = (\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{W})$ be the vector of non–observed data and $\boldsymbol{y} = (y_{ij}, 1 \le i \le N, 1 \le j \le n_i)$ are the observed data. The vector of complete data is then $(\boldsymbol{y}, \boldsymbol{q}) = (\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{W})$ and the complete data log–likelihood can be written as:

$$l(\boldsymbol{\Theta}) = \log p(\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{W}; \boldsymbol{\Theta}) = \log p(\boldsymbol{y}|\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{W}; \boldsymbol{\Theta}) + \log p(\boldsymbol{\omega}|, \boldsymbol{\eta}, \boldsymbol{W}; \boldsymbol{\Theta}) + \log p(\boldsymbol{\eta}; \boldsymbol{\Theta}) + p(\boldsymbol{W}; \boldsymbol{\Theta}) \tag{6}$$

It can be noted that the first term of equation (6) can be ignored as it does not include any information about the parameters. In fact, $\boldsymbol{y}$ is completely specified, without any uncertainty, when $\boldsymbol{\omega}$ is known. It follows that:

$$
\begin{aligned}
l(\boldsymbol{\Theta}) \;=\; & -\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\omega}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i'\boldsymbol{b})^2 \\
& -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log|R(\zeta)| - \frac{1}{2\sigma^2}\boldsymbol{\eta}'R^{-1}(\zeta)\boldsymbol{\eta} \\
& -\frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2}\boldsymbol{W}'\boldsymbol{W}.
\end{aligned}
\tag{7}
$$

3.1 The ML–SAEM algorithm

At iteration k, the SAEM algorithm is composed by the following steps:

- ML Simulation step: we propose sampling missing values $\boldsymbol{q} = (\boldsymbol{\omega}, \boldsymbol{b})$, with $\boldsymbol{b} = (\boldsymbol{\eta}', \boldsymbol{W}')'$, given the observations $\boldsymbol{y}$ and the parameters at iteration k ($\boldsymbol{\Theta}^{(k)}$) using a Gibbs scheme. We therefore only need to sample from the conditional distributions $\boldsymbol{b}_i^{(k+1)}|\boldsymbol{\omega}^{(k)}, \boldsymbol{y}; \boldsymbol{\Theta}^{(k)}$ and $\omega_{ij}^{(k+1)}|\boldsymbol{b}_i^{(k+1)}, \boldsymbol{y}; \boldsymbol{\Theta}^{(k)}$. It can be shown using Henderson's mixed model equations that:

$$
\boldsymbol{b}_i^{(k+1)}|\boldsymbol{\omega}^{(k)}, \boldsymbol{y}; \boldsymbol{\Theta}^{(k)} \sim \mathcal{N}\left(\boldsymbol{V_i}^{(k)}\boldsymbol{Z}_i'(\boldsymbol{\omega}_i^{(k)} - \boldsymbol{X}_i\boldsymbol{\beta}^{(k)}), \boldsymbol{V_i}^{(k)}\right)
\tag{8}
$$

where matrix $\boldsymbol{V_i}^{(k)}$ is defined as: $\boldsymbol{V_i}^{(k)} = [\boldsymbol{Z}_i'\boldsymbol{Z}_i + \boldsymbol{\Gamma}^{-1,(k)}]^{-1}$ where

$$
\boldsymbol{\Gamma}^{(k)} = \begin{pmatrix} \sigma^{2(k)}R(\zeta^{(k)}) & 0 \\ 0 & \tau^{2(k)}I \end{pmatrix}
$$

$$
\boldsymbol{Z}'\boldsymbol{Z} = (\boldsymbol{Z}_1', \cdots, \boldsymbol{Z}_n')\begin{pmatrix} \boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_n \end{pmatrix} = \sum_{i=1}^{n}\boldsymbol{Z}_i'\boldsymbol{Z}_i
$$

where

$$
\boldsymbol{Z}_i'\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{Z}_{i1}^T\boldsymbol{Z}_{i1} & \boldsymbol{Z}_{i1}^T\boldsymbol{Z}_{i2} \\ \boldsymbol{Z}_{i2}^T\boldsymbol{Z}_{i1} & \boldsymbol{Z}_{i2}^T\boldsymbol{Z}_{i2} \end{pmatrix}.
$$

