Bayesian Inference of Spatially Correlated Binary Data Using Skew-Normal Latent Variables with Application in Tooth Caries Analysis

Solaiman Afroughi

Department of Biostatistics and Epidemiology, Social Determinants of Health Research Centre, Yasuj University of Medical Sciences, Yasuj, Iran
Email: safroughi@yahoo.com

Abstract

The analysis of spatially correlated binary data observed on lattices is an interesting topic that catches the attention of many scholars of different scientific fields like epidemiology, medicine, agriculture, biology, geology and geography. To overcome the encountered difficulties upon fitting the autologistic regression model to analyze such data via Bayesian and/or Markov chain Monte Carlo (MCMC) techniques, the Gaussian latent variable model has been enrolled in the methodology. Assuming a normal distribution for the latent random variable may not be realistic and wrong, normal assumptions might cause bias in parameter estimates and affect the accuracy of results and inferences. Thus, it entails more flexible prior distributions for the latent variable in the spatial models. A review of the recent literature in spatial statistics shows that there is an increasing tendency in presenting models that are involving skew distributions, especially skew-normal ones. In this study, a skew-normal latent variable modeling was developed in Bayesian analysis of the spatially correlated binary data that were acquired on uncorrelated lattices. The proposed methodology was applied in inspecting spatial dependency and related factors of tooth caries occurrences in a sample of students of Yasuj University of Medical Sciences, Yasuj, Iran. The results indicated that the skew-normal latent variable model had validity and it made a decent criterion that fitted caries data.

Keywords

Spatial Data, Latent Variable, Autologistic Model, Skew-Normal Distribution, Bayesian Inference, Tooth Caries

1. Introduction

The analysis of spatially correlated binary data observed on lattices is an interesting topic that catches the attention of many scholars of different scientific fields like epidemiology, medicine, agriculture, biology, geology and geography [1]-[11]. A review of the spatial statistical literature reveals that the autologistic regression, as a special case of the conditional autoregressive (CAR) models, is the main tool for analyzing the binary data collected from random fields [12]-[20]. Since the autologistic model normalizing factor does not have a closed form, it is tough to handle it [4] [21] [22]. So, following the traditional estimation methods, other inferential methods such as Bayesian paradigm were proposed to decline computational complications or to augment the efficiency of the estimates of model parameters [4] [12] [23]-[25]. To overcome the encountered difficulties upon fitting the autologistic regression model to analyze such data via Bayesian and/or Markov chain Monte Carlo (MCMC) techniques, the Gaussian latent variable method has been enrolled in the spatial generalized linear models [3] [8] [12] [13] [26]. Assuming a normal distribution for the latent random variable may not be realistic, so wrong, normal assumptions might cause bias in parameter estimates and affect the accuracy of results and inferences of the proposed methodology [1] [5] [6] [27]-[29]. Thus, it entails more flexible prior distributions for the latent variable in the spatial models [30] [31]. A review of the recent literature in spatial statistics shows that there is an increasing tendency in presenting models that are involving skew distributions, especially skew-normal ones [8] [27] [29] [32]-[36].

The univariate skew-normal distribution, a pioneering work started by Azzalini [37] [38], has many similar properties to normal distribution and include an extra parameter which regulates (represents) its skewness [35] [37]-[42]. The multivariate case and its marginal distributions as well as the extended conditional versions of the skew-normal distributions were developed by Azzalini and Dalla Valle [43], Azzalini and Capitanio [44] and Azzalini [45].

The main capability of this class of distributions in applications is its ability in capturing and simplicity modeling departures from symmetry, whilst retaining tractability and closeness [46] [47]. Following the development in spatial domain by Kim *et al*. [33] and Kim and Mallick [34], the authors such as Zhang and El-Shaaravi [29], Mohamadzadeh and Hosseini [31], Flecher *et al*. [27] and Allard and Soubeyrand [2], used the skew-normal distribution in analyzing spatial structure models. Moreover, Kim and Mallick [6] [48] and Karimi and Mohammadzadeh [5], implemented a Bayesian analysis in modeling skew-normal spatial observations. Although Hosseini *et al*. [3] incorporated the skew-normal latent variables in inference from a spatial generalized mixed model, no studies were found considering skew-normal latent variables in analysis of spatially correlated binary data modeled via autologistic regression [8] [49]-[51].

In this study, a skew-normal latent variable modeling in Bayesian analysis of the spatially correlated binary data acquired on uncorrelated lattices will be developed.

