

# Parameter Estimation in Spatial Generalized Linear Mixed Models with Skew Gaussian Random Effects using Laplace Approximation

Seyed Reza Hosseini Shojaei<sup>†</sup>, Yadollah Waghei<sup>†,\*</sup> and  
Mohsen Mohammadzadeh<sup>‡</sup>

<sup>†</sup> University of Birjand

<sup>‡</sup> Tarbiat Modares University

Received: 8/5/2016      Approved: 10/16/2017

**Abstract.** Spatial generalized linear mixed models are used commonly for modelling non-Gaussian discrete spatial responses. We present an algorithm for parameter estimation of the models using Laplace approximation of likelihood function. In these models, the spatial correlation structure of data is carried out by random effects or latent variables. In most spatial analysis, it is assumed that random effects have Gaussian distribution, but the assumption is questionable. This assumption is replaced in the present work, using a skew Gaussian distribution for the latent variables, which is more flexible and includes Gaussian distribution. We examine the proposed method using a real discrete data set.

**Keywords.** Laplace approximation; multivariate skew Gaussian; random effects; SGLM; spatial data.

MSC 2010: 62M30; 62H11.

---

\* Corresponding author

## 1 Introduction

Spatial modelling and prediction is an important problem in many sciences such as climate studies, geographical sciences, petroleum engineering and mining. Usually, the spatial data are considered to be a realization from a random field, and focus is on prediction of the random variable at the new locations or regions. The most common technique for statistical spatial prediction is kriging (Cressie, 1993).

The classical linear models are inappropriate for count and binary data and it is necessary to consider other models. Generalized linear models (GLM) are widely used for analysing counts and other types of discrete data, and thus provide a good starting point (McCullagh and Nelder, 1989). The generalized linear mixed model (GLMM) is a natural extension of the GLM and the linear mixed model which successfully combines their capabilities to model non-Gaussian correlated data. For the classical linear models, the maximum likelihood estimation (MLE) of the model parameters can be analytically found in the closed form. This is not the case for the GLM and the GLMM. Thus numerical methods are often used instead.

The GLMMs with spatially correlated random effects are employed to encounter discrete and continuous spatial data. The random effects for the data are usually considered as a zero mean Gaussian random field. There have been many attempts in making inference for such spatial models. Several approaches have been proposed to find the MLE in GLMMs. Some are approximate inferences as Breslow and Clayton (1993) considered an approximate approach called penalized quasi-likelihood (PQL) for inference in Spatial Generalized Linear Mixed (SGLM) models. Some incorporate Monte Carlo methods into the EM algorithm to obtain the MLE. See, for example, Chan and Ledolter (1995) for time series models with latent correlated random effects, McCulloch (1997) for GLMM with independent random effects, and Zhang (2002) for SGLM models. Diggle et al. (1998) employed an SGLM model with Gaussian latent variables for spatially dependent non-Gaussian variables observed in a continuous region.

Likelihood analysis for SGLM models was proposed by Christensen (2004) based on Markov Chain Monte Carlo (MCMC) algorithms. Christensen and Waagepetersen (2002) and Christensen (2004) illustrated the usefulness of the MCMC method for parameter estimation and model exploration in SGLM models. Both MCMC likelihood and Bayesian algorithms are implemented for Poisson and binomial models in the R (R Core Team, 2015) pack-

age `georglm`. Alternative methods were proposed by Zhang (2002) adopting a spatial expectation maximization algorithm. Recently, Baghishani and Mohammadzadeh (2011) and Torabi (2015) proposed a different approach based on the data clone algorithm (Lele et al., 2010).

The Laplace approximation is frequently applied in the context of longitudinal data analysis, however it is seldom applied in the context of SGLM models (Bonat and Ribeiro, 2015).

Most researchers are satisfied with using a Gaussian distribution for the random effects, but in many applications it is unclear whether or not the Gaussian assumption holds. This assumption is replaced in the present work using a skew Gaussian distribution, which is more flexible and forms a larger class than Gaussian distribution.

