1st Conference on Spatial Statistics 2011

Maximum Likelihood Estimation for Spatial GLM Models

Mohsen Mohammadzadeh\textsuperscript{a,}\textsuperscript{*}, Fatemeh Hosseini\textsuperscript{b}

\textsuperscript{a}Department of Statistics, Tarbiat Modares University, Tehran, Iran.
\textsuperscript{b}Department of Statistics, Semnan University, Semnan, Iran.

Abstract

Spatial generalized linear mixed models are usually used for modelling non-Gaussian and discrete spatial responses. In these models, spatial correlation of the data is usually modelled by spatial latent variables. Although, it is a standard assumption that the latent variables have normal distribution, in practice this assumption may not be valid. The first purpose of this paper is to use a closed skew normal distribution for the spatial latent variables which is more flexible distribution and also includes normal and skew normal distributions. The second is to develop Monte Carlo EM gradient algorithm for maximum likelihood estimation of the model parameters. Then, the performance of the proposed model is illustrated through a simulation study. Finally, the model and algorithm are applied to a case study on a temperature data.

© 2011 Published by Elsevier Ltd. Open access under \textsuperscript{CC BY-NC-ND license.}
Selection and/or peer-review under responsibility of Spatial Statistics 2011.

Keywords: Spatial Generalized Linear Mixed Model; Latent Variables; Closed Skew Normal Distribution;

1. Introduction

In most studies involving spatial generalized linear mixed (SGLM) models, it is assumed that the latent variables are normally distributed. Erroneous normal assumption, however, can affect the estimation of the model parameters and accuracy of spatial predictions. This motivates the search for SGLM models with more flexible distributions of the latent variables. Dominguez-Molina \textit{et al.} \cite{3} generalized the skew normal distribution (Azzalini and Dallavale \cite{1}) to define the closed skew normal (CSN) distribution, that forms a larger and more flexible class than both skew normal and normal distributions. The CSN distribution also remains closed under linear conditioning and marginalization. Hosseini \textit{et al.} \cite{4} used the skew normal distribution for the latent variables and proposed an approximate Bayesian method for the
inference and spatial prediction in SGLM model. Baghishani and Mohammadzadeh [2] proposed a computationally efficient strategy to fit SGLM models based on the data cloning method suggested by Lele et al. [6]. Zhang [9] proposed a Monte Carlo EM gradient (EMG) algorithm for computing the maximum likelihood (ML) estimate of SGLM model. In the present paper we propose an alternative approach for SGLM models, where the normal distribution is replaced by a CSN distribution for the latent variables. Then, the EMG algorithm is also extended for this model. Next, the performance of the proposed model is illustrated through a simulation study. Finally, the model and algorithm are applied on the winter temperature data in North-West provinces of Iran.

2. SGLM with Closed Skew Latent Variables

A random vector \( \mathbf{x} \) distributed according to a multivariate closed skew normal if its pdf is given by

\[
f(\mathbf{x} | \mu, \Sigma, \lambda) = k \phi_n(\mathbf{x}; \mu, \Sigma) \Phi_q(D(\mathbf{x} - \mu); \nu, \Delta),
\]