2. Statistical Models

2.1. Modeling Based on the Autologistic Regression

Let $y_i = \left(y_{i1}, \cdots, y_{ij}, \cdots, y_{in} \right)'$ be the binary (1/0) observations at n sites of the ith element of a set of N uncorrelated and uniform lattices [4] [12] [20] [52]. We suppose that $x_{ij} = \left(x_{ij1}, \cdots, x_{ijm} \right)'$ be the m-vector of covariates related to response variable y_{ij}. Thus, in lattice i, y_i is a random field and the observations $\left(y_{i1}, \cdots, y_{ij}, \cdots, y_{in} \right)'$ are spatially correlated binary data [15] [52], to the extent that neighboring observations affect a site observation. Thus the appropriate model is the autologistic regression which incorporates the effects of neighboring responses and covariates simultaneously and is a Markov random field model [20] defined as the conditional probability p_{ij}, where $y_{ij} = 1$ given all other values y_{ij}:

$$p_{ij}(\boldsymbol{\eta}) = p\left(y_{ij} \mid y_{ik}, j \neq k, k \in N_i(j) \right) = \frac{\exp\left(x'_{ij}\boldsymbol{\beta} + \gamma \sum_{k \in N_i} y_{ik} \right)}{1 + \exp\left(x'_{ij}\boldsymbol{\beta} + \gamma \sum_{k \in N_i} y_{ik} \right)} \tag{1}$$

where x_{ij} as defined before is the vector of covariates of y_{ij} with its first element 1; $\boldsymbol{\beta} = \left(\beta_1, \cdots, \beta_q \right)^t$ is the vector of coefficients of the covariates x_{ij}; and γ is the coefficient of spatial auto covariate $\sum_{k \in N_i} y_{ik}$, such that $N_i(j) = \left\{ k : y_{ik} \text{ is a neighbor of } y_{ij} \right\}$ denotes indices of the first-order neighborhood set of site j in lattice i, which for a rectangular lattice, is defined as adjacent two vertical and two horizontal sites. We assume that the

model (1) is a fixed effect and the spatial associations that are obeying an isotropic function are stationeries [52] [53]. Hence, effects of individual lattices and coefficients are the same at neighboring sites as well as in all directions [54].

Based on the Equation (1), the joint probability model of the $y_{ij}s$ (vector y_i) is

$$p(y_i|\boldsymbol{\beta},\gamma)=(1/\varphi(\boldsymbol{\beta},\gamma))\exp\left[\sum_{j=1}^{n}y_{ij}\left(x'_{ij}\boldsymbol{\beta}+\frac{1}{2}\gamma w_{ij}\right)\right] \tag{2}$$

where $w_{ij}=\sum_{k\in N(j)}y_{ik}$ and $\varphi(\boldsymbol{\beta},\gamma)$ is a normalizing factor obtained by summing overall possible realizations of y_i namely

$$\varphi(\boldsymbol{\beta},\gamma)=\sum_{\text{for all }y_{ij}}\exp\left[\sum_{j=1}^{n}y_{ij}\left(x'_{ij}\boldsymbol{\beta}+\frac{1}{2}\gamma w_{ij}\right)\right],$$

and for all observations $y=(y_1,\cdots,y_N)'$ is

$$L(\boldsymbol{\eta}|y)=p(y|\boldsymbol{\beta},\gamma)=\prod_{i=1}^{N}(1/\varphi(\boldsymbol{\beta},\gamma))\exp\left[\sum_{j=1}^{n}y_{ij}\left(x_{ij}\beta+\frac{1}{2}\gamma w_{ij}\right)\right] \tag{3}$$

In this stage the set of parameters is denoted by $=\{\boldsymbol{\beta},\gamma\}$.

The parameter estimation of the models (1) to (3), which are based on the autologistic model, using traditional [15] [23], Bayesian [24] [25] or Markov chain Monte Carlo methods [19], is time-consuming and might expose limitations and complications [21] [51]. Consequently, the authors such as Afroughi *et al.* [12] [13] with introducing Gaussian latent variables, made easier the computational implications in proposed models. The current study implements a posterior analysis with the help of skew-Gaussian latent variable modeling [6] [37] [38] [43]-[45] [55].