A model based on the skew Gaussian distribution has been proposed by Kim and Mallic (2004) to handle skewed continuous spatial data. Kim and Mallic (2004) developed a skew Gaussian random field using the multivariate skew Gaussian distribution (Azzalini, 1985), and presented an approach for spatial interpolation. Even if their model has an appealing construction, there are some problems in use of such skew Gaussian process (see Zhang and El-Shaarawi, 2010, for more details). To overcome the mentioned problems, some progress has been made by introducing a more general class of distribution. For example, Allard and Naveau (2007) introduced a spatial skew Gaussian process based on the multivariate Closed Skew Gaussian (CSG) and presented approaches for spatial interpolation. The skew Gaussian random field introduced in Kim and Mallic (2004) is a special case of the CSG random field. Hosseini et al. (2011) used an approximate Bayesian approach in SGLM models where they considered a closed skew Gaussian distribution for the spatial latent variables. Mohammadzadeh and Hosseini (2011) applied Monte Carlo EM gradient algorithm to making inference of the SGLM models using a closed skew Gaussian distribution.

In this article, we use Laplace approximation of likelihood function in the context of GLMMs with spatially random effects for geostatistical data. In Section 2, we first describe the SGLM models in general form. Then, we explain how skew Gaussian distribution can be used in SGLM models in terms of random effects and also how to estimate the parameters in Section 3. The performance of maximum likelihood estimation, is evaluated by using a real dataset in Section 4. Finally, some concluding remarks are given in Section 5.

## 2 Spatial Generalized Linear Mixed Model

Spatial generalized linear mixed models are commonly used for non-Gaussian spatial data such as count and binary data acquired over a continuous spatial domain (Diggle et al., 1998). The hierarchical model proposed by Diggle et al. (1998) is the prime example for SGLM model. Let  $Y(\mathbf{s}_i)$  be a spatial random variable at location  $\mathbf{s}_i$  for  $i = 1, \dots, n$ . It is assumed that the components of  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$  are conditionally independent, given the latent variable  $\delta(\mathbf{s}_i)$ , with Probability Density Function (pdf)

$$f(y(\mathbf{s}_i); \eta(\mathbf{s}_i), \psi(\mathbf{s}_i)) = \exp \left\{ \frac{(y(\mathbf{s}_i)\eta(\mathbf{s}_i) - b(\eta(\mathbf{s}_i)))}{a(\psi(\mathbf{s}_i))} + c(y(\mathbf{s}_i), \psi(\mathbf{s}_i)) \right\}, \quad (1)$$

for some specific functions  $a(\cdot)$ ,  $b(\cdot)$  and  $c(\cdot)$  (McCullagh and Nelder, 1989). The pdf (1) covers well-known distributions such as Gaussian, Binomial and Poisson. The pdf (1) is parameterized with respect to the canonical parameters  $\eta(\mathbf{s}_i)$  and scale parameters  $\psi(\mathbf{s}_i)$ . If  $\psi(\mathbf{s}_i)$  is known, this is a one parameter exponential family model with canonical parameter  $\eta(\mathbf{s}_i)$  which is equal to linear predictor. The linear predictor is related to the distribution parameter through the link function. Therefore, in a SGLM model, relation between the spatial variables and known location-dependent covariates  $\mathbf{x}(\mathbf{s}_i) = (x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i))^T$ , with  $x_1(\mathbf{s}_i) = 1$  is

$$g(E(Y(\mathbf{s}_i))) = \boldsymbol{\beta}^T \mathbf{x}(\mathbf{s}_i) + \delta(\mathbf{s}_i),$$

where  $g$  is the link function and  $\boldsymbol{\beta} \in \mathbb{R}^p$  are unknown regression parameters.

In most studies, it is assumed that the latent variables  $\boldsymbol{\delta} = (\delta(\mathbf{s}_1), \dots, \delta(\mathbf{s}_n))^T$  are Gaussianly distributed in SGLM models. Based on these assumptions, the random part of the linear predictor is Gaussian with zero mean and covariance matrix  $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2 \mathbf{R}(\varphi)$ , where  $\mathbf{R}$  is a correlation matrix with elements  $R_{ij} = \rho(\mathbf{s}_i - \mathbf{s}_j; \varphi)$  and  $\rho(\cdot; \varphi)$  is a valid spatial correlation function on  $\mathbb{R}^2$  indexed by a parameter  $\varphi$ . The parameters  $\boldsymbol{\theta} = (\sigma^2, \varphi)$  are sometimes called the partial sill and range, respectively.