where \( k = \Phi_q^{-1}(0; \nu, \Delta + D \Sigma D') \), \( \phi_n(\cdot; \mu, \Sigma) \) is the n-dimensional normal density with mean vector \( \mu \) and covariance matrix \( \Sigma \), \( \Phi_q(\cdot; \nu, \Delta) \) is q-dimensional normal cumulative distribution function with mean \( \nu \) and covariance matrix \( \Delta \), and \( D \) is a \( q \times n \) matrix whose elements are the skewness parameters. In short, we denote this distribution by \( \text{CSN}(\mathbf{x}; \mu, \Sigma, \Delta, \lambda) \). For \( q = 1 \), \( \nu = 0 \), \( \Delta = 1 \) and \( D = \lambda \Sigma^{-1/2} \) the CSN density reduces to that of the SN distribution. When \( D \) is a zero matrix, the density in equation (1) reduces to the density of a multivariate normal distribution. The CSN distribution has some desirable properties, similar to those of the normal distribution. For instance, the CSN distribution is closed under marginalization, conditioning, and linear transformations (full column or row rank), see Dominguez-Molina et al. [3]. The first moment of the CSN distribution is \( \mathbb{E}(X) = \mu + \Sigma D' \psi \), where \( \psi = \Phi_q'(0; \nu, \Delta + D \Sigma D') / \Phi_q(\nu; \nu, \Delta + D \Sigma D') \), \( \Phi_q'(r; v, \Omega) = [\nabla \Phi_q(r; v, \Omega)]' \), such that \( \Omega \) is a positive definite matrix and \( \nabla_r \) gradient operator as \( \nabla_r = (\partial / \partial r_1, \ldots, \partial / \partial r_p)' \). Now, let \( x = (x_1, \ldots, x_p)' \) be spatial latent variables at \( n \) sites \( \{s_1, \ldots, s_n\} \) with distribution \( \text{CSN}_{n,q}(\mathbf{Z}; \Sigma_0, D, \nu, \Delta) \), from equation (1) we have

\[
f(\mathbf{x} | \eta) = k (2\pi)^{-n/2} |\Sigma_0|^{-1/2} \exp\left(-\frac{(\mathbf{x} - \mathbf{Z}\eta)' \Sigma_0 (\mathbf{x} - \mathbf{Z}\eta)}{2}\right) \Phi_q(D(\mathbf{x} - \mathbf{Z}\eta); \nu, \Delta),
\]

where \( k = \Phi_q^{-1}(0; \nu, \Delta + D \Sigma D') \), \( \mathbf{Z} \) is an \( n \times (p+1) \) observations matrix of \( p + 1 \) explanatory variables, \( \eta = (\beta_0, \ldots, \beta_p)' \) are regression parameters, \( \Sigma_0 \) is a positive definite \( n \times n \) covariance matrix, with two dimensional parameter \( \theta = (\sigma, \rho) \) indicative of the scale and spatial ‘correlation’ length, respectively, and \( \eta = (\beta, \theta, D, \nu, \Delta) \). We consider the situation where sites \( \{s_1, \ldots, s_k\} \) are observation sites, while one of our goals is to predict the spatial latent variables at the unobserved sites \( \{s_{k+1}, \ldots, s_n\} \). The latent variables at the \( k \) observed sites are denoted by \( \mathbf{x}^{obs} = \mathbf{A} \mathbf{x} \), where \( \mathbf{A} = [I_k | 0_{k \times (n-k)}] \). Thus, the matrix \( A \) decomposes \( \mathbf{x} \) into \( \mathbf{x} = (\mathbf{x}^{obs}, \mathbf{x}^{pred})' \), where \( \mathbf{x}^{obs} = (x_1, \ldots, x_k)' \) are latent variables at \( k \) observation sites \( \{s_1, \ldots, s_k\} \) and \( \mathbf{x}^{pred} = (x_{k+1}, \ldots, x_n)' \), are latent variables at \( n-k \) prediction sites.

Also, let \( \mathbf{y} = (y_1, \ldots, y_n)' \) represent the discrete spatial response variables at \( k \) sites \( \{s_1, \ldots, s_k\} \), \( k < n \). We assume that the measurements are conditionally independent with likelihood \( \pi(y | x) \) of an exponential family (McCullagh and Nelder [7]), given by