2.2. Using Skew-Normal Latent Variables Model

Let $z_i=(z_{i1},\cdots,z_{in})'$ be the vector of latent variables corresponding to y_i, so that for every observed binary variable y_{ij}, $p(y_{ij}=1)=p(z_{ij}>d_0)$ and $p(y_{ij}=0)=p(z_{ij}<d_0)$. We assume that based on the Azzalini and Dalla Valle [43], Azzalini and Capitanio [44], Liseo and Loperfido [46], Ashur and Abdel-Hameed [56], Gupta *et al.* [41], Liseo and Parisi [57] and Figueiredo and Gomes [58], the *n*-dimensional random vector z_i has a multivariate skew-normal distribution with *n*-dimentional location parameters vector $\boldsymbol{\mu}_i=x'_i\boldsymbol{\beta}+\gamma w_i$, $n\times n$ positive-definite variance-covariance matrix Σ_θ and skewness parameters vector a_n, written as $z_i\sim SN_n(x'_i\boldsymbol{\beta}+\gamma w_i,\Sigma_\theta,a_n)$, with probability density function

$$f(z_i|\beta,\gamma)=2\phi_n(x'_i\beta+\gamma w_i,\Sigma)\Phi(a'_n\omega^{-1}(z_i-(x'_i\boldsymbol{\beta}+\gamma w_i))) \tag{4}$$

where ϕ_n and Φ are *n*-dimensional normal density and standard normal cumulative distribution functions, respectively, $a_n=a1_n$ such that 1_n is a $n\times 1$ vector of 1 and $a\in R$, $x_i=(x'_{i1},x'_{i2},\cdots,x'_{in})'$, $w_i=(w_{i1},w_{i2},\cdots,w_{in})'$ and ω is a diagonal matrix such that $\Sigma_\theta=\omega^{-1}\Sigma_\theta\omega$, where Σ_θ is the correlation (positive definite) matrix of the vector z_i, the spatially correlated variables z_{ij}. If the element of Σ is denoted by σ_{ij} and $\sigma_{jj}=\sigma_j^2=1$, then $\omega=\text{diag}(\sigma_{11},\cdots,\sigma_{nn})^{1/2}$. The element of Σ_θ in lattice *i* is defined in the equation below, such that

$$\text{corr}(z_{ij},z_{ik})=k_\theta(d_{jk})=\theta^{d_{jk}} \tag{5}$$

where d_{jk} is the Euclidian distance between sites *k* and *j* in the lattice *i*, and θ is the correlation parameter which measures smoothness of the correlation function k_θ [6]. Thus, the final states of the model and its parameters are

$$f(z_i|\eta)=2\phi_n(x'_i\beta+\gamma w_i,\Sigma_\theta)\Phi(a1'_n(z_i-(x'_i\boldsymbol{\beta}+\gamma w_i)) \tag{6}$$

and $\boldsymbol{\eta}=\{\boldsymbol{\beta}',\gamma,\theta,a\}$, respectively.

2.3. Bayesian Inference

The latent variable z is augmented to the acquired data y, so the posterior function of parameters and latent variables are $f(\eta, z|y)$. Since, this function has a complicated form, sample generations could be done by full conditional functions $f(\eta|z, y)$ and $f(z|\eta, y)$, using Gibbs sampler technique [59]. To generate i.i.d. samples from $f(\eta|z, y)$, first sample generation must be down from $f(z|\eta, y)$. The analytic process for exploiting the above functions and/or related conditional distributions with Markov chain Monte Carlo methods is presented as follows.

a) Inlattice i, the distribution of the skew-normal (latents) vector z_i is

$$f(z_i|\eta, y_i) = 2\phi_n(z_i, \Sigma_\theta)\Phi(a_n^t z_i)\prod_{i=1}^{n}\{I(z_{ij} > 0)I(y_{ij} = 1) + I(z_{ij} < 0)I(y_{ij} = 0)\} \tag{7}$$

which is a multivariate truncated skew-normal distribution [60]. Based on the Gibbs sampling technique, the related full conditionals, which are univariate truncated skew-normal distributions, are given as below:

$$f(z_{ij}|z_{i-j}, \eta, Y_{ij}) = \phi(z_{ij} - \mu_{2.1}, 1 - \Sigma_{21}^t \Sigma_{22}^{-1} \Sigma_{21}) \times \Phi(a(z_{ij} - \mu_{2.1}) + \varphi\sqrt{1 + a^2(L_{2.1})})\Big/\Phi(\varphi)$$
$$\times\Big[I(z_{ij} > 0)I(y_{ij} = 1) + I(z_{ij} < 0)I(y_{ij} = 0)\Big], \tag{8}$$

where,

$$z_{i-j} = (z_{i1}, \cdots, z_{ij-1}, z_{ij+1} \cdots, z_{in})^t,$$

$$\mu_{2.1} = \Big[x_{ij}^t \beta + w_{ij}\gamma + \Sigma_{12}\Sigma_{22}^{-1}(z_{i-j} - x_{i-j}^t \beta_{-j})\Big],$$