In this paper, we illustrate parameter estimation can be affected in SGLM models with more flexibility in distributions of the spatial latent variables. Thus, we suppose that  $\boldsymbol{\delta}$  to be realizations from a skew Gaussian process and call them Skew-SGLM models. The skew Gaussian processes are a collection of random variables where any finite collection of them has a joint multivariate skew Gaussian distribution (that is, every linear combination has a skew Gaussian distribution).

## 2.1 Skew Gaussian Distribution

The univariate skew Gaussian distribution was introduced by Azzalini (1985) and the multivariate skew Gaussian distribution is defined by Azzalini (1985), as a generalization of the multivariate Gaussian distribution. Several authors have proposed different multivariate skew Gaussian distributions and many of these generalizations are summarized in Arellano-Valle and Azzalini (2006). We consider the multivariate skew Gaussian distribution introduced in Azzalini and Capitanio (2014).

A random vector  $\boldsymbol{\delta}$  is said to be distributed according to the multivariate skew Gaussian probability law if its pdf is given by

$$f_{\boldsymbol{\delta}}(\boldsymbol{\delta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_n(\boldsymbol{\delta}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu})), \quad (2)$$

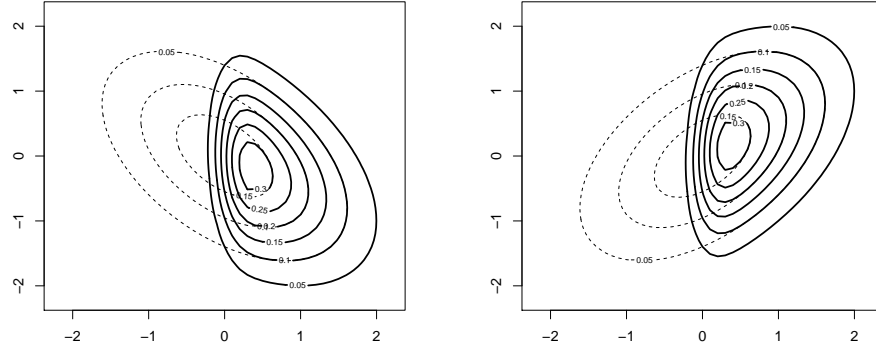
where  $\phi_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the multivariate Gaussian density with mean  $\boldsymbol{\mu} \in \mathbb{R}^n$  and covariance matrix  $\boldsymbol{\Sigma}$ ,  $\Phi(\cdot)$  represents the cumulative distribution function of a univariate standard Gaussian distribution,  $\boldsymbol{\lambda}$  is a  $n$ -dimensional skewness parameter, and  $\boldsymbol{\omega}$  is square root of diagonal of  $\boldsymbol{\Sigma}$  that can be written as  $\boldsymbol{\omega} = (\boldsymbol{\Sigma} \odot I_n)^{1/2}$ , where  $I_n$  is an  $n \times n$  identity matrix. We denote a random vector  $\boldsymbol{\delta}$  with pdf as in (2) by  $\boldsymbol{\delta} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ . The skew Gaussian distribution has some desirable properties (see for example Azzalini and Capitanio, 2014, for more details). When  $\boldsymbol{\lambda} = \mathbf{0}$ , the skew Gaussian distribution reduces to the Gaussian distribution.

The shape of the multivariate skew Gaussian density depends on the combined effect of  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\lambda}$ . For the bivariate case, a graphical illustration of the interplay between these components is provided in Figure 1, which shows the contour plots of two densities having the same parameter set except  $\Sigma_{12}$ . In each panel the dashed grey line represents the contour plot of the corresponding modulated bivariate Gaussian distribution.

The class of CSG as an extension of the multivariate skew Gaussian distribution inherits important properties from the multivariate Gaussian distribution, such as being closed under marginalization, conditioning, and linear transformations (González-Farías et al., 2004). However, we have not considered this class in the paper.

## 3 Parameter Estimation

To estimate the model parameters, we maximize the likelihood function. The joint distribution  $f_{\mathbf{Y}}(\mathbf{y})$  of the Skew-SGLM models is given by



**Figure 1.** Contour plot of bivariate skew Gaussian density functions when  $\boldsymbol{\mu} = (0, 0)$ ,  $\boldsymbol{\lambda} = (5, 0)$ ,  $\Sigma_{11} = \Sigma_{22} = 1$  and left:  $\Sigma_{12} = -0.5$ , right:  $\Sigma_{12} = 0.5$ .