\[
f(y_i | x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}, \quad i = 1, \ldots, k,
\]
where \( b(*) \) and \( c(*) \) are known functions and \( b(x_i) \) is the cumulant function, such that
\[
\frac{\partial b(x_i)}{\partial x_i} = E(y_i \mid x_i), \quad \frac{\partial^2 b(x_i)}{\partial x_i^2} = \text{Var}(y_i \mid x_i).
\]
The mean \( E(y_i \mid x_i) \) and \( x_i \) are related by a link function \( g(*) \), i.e. \( E(y_i \mid x_i) = g^{-1}(x_i) \). Therefore, the model is given by
\[
f(y, x \mid \eta) = f(y \mid x)f(x \mid \eta)
\]
\[
x \sim k \mid \Sigma_0 \mid ^{-1/2} \exp \left[ \sum_{i=1}^{n} [y_i - b(x_i) + c(y_i)] - (x - Z\beta)'(x - Z\beta) / 2 \right] \Phi_q \left( D(x - Z\beta); \nu, \Delta \right).
\]

### 3. Prediction for SGLMM with Closed Skew Latent Variables

An important problem in SGLM model is the prediction of the spatial latent variables at \( n - k \) sites \{\( s_{k+1}, \ldots, s_n \)\}, using observations \( y \) and \( Z \). First, we assume that the model parameters are known. Considering \( x = (x^{obs}, x^{pred}) \), we can write
\[
Z\beta = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma_0 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}.
\]

On the other hand, when the joint distribution of \( x \) is CSN, the conditional distribution of \( x^{pred} \mid x^{obs} \), is also within the CSN class (Domínguez-Molina et al. [3]). Therefore,
\[
x^{pred} \mid x^{obs} \sim \text{CSN}_{1, q}(\mu_2 + \Sigma_{21}^{-1}(x^{obs} - \mu_1), \Sigma_{22}^{-1}, D_2, v - D^*(x^{obs} - \mu_1), \Delta),
\]
where \( \Sigma_{22} = \Sigma_{22} - \Sigma_{21}^{-1}\Sigma_{12} \) and \( D^* = D_1 + D_2\Sigma_{21}^{-1} \). Using the expectation of the CSN distribution, we get
\[
E(x^{pred} \mid x^{obs}) = \mu_2 + \Sigma_{21}^{-1}(x^{obs} - \mu_1) + \Sigma_{22}D_2\Psi,
\]
where \( \Psi = \Phi_q'(r; v - D^*(x^{obs} - \mu_1), \Delta + D_2\Sigma_{22}D_2, z) / \Phi_q'(r; v - D^*(x^{obs} - \mu_1), \Delta + D_2\Sigma_{22}D_2, z) \). Due to the model formation, the conditional distribution of \( y \) given \( x \) is the conditional distribution of \( y \) given \( x^{obs} \), this implies \( E(x^{pred} \mid x^{obs}) = E(x^{pred} \mid x^{obs}, y) \) and from the well-known fact \( E(x^{pred} \mid y) = E(E(x^{pred} \mid x^{obs}, y) \mid y) \) (Shiryayev [8], p. 214). Now, by minimizing the mean squared error the MMSE prediction of \( x^{pred} \) at \( n - k \) sites \{\( s_{k+1}, \ldots, s_n \)\} is given by
\[
E(x^{pred} \mid y) = \mu_2 + \Sigma_{21}^{-1}(E(x^{obs} \mid y) - \mu_1) + \Sigma_{22}D_2'E(\Psi \mid y),
\]
where \( E(x^{obs} \mid y) \) and \( E(\Psi \mid y) \) can not be given in closed form but can be approximated by Monte Carlo samples.
4. Maximum Likelihood Estimations