$$\varphi = a \times a_z^t(z_{i-j} - x_{i-j}^t \beta_{-j}),$$

$$L_{2.1} = 1 - \Sigma_{12}^t \Sigma_{22}^{-1} \Sigma_{12}, \quad \alpha_z^t = a\Big(1_{(n-1)}^t + \Sigma_{22}^{-1}\Sigma_{21}I_{n-1}\Big)^t\Big/\sqrt{1 + a^2(1 - L_{2.1})},$$

$$x_{ij} = \Big[x_{ij1}, \cdots, x_{ijk}, \cdots, x_{ijq}\Big]',$$

$$x_i = \Big[x_{i1}, \cdots, x_{ij}, \cdots, x_{in}\Big]^t$$

$$x_{i-j} = \Big[x_{i1}^t, \cdots, x_{ij-1}^t, x_{ij+1}^t, \cdots, x_{in}^t\Big]'$$

$$\beta = \Big[\beta_1, \beta_2, \cdots, \beta_q\Big]',$$

$$\beta_{-j} = \Big[\beta_1, \beta_2, \cdots, \beta_{j-1}, \beta_{j+1}, \cdots, \beta_q\Big]',$$

and Σ_{12}, Σ_{22} and Σ_{21} are the submatrixes of the $n \times n$ partitioned correlation matrix Σ of the vector $(z_{ij}, z_{i-j})'$ such that $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

b) The posterior distribution $f(\eta|z, y)$ is equal to $f(\eta|z)$ or $f((\beta, \theta, \gamma, a)|z)$.

Again, based on the Gibbs sampling technique [61] and assumption of prior independence of parameters, the related full conditional functions are given as follow:

b-1):

$$f(\beta|z, \theta, a, \gamma) \propto f(z|\beta, \theta, a, \gamma)f(\beta|\theta, \gamma, \alpha)$$
$$\propto f(z|\beta, \theta, a, \gamma)f_0(\beta) \tag{9}$$

where $f\left(z|\boldsymbol{\beta},\theta,a,\gamma\right)=\prod_{i=1}^{N}?\left(z_{i}|\boldsymbol{\eta}\right)$ in which $z_{i}\sim SN_{n}\left(x_{i}'\boldsymbol{\beta}+\gamma w_{i},\Sigma_{\theta},a_{n}\right)$

Such that

$$f\left(z_{i}|\boldsymbol{\beta},\gamma,\theta,a\right)=2\phi_{n}\left(x_{i}'\boldsymbol{\beta}+\gamma w_{i},\Sigma_{\theta}\right)\Phi\left(a_{n}^{t}z_{i}\right)$$

or in other form

$$f\left(z_{i}|\boldsymbol{\beta},\gamma,\theta,a\right)\propto 2\phi_{n}\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right),\Sigma_{\theta}\right)\Phi\left(a_{n}^{t}\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right)$$

$$=2\left(2\pi\right)^{-n/2}\left|\Sigma_{\theta}\right|^{-1/2}\exp\left[\frac{-1}{2}\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)'\Sigma_{\theta}^{-1}\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right]\int_{-\infty}^{\alpha\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)}\frac{1}{\sqrt{2\pi}}e^{-t^{2}/2}\mathrm{d}t,$$

since

$$f\left(z|n\right)=\prod_{i=1}^{N}f\left(z_{i}|n\right)$$

$$\propto 2^{N}\left(2\pi\right)^{\frac{-nN}{2}}\left|\Sigma_{\theta}\right|^{-N/2}\exp\left[\frac{-1}{2}\sum_{i=1}^{N}\left[\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)'\Sigma_{\theta}^{-1}\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right]\right]\prod_{i=1}^{N}\int_{-\infty}^{\alpha\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)}\frac{1}{\sqrt{2\pi}}e^{-t^{2}/2}\mathrm{d}t,$$

Then

$$f\left(\boldsymbol{\beta}|z,\theta,\alpha,\gamma\right)\propto f\left(z|\boldsymbol{\eta}\right)f_{0}\left(\boldsymbol{\beta}\right)$$

where $f_{0}\left(\boldsymbol{\beta}\right)$ is the prior distribution of the vector $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}\sim\phi_{q}\left(\boldsymbol{\beta}_{0},V_{0}\right)$.