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \int_{\mathbb{R}^n} f_{\mathbf{Y}|\boldsymbol{\delta}}(\mathbf{y}; \boldsymbol{\beta}) f_{\boldsymbol{\delta}}(\boldsymbol{\delta}; \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}, \boldsymbol{\lambda}) d\boldsymbol{\delta}. \quad (3)$$

We assume that the distribution  $f_{\mathbf{Y}|\boldsymbol{\delta}}(\cdot)$  can be written in the form of the exponential family, as follows:

$$f_{\mathbf{Y}|\boldsymbol{\delta}}(\mathbf{y}; \boldsymbol{\beta}) = \exp \{ \mathbf{y}^T (X\boldsymbol{\beta} + \boldsymbol{\delta}) - \mathbf{1}^T b(X\boldsymbol{\beta} + \boldsymbol{\delta}) + \mathbf{1}^T c(\mathbf{y}) \},$$

where  $b(\cdot)$  and  $c(\cdot)$  are known functions and  $\mathbf{1}$  denotes an  $n$ -dimensional column vector whose elements are all 1. Also, we consider  $f_{\boldsymbol{\delta}}(\cdot)$  to be a zero-centered multivariate skew Gaussian pdf is given by the following:

$$f_{\boldsymbol{\delta}}(\boldsymbol{\delta}; \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}, \boldsymbol{\lambda}) = 2(2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\delta}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta} \right\} \Phi(\boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta}).$$

Then the marginal likelihood function (3) is as follow,

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \int_{\mathbb{R}^n} \exp \{ q(\boldsymbol{\delta}) \} d\boldsymbol{\delta},$$

where

$$q(\boldsymbol{\delta}) = \mathbf{y}^T(X\boldsymbol{\beta} + \boldsymbol{\delta}) - \mathbf{1}^T b(X\boldsymbol{\beta} + \boldsymbol{\delta}) + \mathbf{1}^T c(\mathbf{y}) + \log(2) \\ - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_{\boldsymbol{\theta}}| - \frac{1}{2} \boldsymbol{\delta}^T \Sigma_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta} + \log(\Phi(\boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta})).$$

It is clear from (3) that the likelihood function involves intractable integrals. The computational burden also increases with the number of observations as the dimension of the random field is equal to the number of observations. The integral has a high dimension and consequently it is intractable to find the MLE by directly maximizing  $L$  or  $\log(L)$ .

### 3.1 Laplace Approximation

The Laplace method has been designed to approximate integrals in the form as follows:

$$\int_{\mathbb{R}^n} \exp\{q(\boldsymbol{\delta})\} d\boldsymbol{\delta} \approx (2\pi)^{n/2} \left| -H(\hat{\boldsymbol{\delta}}) \right|^{-1/2} \exp\{q(\hat{\boldsymbol{\delta}})\},$$

where

$$H(\hat{\boldsymbol{\delta}}) = \left. \frac{\partial^2 q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right|_{\boldsymbol{\delta}=\hat{\boldsymbol{\delta}}},$$

and  $q(\boldsymbol{\delta})$  is a known, uni-modal, and bounded function of a  $n$ -dimensional variable  $\boldsymbol{\delta}$  and  $\hat{\boldsymbol{\delta}}$  is the value for which  $q(\boldsymbol{\delta})$  is maximized.

The idea is to approximate the integrand to obtain a tractable closed-form expression of the integral, allowing the numerical maximization of the marginal likelihood.

We adopt an efficient Newton-Raphson (NR) algorithm to find  $\hat{\boldsymbol{\delta}}$  that maximize  $q(\boldsymbol{\delta})$ , although other numerical maximization methods can be used. The NR algorithm consists of an iterative scheme as follows:

$$\boldsymbol{\delta}_{i+1} = \boldsymbol{\delta}_i - H(\boldsymbol{\delta}_i)^{-1} J(\boldsymbol{\delta}_i), \quad \text{for } i = 1, 2, \dots,$$

until convergence, which gives  $\hat{\boldsymbol{\delta}}$ . The generic expressions for the derivatives required by the NR algorithm are given by

$$J(\boldsymbol{\delta}) = \frac{\partial q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = \mathbf{y} - \frac{\partial b(X\boldsymbol{\beta} + \boldsymbol{\delta})}{\partial \boldsymbol{\delta}} - \Sigma_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta} + \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\lambda} k,$$

and

$$H(\boldsymbol{\delta}) = \frac{\partial^2 q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = -\text{diag} \left\{ \frac{\partial^2 b(X\boldsymbol{\beta} + \boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right\} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - k (\boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta} + k) \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\lambda} \boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1},$$

where

$$k = \frac{\phi(\boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta})}{\Phi(\boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\delta})},$$

is known as reversed hazard function in reliability theory.