In this section we explore the Monte Carlo EMG algorithm for ML estimation of model parameters when the latent variables have a CSN distribution. We assume an isotropic exponential covariance function for the entries in the matrix $\Sigma$. This entails that $C(s_i, s_j; \theta) = \sigma^2 \exp(-\|s_i - s_j\|/\phi)$, where $\|s_i - s_j\|$ is the Euclidean norm, $\phi$ is the spatial correlation parameter and $\sigma^2$ is the variance. Also, we assume that the vector of skewness parameters is of the form $\lambda = \lambda_0 1$, where $1$ is a $n \times 1$ vector of ones. Then $\eta = (\beta, \theta, \lambda_0)$, where $\theta = (\sigma, \phi)$. The likelihood function (3), cannot usually be given in a closed form. The EM algorithm is a general algorithm for ML estimation where the data are incomplete or the likelihood function involves latent variables. In many practical applications, unfortunately, convergence of the EM algorithm can be extremely slow. Lange [5] proposed the EMG algorithm to accelerate convergence of the EM algorithm. This algorithm substitutes one step of Newton-Raphson algorithm for the M-step. From (3) the complete log-likelihood function $f(\eta) = \sum_{i=1} f(y_i | x_i) + \ln f(x | \eta)$. Now, from the EMG algorithm we have

$$\eta^{(m+1)} - \eta^{(m)} = \left[ E(\partial^2 \ln f(x | \eta) / \partial \eta \partial \eta^t | y)) \right]^{-1} \left[ E(\partial \ln f(x | \eta) / \partial \eta | y) \right]_{\eta=\eta^{(m)}}.$$

The conditional expectation in (5) cannot usually be evaluated in closed form, but can be approximated using the N Monte Carlo samples $x(1), \ldots, x(N)$ from the Metropolis-Hastings algorithm under the current estimates $\eta^{(m)}$.

5. Simulation Study

In this section, a simulation study is performed to assess the proposed model. First, $n = 100$ locations are generated on a $10 \times 10$ regular grid as $\{(\ell, k), \ell, k = 1, \ldots, 10\}$. We fix parameters of the CSN model and draw latent variables from distribution $CSN_{100,1}(\beta \Sigma, \Sigma, \lambda_0 \Sigma^{-1/2}, 0)$. In the simulation we set $\beta = -0.5$, $\theta = (\sigma, \phi) = (2, 4)$, using an isotropic exponential covariance function, the $\lambda_0 = 2$. The covariate variable at $(\ell, k)$ th site is $z_{(\ell, k)} = \log(1 + \ell)$. Given the latent variables, the binomial responses $y_{i(k)}$ are generated according to $Bin(100, \exp(x_{i(k)}) / 1 + \exp(x_{i(k)}))$. Simulated data set is analyzed by the SGLMM under the assumptions of CSN and Normal latent variables. We generated 100 datasets to estimate MSE and mean square prediction errors (MSPE). For the ML estimation of the parameters, we run the Monte Carlo EMG algorithm. With respect to the ML estimates, biases and MSE of estimated parameters, reported in Table 1, then, CSN approach leads to the predictions which are often more accurate than the predictions provided by the Normal approach. To check the accuracy of the MMSE prediction, we use all 100 datasets to calculate the MSPE at an arbitrary site $(5.5, 5.5)$ for CSN and Normal SGLMMs. The obtained values 1.5432 and 1.9241, show that the SGLMM with CSN latent variables is more accurate than the normal latent variables.

| Parameters | Real Value | CSN SGLM | N SGLM |
|------------|------------|----------|--------|
| $\beta$    | -0.5       | -0.0763  | -0.0887|
| $\sigma^2$ | $\sqrt{2}$ | -0.0305  | -0.1236|
| $\varphi$  | 4          | 0.1523   | 0.3919 |
| $\lambda$  | 2          | -0.4574  | 1.2841 |
6. Temperature Data

Temperature, an important meteorological parameter, is useful for the researchers working on many disciplines. For example, the late chilling effected by the winter temperature usually causes decrease of some agricultural yields. In this study, spatial prediction is implemented using SGLMM with skew normal latent variables to study the cold severity on January 2009 in the North-West provinces of Iran.