Therefore,

$$f\left(\boldsymbol{\beta}|z,\theta,\gamma,\alpha\right)$$

$$\propto\left|\Sigma_{\theta}\right|^{-N/2}\exp\left[\frac{-1}{2}\sum_{i=1}^{N}\left[\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)'\Sigma_{\theta}^{-1}\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right]\right]\prod_{i=1}^{N}\Phi\left(\alpha\left(z_{i}-\left(x_{i}'\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right)f_{0}\left(\boldsymbol{\beta}\right)\quad(10)$$

$$\propto\phi_{q}\left(\boldsymbol{\beta}-\boldsymbol{u}_{\beta},\boldsymbol{v}_{\beta}\right)\prod_{i=1}^{N}\Phi\left(\alpha\left(z_{i}-\left(x_{i}^{t}\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right),$$

where

$$\boldsymbol{u}_{\beta}=\sum_{i=1}^{N}v_{i}\left(x_{i}^{t}\Sigma_{\theta}^{-1}x_{i}\hat{\boldsymbol{\beta}}_{i}+V_{0}^{-1}\boldsymbol{\beta}_{0}\right),\quad v_{i}=\left(x_{i}^{t}\Sigma_{\theta}^{-1}x_{i}+V_{0}^{-1}\right)^{-1}$$

$\boldsymbol{v}_{\beta}=\sum_{i=1}^{N}v_{i}$ and $\hat{\boldsymbol{\beta}}_{i}=\left(x_{i}^{t}\Sigma_{\theta}^{-1}x_{i}\right)^{-1}x_{i}^{t}\Sigma_{\theta}^{-1}z_{i}$

b-2):

$$f\left(\gamma|z,\theta,\boldsymbol{\beta},\alpha\right)\propto f\left(z|\gamma,\boldsymbol{\beta},\theta,\alpha\right)f\left(\gamma|\boldsymbol{\beta},\theta,\alpha\right)=f\left(z|\gamma,\boldsymbol{\beta},\theta,\alpha\right)g_{0}\left(\gamma\right)\quad(11).$$

where the above equations $g_{0}\left(\gamma\right)$ is the prior distribution of γ and assumed Gamma distribution $G\left(a_{0},b_{0}\right)$. So

$$f\left(\gamma|z,\boldsymbol{\beta},\theta,\alpha\right)\propto f\left(z|\boldsymbol{\eta}\right)g_{0}\left(\gamma\right)=\prod_{i=1}^{N}f\left(z_{i}|n\right)G\left(a_{0},b_{0}\right)$$

$$=2^{N}\left(2\pi\right)^{-Nn/2}\left|\Sigma_{\theta}\right|^{-N/2}\exp\left(\frac{-1}{2}\sum_{i=1}^{N}\left(z_{i}-\left(x_{i}\beta+\gamma w_{i}\right)\right)^{t}\Sigma_{\theta}^{-1}\left(z_{i}-\left(x_{i}\beta+w_{i}\gamma\right)\right)\prod_{i=1}^{N}\Phi\left(\alpha\left(z_{i}-\left(x_{i}^{t}\boldsymbol{\beta}+w_{i}\gamma\right)\right)\right)\right)$$

$$\times G\left(a_{0},b_{0}\right)\quad(12)$$

$$=2^{N}\left(2\pi\right)^{-Nn/2}\left|\Sigma_{\theta}\right|^{-N/2}\exp\left(\frac{-1}{2}\sum_{i=1}^{N}\left(z_{i}-\left(x_{i}\beta+\gamma w_{i}\right)\right)^{t}\Sigma_{\theta}^{-1}\left(z_{i}-\left(x_{i}\beta+w_{i}\gamma\right)\right)\prod_{i=1}^{N}\int_{-\infty}^{\alpha^{t}\left(z_{i}-\left(x_{i}^{t}\boldsymbol{\beta}+w_{i}\gamma\right)\right)}\frac{1}{\sqrt{2\pi}}e^{-t^{2}/2}\mathrm{d}t\right.$$

$$\times\frac{b_{0}^{a_{0}}}{\Gamma\left(a_{0}\right)}\gamma^{a_{0}-1}\exp\left(-b_{0}\gamma\right),$$

As it is evident, $f\left(\gamma|z,\boldsymbol{\beta},\theta,\alpha\right)$ does not have a specified form and we can sample again from this distribution through Metropolis-Hastings [59] steps with proposed density $q\left(.|\gamma\right)=N\left(\hat{\gamma},\widehat{\sigma_{\gamma}^{2}}\right)$ where $\hat{\gamma}$ and $\widehat{\sigma_{\gamma}^{2}}$

(**Appendix A**) are derived from $\partial \ln f\left[\left(z|\beta,\gamma,\theta,a\right)/\partial\gamma\right]$ and $-\partial^2\ln f\left[\left(z|\beta,\gamma,\theta,a\right)/\partial^2\gamma\right]$, respectively.