The conditional distribution for the latent variable can be defined with a truncated normal distribution as follows:

$$
\omega_{ij}^{(k+1)}|\boldsymbol{b}_i^{(k+1)}, \boldsymbol{y}; \boldsymbol{\Theta}^{(k)} \sim \mathcal{N}_{\mathcal{T}}\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta}^{(k)} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i^{(k+1)}, 1; 0\right)
\tag{9}
$$

A right truncated normal distribution $\mathcal{N}_{\mathcal{T}-}(.)$ will be used if observation y_{ij} is 0 and a left truncated normal distribution $\mathcal{N}_{\mathcal{T}+}(.)$ will be used if observation y_{ij} is equal to 1.

- ML Stochastic step: update the following statistics:

$$s_1^{(k+1)} = s_1^{(k)} + \gamma_{k+1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i(\boldsymbol{\omega}_i^{(k+1)} - \boldsymbol{Z}_i\boldsymbol{b}_i^{(k+1)}) - s_1^{(k)}\right)$$

$$s_2^{(k+1)} = s_2^{(k)} + \gamma_{k+1}\left(\boldsymbol{\eta}^{(k+1)} - s_2^{(k)}\right)$$

$$s_3^{(k+1)} = s_3^{(k)} + \gamma_{k+1}\left(\boldsymbol{W}^{(k+1)\prime}\boldsymbol{W}^{(k+1)} - s_3^{(k)}\right)$$

Parameter γ_{k+1} is a smoothing parameter, i.e. a decreasing sequence of positive numbers, as proposed by Kuhn and Lavielle (2004) to improve convergence in the SAEM algorithm:

$$\gamma_k = \begin{cases} 1 & \text{for } 1 \leq k \leq K_b \\ (k - K_b)^{-1} & \text{for } k \geq K_b + 1. \end{cases}$$

Here, $K_b + 1$ is the number of iterations to perform before starting the smoothing phase of the SAEM algorithm.

- ML Maximization step: Parameters are then updated at iteration k as follows

$$\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i'\left(s_1^{(k+1)}\right)$$

$$\sigma^{2(k+1)} = \frac{1}{n}s_2^{(k+1)\prime}R^{-1}(\zeta^{(k)})s_2^{(k+1)}$$

$$\tau^{2(k+1)} = \frac{1}{n}s_3^{(k+1)}$$

$$\zeta^{(k+1)} = \arg\min_{\zeta}\left\{-\frac{1}{2}\log|R(\zeta)| - \frac{1}{2\sigma^{2(k+1)}}s_2^{(k+1)\prime}R(\zeta)^{-1}s_2^{(k+1)}\right\}.$$

3.2 The REML–SAEM algorithm

In the framework of GLMM, the approximate likelihood techniques such as the penalized quasi–likelihood (PQL, Breslow and Clayton, 1993), are known to produce severely biased estimates of both regression parameters and variance components, particularly when the response is binary and/or the variance components are large (Breslow and Lin, 1995; Lin and Breslow, 1996; Neuhaus and Segal, 1997). The Restricted Maximum Likelihood (REML) estimation procedure permits, in linear and nonlinear mixed effects models, a reduction of the bias observed with the Maximum Likelihood (ML) estimation for variance components. Following Meza et al. (2009) and Meza et al. (2007), we use a conditional likelihood approach to implement REML in this kind of model. We rely on the concept of integrated likelihood to eliminate nuisance parameters by integrating out fixed effects (Harville, 1974; Berger and Wolpert, 1999). We consider the fixed effects as random, assuming that the prior distribution of $\boldsymbol{\beta}$ is noninformative. The vector

of parameters $\boldsymbol{\Theta}$ becomes $\boldsymbol{\Theta}^\star = (\sigma^2, \tau^2, \zeta)$, and the vector of missing data now includes $\boldsymbol{\beta}$ and will be noted $\boldsymbol{q}_{REML} = (\boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\beta})$.