Note that up to this stage, all parameters are considered known. After convergence, the value of the integral is computed to evaluate the log-likelihood.

Finally, the Laplace approximation for the log-likelihood is as follows:

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}) &\approx \frac{n}{2} \log(2\pi) - \frac{1}{2} \log \left| -H(\hat{\boldsymbol{\delta}}(\boldsymbol{\theta})) \right| + \mathbf{y}^T (X\boldsymbol{\beta} + \hat{\boldsymbol{\delta}}(\boldsymbol{\theta})) \\ &\quad - \mathbf{1}^T b(X\boldsymbol{\beta} + \hat{\boldsymbol{\delta}}(\boldsymbol{\theta})) + \mathbf{1}^T c(\mathbf{y}) \\ &\quad + \log(2) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2} \hat{\boldsymbol{\delta}}(\boldsymbol{\theta})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\delta}}(\boldsymbol{\theta}) \\ &\quad + \log \left( \Phi(\boldsymbol{\lambda}^T \boldsymbol{\omega}_{\boldsymbol{\theta}}^{-1} \hat{\boldsymbol{\delta}}(\boldsymbol{\theta})) \right), \end{aligned}$$

that can be numerically maximized over the model parameters. We use the notation  $\hat{\boldsymbol{\delta}}(\boldsymbol{\theta})$  to emphasize that  $\hat{\boldsymbol{\delta}}$  is a function of the model parameters  $\boldsymbol{\theta}$ . For the maximization step, we typically use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, implemented by the function `optim` in R, to estimate the model parameters  $(\boldsymbol{\beta}, \sigma^2, \varphi, \boldsymbol{\lambda})$ . The choice of initial values may be critical, then we follow the simple strategy proposed by Bonat and Ribeiro (2015).

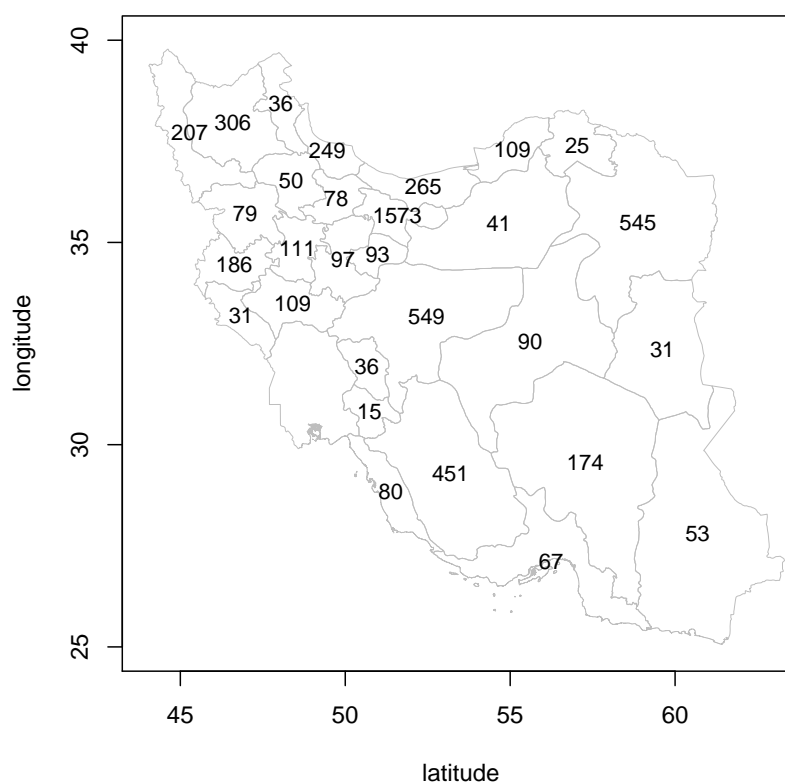
## 4 Data Analysis

In this section a real data set is considered to evaluate the performance of the proposed model. In order to compare the models we consider the  $-\log\text{Lik} = -\log L(\hat{\boldsymbol{\vartheta}})$  values and the Akaike information criterion (AIC), which are defined by  $-2\log\text{Lik} + 2\kappa$ , where  $\hat{\boldsymbol{\vartheta}}$  is the MLEs vector,  $\kappa$  is the number of estimated parameters. The best model corresponds to lower  $-\log\text{Lik}$  and AIC values.