We used the temperature data of 31 days in January 2009 at 29 weather stations located in the North-West provinces of Iran. The data consist of number of days with minimum temperature less than or equal to $-4^\circ$, observed at 29 sites. The average of dewpoint temperature ($z$) is used as covariate. The responses are assumed to be conditionally independent Binomial variable, i.e.

$$\pi(y_i | x_i) = \exp\{y_i x_i - u_i \log(1 + \exp(x_i))\},$$

where $u_i = 31$, $i = 1, \ldots, 29$. A CSN distribution is used for latent variables, i.e. $(x | \eta) \sim CSN_{29/1}(\beta \Sigma, \lambda^\gamma 1 \Sigma^{-1/2}, 0, 1)$. Also, we use Normal prior for the latent variables and compare the prediction accuracy with the case of CSN prior. We also use an isotropic exponential covariance function with parameter $\theta = (\varphi, \sigma)$. Using the Monte Carlo EMG algorithm with a Monte Carlo sample size 10000, we obtained the estimates $\hat{\beta} = (18, 1.0971)$, $\hat{\theta} = -0.5152$ and $\hat{\lambda} = 2.1365$ for SGLM model with CSN latent variable and $\hat{\beta} = (21, 1.4015)$ and $\hat{\theta} = -0.6321$ for SGLM model with Normal latent variable. The CVMSE of the MMSE prediction based on CSN and Normal models are derived as 0.31079 and 0.3317, respectively. These values show that the CSN model provides more accurate results than the Normal model. The prediction maps of $x$ with their prediction at thirteen unobserved sites, that have dewpoints, are shown in Figure 1. We can see the cold severity on January in the North-West provinces of Iran.

![Fig.1. Prediction map of temperature data in the North-West of Iran for $x$.](image)
7. Conclusion

It is shown that misspecification of spatial latent variables in SGLMM affects on the estimation of the parameters as well as the spatial prediction of the spatial latent variables. However, Normal prior is often used for latent variables in SGLMM, here a CSN prior, that is more general than a Normal prior, is proposed. Then, MMSE spatial prediction and ML estimation of the model parameters are derived by Monte Carlo EMG algorithm. In two examples, we compared results obtained from the CSN prior with that of the Normal prior by using CVMSE and MSE criteria. The CVMSE of the prediction and MSE of the estimations reveal the high accuracy of the CSN SGLMM comparing with the Normal SGLMM.

Acknowledgments

Partial support from Ordered and Spatial Data Centre of Excellence of Ferdowsi University of Mashhad is acknowledged.

References

[1] Azzalini A., Dalla Valle A. The Multivariate Skew Normal Distribution. *Biometrika*, 1996; 83: 715-26.

[2] Baghishani H., Mohammadzadeh M. A Data Cloning Algorithm for Computing Maximum Likelihood Estimates in Spatial Generalized Linear Mixed Models. *Computational Statistics and Data Analysis*, 2011; 55: 1748–59.

[3] Dominguez-Molina J., González-Farias G., Gupta A. The Multivariate Closed Skew Normal Distribution. *Technical Report 3-12*, Department of Mathematics and Statistics, Bowling Green State University, 2003.

[4] Hosseini F., Eidsvik J., Mohammadzadeh M. Approximate Bayesian Inference in Spatial GLMM with Skew Normal Latent Variables. *Computational Statistics and Data Analysis*, 2011; 55: 1791–806.

[5] Lange K. A Gradient Algorithm Locally Equivalent to the EM Algorithm. *Journal of the Royal Statistical Society*, 1995; Series B 57: 425-37.

[6] Lele, S.R., Dennis, B., Lutscher, F., Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models using Bayesian Markov Chain Monte Carlo Methods. *Ecology Letters, 2007; 10*: 551–63.

[7] McCullagh P., Nelder J.A. *Generalized Linear Models*. Chapman and Hall, London; 1989.

[8] Shiryayev A.N. *Probability*. New York: Springer Verlag; 1984.

[9] Zhang H. On Estimation and Prediction for Spatial Generalized Linear Mixed Models. *Biometrics*, 2002; 58. No.1, 129-36.