To perform the simulation step of the algorithm, we use Henderson's mixed model equations, i.e.

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z} + \boldsymbol{\Gamma}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'\omega} \\ \boldsymbol{Z'\omega} \end{bmatrix}, \tag{10}$$

where $\hat{\boldsymbol{\beta}}$ is the Generalized Least Square (GLS) estimate of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{b}}$ is $\mathbb{E}(\boldsymbol{b}|\boldsymbol{\omega}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{y}; \boldsymbol{\Theta}^{\star(k)})$. Let $\boldsymbol{A} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{b}} \end{bmatrix}$ and $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z} + \boldsymbol{\Gamma}^{-1} \end{bmatrix}^{-1}$ the inverse of the coefficient matrix.

At iteration k of the proposed the SAEM–REML algorithm, the stochastic approximation step as follows:

$$\begin{aligned} s_{1_{REML}}^{(k+1)} &= s_{1_{REML}}^{(k)} + \gamma_{k+1} \left(\boldsymbol{\eta}^{(k+1)} - s_{1_{REML}}^{(k)} \right) \\ s_{2_{REML}}^{(k+1)} &= s_{2_{REML}}^{(k)} + \gamma_{k+1} \left(\boldsymbol{W}^{(k+1)\prime} \boldsymbol{W}^{(k+1)} - s_{2_{REML}}^{(k)} \right) \end{aligned}$$

As previously, a Gibbs scheme was used to draw the non–observed data $\boldsymbol{q}_{REML,m} = (\boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\beta})$ from the conditional distribution given the observations $\boldsymbol{y}$ and the parameters at iteration k ($\boldsymbol{\Theta}^{\star(k)}$). In this case, it can be shown that:

1) $(\boldsymbol{b}_i^{(k+1)}, \boldsymbol{\beta}^{(k+1)})|\boldsymbol{\omega}^{(k)}, \boldsymbol{y}; \boldsymbol{\Theta}^{\star(k)} \sim \mathcal{N}(\boldsymbol{A}, \boldsymbol{C})$ since $\boldsymbol{y}$ does not bring any additional information given the underlying variable $\boldsymbol{\omega}$.

2) $\omega_{ij}^{(k+1)}|\boldsymbol{b}_i^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{y}; \boldsymbol{\Theta}^{\star(k)} \sim \mathcal{N}_{\mathcal{T}}\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta}^{(k)} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i^{(k+1)}, 1; 0 \right)$, where $\mathcal{N}_{\mathcal{T}}$ is the truncated normal distribution as previously defined.

Parameters are then updated, at iteration k, as follows

$$\begin{aligned} \sigma^{2(k+1)} &= \frac{1}{n} s_{1_{REML}}^{(k+1)\prime} R^{-1}(\zeta^{(k)}) s_{1_{REML}}^{(k+1)} \\ \tau^{2(k+1)} &= \frac{1}{n} s_{2_{REML}}^{(k+1)} \\ \zeta^{(k+1)} &= \arg\min_{\zeta} \left\{ -\frac{1}{2}\log|R(\zeta)| - \frac{1}{2\sigma^{2(k+1)}} s_{1_{REML}}^{(k+1)\prime} R(\zeta)^{-1} s_{1_{REML}}^{(k+1)} \right\}. \end{aligned}$$

As proposed by Meza et al. (2007), estimation of fixed effects can be obtained in this SAEM-REML algorithm by their conditional expectation given the data vector and the variance-covariance components equal to their REML estimates, i.e. $\hat{\boldsymbol{\beta}}_{REML} = E(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\theta} = \hat{\boldsymbol{\Theta}}^\star)$. This estimator makes sense in an Empirical Bayes framework.