## Breast Cancer Data

Cancer is the third cause of mortality after car accident and cardiovascular disease in Iran, so cancer is an important problem in public health in Iran. The case of interest was number of breast cancer patients as a count variable registered between the Apr 2007 to Mar 2008. Breast cancer is the leading cancer among women in Iran. We used the 2006 population size at each location as an offset. Data recorded on incident cases of cancer were obtained from ministry of health and medical education (in the non-infectious diseases management center) of Iran.



**Figure 2.** Map of the provinces of Iran with number of patients as labels.

The geographical units are 30 provinces with different size and shapes displayed in Figure 2. We display some collected data included province name in Table 1, where  $s_i$  for  $i = 1, \dots, 30$  identifies a geographical location. Following most authors, entities (e.g. counties, states, electoral ward) are

**Table 1.** Some collected data from different provinces and related features.

$i$	Province Name	$\mathbf{s}_i$	Offset	Number of Patients
1	Ardebil	(48.05, 38.45)	1209968	36
2	Bushehr	(51.38, 28.85)	866490	80
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
30	Zanjan	(48.36, 36.52)	942818	50

Table 2. Parameter estimates using Laplace method in three models for the Breast cancer data.

Parameter Estimates	GLM	SGLM	Skew-SGLM
$\hat{\beta}_0$	-9.34	-9.63	-9.59
$\hat{\sigma}^2$	-	0.17	0.17
$\hat{\varphi}$	-	0.79	0.73
$\hat{\lambda}$	-	-	-7.41
logLik	-435.25	-152.45	-151.82

referenced geographically by their centroids (or seats) with the spatial coordinates, which means that the actual spatial support (i.e. size and shape of the county or ward) is ignored in the analysis (Goovaerts, 2005). Then, we use spatial coordinates of the province geographic centroids. We arrive with missing data at  $i = 14$ , because it was not reported. The maximized log-likelihood value was  $-435.27$  for a fitted Poisson GLM using one unknown regression parameter  $\beta_0$  and without considering spatial effects.

Table 2 presents parameter estimates from Laplace method, along with the maximized log-likelihood values obtained by Laplace approximation. We assumed an isotropic exponential covariance function for the entries in the matrix  $\Sigma_{\theta}$ . This entails that  $C(\mathbf{s}_i - \mathbf{s}_j; \theta) = \sigma^2 \exp(-\|\mathbf{s}_i - \mathbf{s}_j\|/\varphi)$ , where  $\|\mathbf{s}_i - \mathbf{s}_j\|$  is the Euclidean norm,  $\varphi$  is the spatial correlation parameter and  $\sigma^2$  is the variance. Also, we assumed that the vector of skewness parameters is of the form  $\lambda = \lambda \mathbf{1}$ . The maximized log-likelihood values for a Poisson GLM, SGLM and Skew-SGLM models fitted for the breast cancer data are  $-435.25$ ,  $-152.45$  and  $-151.82$ , respectively. Based on these values, it is indicated the better fit of SGLM and Skew-SGLM models over the GLM for this data set.

## 5 Conclusion

It is shown that misspecification of spatial latent variables in an SGLM model (called GLM) affects estimation of the parameters. Since the skew Gaussian is more general than a Gaussian distribution, and usually it is not possible to check the latent variables distribution, it is proposed to use the Skew-SGLM models for the spatial count data. We investigated the performance of a simple algorithm to fit Skew-SGLM models based on the Laplace approximation. We illustrated based on maximized log-likelihood values. If there is not a great difference between the results, the Gaussian distribution appears sufficient. Possible topics for further investigation and extensions include analysing other datasets, performing simulations studies, improving the results by using another distribution for latent variables such as the CSG distribution.