b-3):

$$f\left(\theta|z,\boldsymbol{\beta},\alpha,\gamma\right) \propto f\left(z|\boldsymbol{\beta},\gamma,\theta,\alpha\right)f\left(\theta|\beta,\alpha,\gamma\right) \propto f\left(z|\beta,\gamma,\theta,\alpha\right)f_0\left(\theta\right)$$

$$= 2^N\left(2\pi\right)^{-Nn/2}\left|\Sigma_\theta\right|^{-N/2}\exp(\frac{-1}{2}\sum_{i=1}^{N}\left(z_i-\left(x_i\beta+\gamma w_i\right)\right)'\Sigma_\theta^{-1}\left(z_i-\left(x_i\beta+w_i\gamma\right)\right)\prod_{i=1}^{N}\int_{-\infty}^{\alpha\ln(z_i-(x_i\boldsymbol{\beta}+w_i\gamma))}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt \quad (13)$$

$$\times f_0\left(\theta\right),$$

where f_0 is the prior distribution of θ and assumed uniform in $(0, 1)$. In order to generate a sample from $f\left(\theta|z,\gamma,\boldsymbol{\beta},\alpha\right)$, which has not a closed form, we use Metropolis-Hastings algorithm [59] as follows. The correlation parameter θ is changed to $\xi=\text{logit}\left(\theta\right)$ and ξ' is generated from the proposed density

$$q_\xi\left(.|\xi\right)= N\left(\xi=\xi_0,\Psi^2=0.81\right).$$

The value

$$\theta'=\exp\left(\xi'\right)/\left(1+\exp\left(\xi'\right)\right)$$

is accepted with regard to $q_\xi\left(\xi|\xi'\right)/q_\xi\left(\xi'|\xi\right)=\left[\theta'\left(1-\theta\right)/\theta\left(1-\theta'\right)\right]$ and probability of

$$\alpha=\min\left\{\left(\frac{f\left(\theta'|\beta,\gamma,a,z\right)}{f\left(\theta|\beta,\gamma,a,z\right)}\right)\times\left(\frac{\theta'\left(1-\theta\right)}{\theta\left(1-\theta'\right)}\right),1\right\}.$$

b-4):

$$f\left(a|z,\beta,\theta,\gamma\right) \propto f\left(z|\beta,\theta,\gamma,a\right)g\left(a|\beta,\theta,\gamma\right)$$

$$= f\left(z|\beta,\theta,\gamma,a\right)f_0(a)$$

$$\propto f_0\left(a\right)\exp\left[\frac{-1}{2}\sum_{i=1}^{N}\left[\left(z_i-\left(\boldsymbol{x}_i'\boldsymbol{\beta}+\boldsymbol{w}_i\gamma\right)\right)'\Sigma_\theta^{-1}\left(z_i-\left(\boldsymbol{x}_i'\boldsymbol{\beta}+\boldsymbol{w}_i\gamma\right)\right)\right]\right]\prod_{i=1}^{N}\Phi\left(a\left(z_i-\left(\boldsymbol{x}_i^t\boldsymbol{\beta}+\boldsymbol{w}_i\gamma\right)\right)\right) \quad (14)$$

$$\propto f_0\left(a\right)\prod_{i=1}^{N}\Phi\left(a\left(z_i-\left(\boldsymbol{x}_i^t\boldsymbol{\beta}+\boldsymbol{w}_i\gamma\right)\right)\right),$$

where f_0 is the prior distribution of a and since we assumed priori $\alpha \sim N\left(a_0=2,\sigma^2=1\right)$, then

$$f\left(a|z,\beta,\theta,\gamma\right) \propto \exp\left[-\frac{1}{2}\left(a-a_0\right)^2\right]\prod_{i=1}^{N}\Phi\left(a\left(z_i-\left(\boldsymbol{x}_i^t\boldsymbol{\beta}+\boldsymbol{w}_i\gamma\right)\right)\right) \quad (15)$$

this distribution does not have a closed form and we could sample from it again through Metropolis-Hasting technique with proposed density $q\left(.|a\right)=N\left(0,4\right)$.

3. Application

In this section, the proposed methods are applied to the actual data-set that were collected in a study designed to explore spatial dependency and related factors in tooth caries experiences in a random sample of size 132 taken from students of Yasuj University of Medical Sciences, Iran, in 2012. A team of oral health hygienists working in a dentistry centre gathered data as follows. First, information regarding demographic, social and mouth healthcare were obtained through a questionnaire. Then, each tooth along with its periodontal in the complete (32) teeth set of every student were assessed and the presence of caries in each surface of a tooth and gingivitis in its periodontal was diagnosed based on the clinical methods [62] like light, mirror and sound. Additionally, the teeth of each subject were stratified as sound, carious, missing and/or filled due to caries, and a chart was prepared denoting the caries status and site position of each tooth in his/her mouth. The above steps were approved by adentist and a professor of pediatric dentistry.