4. Simulation study

In order to investigate the performance of the proposed REML procedure, we simulate $M = 200$ data sets from a specific spatial model. We implement the SAEM–ML and SAEM–REML algorithms presented above and apply them to the simulated data. We first simulate $n = 15$ spatial

Figure 1: Simulations study: Simulated spatial locations.
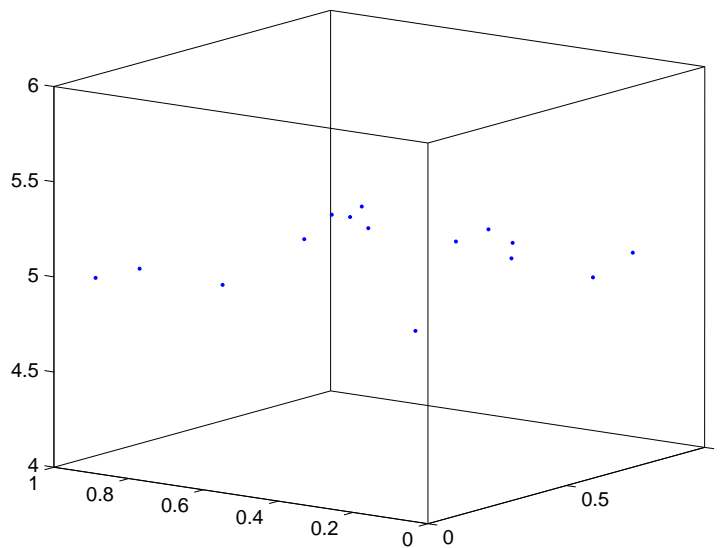
locations $s_i = (s_{i1}, s_{i2})$ with $s_{i1}, s_{i2} \sim_{i.i.d.} Uniform(0,1)$. The resulting spatial location s_i is thus a random point in the unit-square that we fix for all replications.

In this study, we consider the following structure for the spatial correlation:

$$\mathbf{\Gamma}(\theta) = \sigma^2 exp(-\zeta d)$$

with $\sigma^2 = 3$, $\zeta = 13$. In the model (5), we consider that $\beta = \beta_0 = -1$, $X_{ij} = 1$ and we have $n_i = m = 5$ observations in each location s_i. Figure 4 shows the simulated spatial locations .

In this example, for both ML and REML versions of SAEM, we used 200 iterations, and the following sequence (γ_k): $\gamma_k = 1$ for $1 \leq k \leq 100$ and $\gamma_k = 1/(k - 200)$ for $101 \leq k \leq 200$. Summary statistics for ML and REML estimates obtained for these 200 simulated data sets are given in Table 1.

The true values of the parameters used in the simulation, the means and standard deviations of the estimates are provided. It can be seen that the mean values for the REML estimates were closer to the simulated values for parameters ζ and β_0. In general, the ML estimates were found to be slightly more variable than the REML estimates and the covariance matrix $\mathbf{\Gamma}(\theta)$ is biased downwards using ML. Indeed, the bias observed for parameter ζ using SAEM–ML has a huge impact in the estimation of $\mathbf{\Gamma}(\theta)$ due to the exponential correlation function used in this example.

Figure 2 give a graphical representation of these results showing the density estimates obtained with ML and REML, for the parameters σ^2 and ζ.

Table 1: Simulations study: Summary statistics.