## References

- Allard, D. and Naveau, P. (2007). A New Spatial Skew-normal Random Field Model. *Communications in Statistics - Theory and Methods*, **36**, 1821-1834.
- Arellano-Valle, R. and Azzalini, A. (2006). On the Unification of Families of Skew Gaussian Distributions. *Scandinavian Journal of Statistics*, **33**, 561-574.
- Azzalini, A. (1985). A Class of Distributions which Includes the Gaussian Ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- Azzalini, A. and Capitanio, A. (2014). *The Skew Gaussian and Related Families*. Cambridge University Press, New York.
- Azzalini, A. and Dalla-Valle, A. (1996). The Multivariate Skew Gaussian Distribution. *Biometrika*, **83**, 715-726.
- Baghishani, H. and Mohammadzadeh, M. (2011). A Data Cloning Algorithm for Computing Maximum Likelihood Estimates in Spatial Generalized Linear Mixed Models. *Computational Statistics and Data Analysis*, **55**, 1748-1759.
- Bonat, W.H. and Ribeiro, P.J. (2015). Practical Likelihood Analysis for Spatial Generalized Linear Mixed Models. *Environmetrics*, **27**, 83-89.
- Breslow, N.E. and Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9-25.
- Chan, K.S. and Ledolter, J. (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts. *Journal of the American Statistical Association*, **90**, 242-252.

- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley and Sons, Inc, New York.
- Christensen, O.F. (2004). Monte Carlo Maximum Likelihood in Model-based Geostatistics. *Journal of Computational and Graphical Statistics*, **13**, 702-718.
- Christensen, O.F. and Waagepetersen, R. (2002). Bayesian Prediction of Spatial Count Data using Generalized Linear Mixed Models. *Biometrics*, **58**, 280-286.
- Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998). Model-based Geostatistics (with Discussion). *Journal of the Royal Statistical Society: Series C*, **47**, 299-326.
- González-Farías, G., Domínguez-Molina, J. and Gupta, A. (2004). Additive Properties of Skew Normal Random Vectors. *Journal of Statistical Planning and Inference*, **126**, 521-534.
- Goovaerts, P. (2005). Geostatistical Analysis of Disease Data: Estimation of Cancer Mortality Risk from Empirical Frequencies Using Poisson Kriging. *International Journal of Health Geographics*, **4**, 31.
- Kim, H.M. and Mallick, B.K. (2004). A Bayesian Prediction Using the Skew Gaussian Distribution. *Journal of Statistical Planning and Inference*, **120**, 85-101.
- Lele, S.R., Nadeem, K. and Schmuland, B. (2010). Estimability and Kikelihood Inference for Generalized Linear Mixed Models Using Data Cloning. *Journal of the American Statistical Association*, **105**, 1617-1625.
- McCulloch, C.E. (1997). Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **92**, 162-170.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd ed., Chapman and Hall, London.
- Mohammadzadeh, M. and Hosseini, F. (2011). Maximum Likelihood Estimation for Spatial GLM Models. *Procedia Environmental Sciences*, **3**, 63-68.
- Hosseini, F., Eidsvik, J. and Mohammadzadeh, M. (2011). Approximate Bayesian Inference in Spatial GLMM with Skew Gaussian Latent Variables. *Computational Statistics and Data Analysis*, **55**, 1791-1806.
- Torabi, M. (2015). Likelihood Inference for Spatial Generalized Linear Mixed Models. *Communications in Statistics - Simulation and Computation*, **44**, 1692-1701.
- Zhang, H. (2002). On Estimation and Prediction for Spatial Generalized Linear Mixed Models. *Biometrics*, **58**, 129-136.
- Zhang, H. and El-Shaarawi, A. (2010). On Spatial Skew-Gaussian Processes and Applications. *Environmetrics*, **21**, 33-47.

**Seyyed Reza Hosseini Shojaei**

Department of Statistics,

University of Birjand,

Birjand, Iran.

email: *shojae87@birjand.ac.ir*

**Yadollah Waghei**

Department of Statistics,

University of Birjand,

Birjand, Iran.

email: *ywaghei@birjand.ac.ir*

**Mohsen Mohammadzadeh**

Department of Statistics,

Tarbiat Modares University,

Tehran, Iran.

email: *mohsen\_m@modares.ac.ir*