The autologistic model (1) is fitted to data such that y_{ij} is the binary response variable of tooth j in subject i where $y_{ij}=1$ indicates that a tooth is decayed, missed or filled due to caries, and $y_{ij}=0$ if it is sound [49].

Further x_{ij} denotes that the covariates in subject and tooth levels including age (in years), gender (1 = male, 0 = female), gingivitis (gingv) (1 = present, 0 = otherwise), father occupation (ocf) (1 = official worker, 0 = otherwise), father education (edf) (1 = university educalated, 0 = otherwise), teeth were check up at least once in every 6 months by a dentist (vizd6) (1 = yes, 0 = no) at least brushing the teeth once in a day (tbr) (1 = yes, 0 = no), and every day tooth flossing (tfl) (1 = yes, 0 = no). Additionally, the sum of three responses in the first-order neighborhood teeth (**Figure 1**) of a y_{ij}, constitutes the spatial autocovriate (spacov) $\left(w_{ij}\right)$. The coefficients of explanatory covariates and spatial autocovariate are denoted by $\boldsymbol{\beta}=\left(\beta_0,\beta_1,\cdots,\beta_8\right)'$ and γ respectively. As was explained in previous work [54], the teeth caries statuses $\left(y_{ij}s\right)$ in a mouth are spatially correlated binary data clustered in uncorrelated (subjects) lattices. Accordingly, the posited extended autologistic model (1) based on the logit link is as follows.

$$\text{logit}\left(p_{ij}\right)=\beta_0+\beta_1\text{ginv}+\beta_2\text{gend}+\beta_3\text{age}+\beta_4\text{ocf}+\beta_5\text{edf}+\beta_6\text{vzd6}+\beta_7\text{tbr}+\beta_8\text{tfls}+\gamma\text{spacov} \qquad (16)$$

The Bayesian estimations of parameters using Gibbs sampling and MCMC technique were obtained through programming coded in freeware R [63] version 3.1 as follows. First based on the adopted prior distributions $N_9\left(\boldsymbol{\mu}_0=\left(0,0,0,0,0,0,0,0,0\right),v_0=2I_9\right)$, $\text{Gamma}\left(1,2\right)$, $U\left(0,1\right)$ and $N\left(0,10\right)$ for the parameters $\boldsymbol{\beta}$, γ, θ, a, respectively, a sample of $z_{ij}s$ were generated [51] from skew-normal distribution (8). Then, implementing these $z_{ij}s$ in each of 50 iterations a sample of size 10000 of parameters $\boldsymbol{\eta}=\left\{\boldsymbol{\beta},\gamma,\theta,a\right\}$ were simulated from full conditional distributions (10), (12), (13) and (15), and after discarding first 4000 samples, the rest were applied to estimate the parameters.

The estimates of the parameters such as mean, standard errors (SE) as well as the 95% credible intervals are summarized in **Table l**. As is shown from this table, the coefficients of constant value, spatial auto covariate, the covariates gender, age, father occupation and father education and the skewness parameter are significantly different from zero. These findings demonstrated that caries statuses in neighboring teeth had influenced in caries occurrence in a tooth, female and younger students were more susceptible to tooth decaying and students whose fathers were official workers and/or university educated had lower tooth caries experiences. Further, as the contents of this table indicate, the gingivitis in a student notably arises caries occurrence in his (or her) teeth. Although the tooth brushing prevents the dental caries, yet, it doesn't have a significant impact. Furthermore, the estimate of the skewness parameter is different from (below) zero in a considerable case which indicates that the latent variable is highly skewed to the right side.
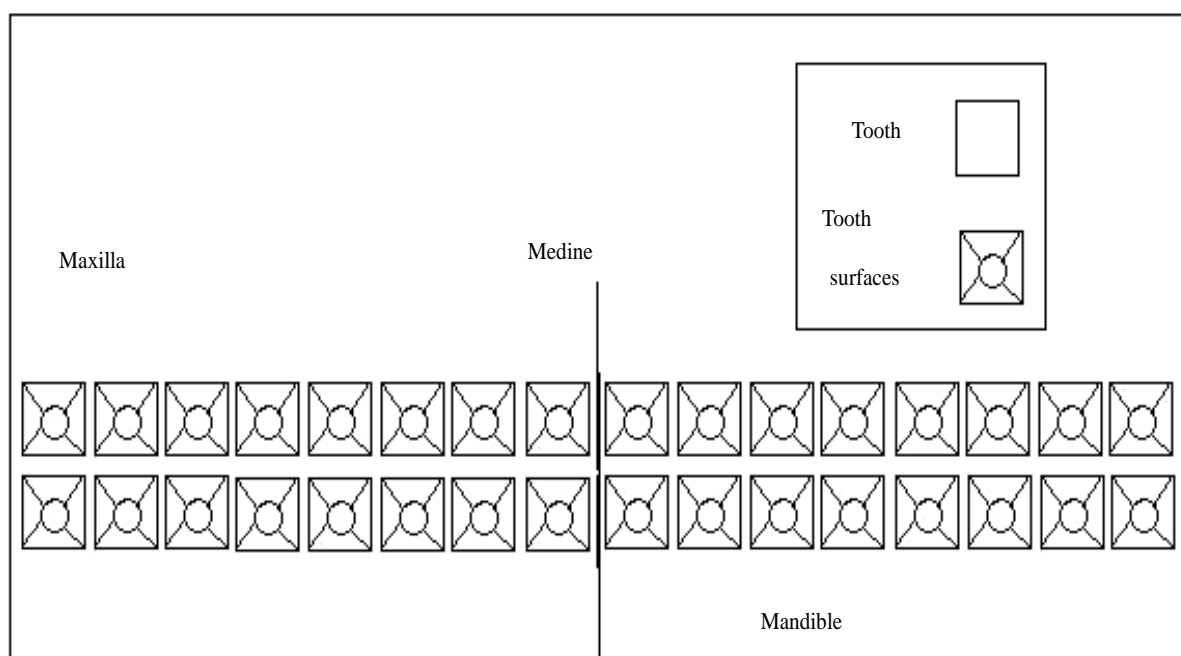