	Method	$\beta_0^\star$	σ^2	ζ
True Value		-1	3	13
Mean	ML	-0.840	2.906	19.912
	REML	-0.962	3.609	17.261
Standard dev.	ML	0.501	1.649	11.292
	REML	0.436	2.047	9.982

$\star$: empirical conditional mean for REML

Figure 2: Simulations study: Density estimates obtained with ML and REML for σ^2 and ζ.
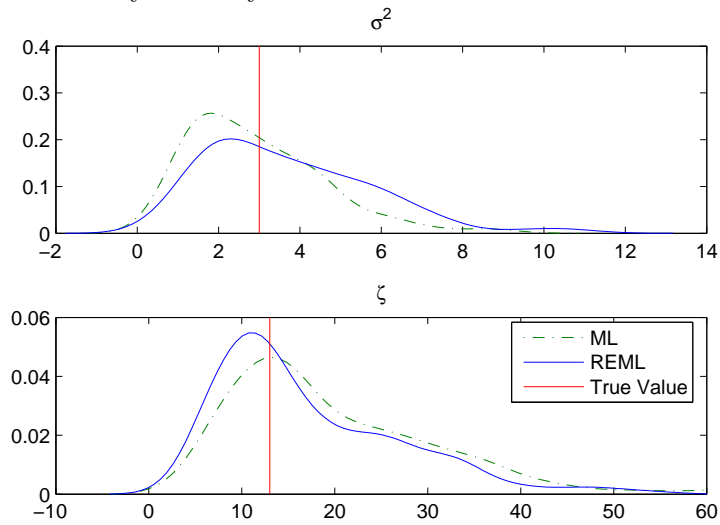
References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.

Ashford, J. and Sowden, R. (1970). Multivariate probit analysis. *Biometrics*, 26:535–546.

Berger, J. O., L. B. and Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14:1–28.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.*, 88:9–25.

Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82:81–91.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27:94–128.

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

Gemperli, A. and Vounatsou, P. (2003). Fitting generalized linear mixed models for point–referenced spatial data. *Journal of Modern Applied Statistical Methods*, 2:481–495.

Harville, D. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61:383–385.

Hausman, J. and Wise, D. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogenous preferences. *Econometrica*, 46:403–426.

Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM P&S*.

Lin, X. and Breslow, N. E. (1996). Bias correction in GLMM with multiple components of dispersion. *J. Amer. Statist. Assoc.*, 91:1007–1016.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *J. Amer. Statist. Assoc.*, 92:162–170.

McFadden, D. (1989). A method of simulated moments for estimation of discrete repsonse models without numerical integration. *Econometrica*, 57:995–1026.

Meza, C., Jaffrézic, F., and Foulley, J.-L. (2007). REML estimation of variance parameters in nonlinear mixed effects models using the SAEM algorithm. *Biometrical Journal*, 49:876–888.

Meza, C., Jaffrézic, F., and Foulley, J.-L. (2009). Estimation in the probit normal model for binary outcomes using the SAEM algorithm. *Computational Statistics and Data Analysis*, 53:1350–1360.

Natarajan, R., M. C. and Kiefer, N. (2000). A Monte Carlo EM method for estimating multinomial probit models. *Computational Statistics and Data Analysis*, 34:33–50.

Neuhaus, J. and Segal, M. (1997). An assessment of approximate maximum likelihood estimators in generalized linear models. In Gregoire, T. G., editor, *Modelling longitudinal and spatially correlated data: Methods, Applications and future directions. Lecture notes in Statistics, v. 122.* Springer, New York.

Raudenbush, S. W., Yang, M.-L., and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high–order, multivariate laplace approximation. *J. Comp. Graph. Stat.*, 9:141–157.

Rodríguez, G. and Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *J. Roy. Statist. Soc. A*, 158:73–89.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78:719–727.

Wolfinger, R. (1992). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80:791–795.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58:129–136.

Zhao, Y., Staudenmayer, J., Coull, B., and Wand, M. (2006). General design Bayesian generalized linear mixed models with applications to spatial statistics. *Statistical Science*, 21:35–51.

Zhou, X. and Liu, X. (2008). The Monte Carlo EM method for estimating multinomial probit latent variable models. *Computational Statistics*, 23:277–289.