Figure 1. Spatial locations of complete (32) teeth in lattice system.

Table 1. Results of parameter estimations of the skew-normal latent variable model.

Covariate	Parameter	Estimate	Standard error	95% Cridble interval
Constant	B0	−0.1037	0.044	(−0.1477, −0.097)
Ginv	B1	0.0713	0.027	(0.0173, 0.1253)
Gend	B2	−0.1544	0.057	(−0.2684, −0.0404)
Age	B3	−3.1108	0.65	(−4.4108, −2.8108)
Ocf	B4	−0.0828	0.028	(−0.1388, −0.0268)
Edf	B5	−0.0550	0.011	(−0.077, −0.033)
Vzd6	B6	0.0715	0.038	(−0.0048, 0.1475)
Tbr	B7	−0.1446	0.056	(−0.2566, 0.0326)
Tfls	B8	0.0276	0.03	(−0.324, 0.0876)
Spacov	γ	0.23	0.032	(0.166, 0.294)
−	θ (correlation)	0.58	0.051	(0.478, 0.682)
−	a (skewness)	−2.75	0.027	(−2.794, −2.696)

4. Assessing the Model

To check the accuracy and validity of the presented skew-normal latent variable modeling, implementing a normal-latent variable methodology in Bayesian analysis of spatially correlated binary data was considered. The parameters estimates of this model using MCMC sampling were obtained and shown in **Table 2**. The findings in this table indicates that the estimates of parameters in the two methodology are often near to each other or in the same direction, while standard errors of the estimates in the methodology using skewed normal latent variable is lower than other case. Furthermore the Bayesian information criterion (BIC) [64] based the $\mathrm{BIC} = -2\ln\left(L(\hat{\eta})\right) + q\log m$, where L, q, m and $\hat{\eta}$ are likelihood function, number of parameter sample size and the estimated parameters, respectively, were computed for two models. For skew-normal latent model BIC = 1162036.2 and for normal latent variable model BIC = 1225318.8 were obtained. As is seen, the model using skew-normal latent variable in analysis of spatially correlated binary data is the better one.

5. Discussion

In this paper, a skew-normal latent variable methodology has been developed in Bayesian analysis of the spatially correlated binary data using autologistic regression model. Parameter estimation for the autologistic model is an extremely difficult job since its likelihood function has a normalizing factor which doesn't have a closed form [22] [23]. On the other hand, the traditional pseudo-likelihood estimation method, due to spatial dependency in data, especially when the lattice size is small, is inefficient [54]. Furthermore, the likelihood-based Markov chain Monte Carlo and Bayesian approaches often encounter with complicated and rigorous computations. Recently, authors such as Hughes *et al*. [4] have presented a method for easier and faster implementing the autologistic model. But with introduced excess parameters in their model, the inferential complications have not been reduced and their presented method is appropriate only for large size lattices. Also, Hossaini *et al*. [3] have researched the using of skew-normal latent variable approach in analyzing spatial data. However, they haven't introduced the autologistic model and a covariate capturing the influences of the neighboring sites of the response variable. Moreover, in the mentioned studies only data from one lattice were investigated.

The application of the presented methodology in tooth caries analysis demonstrates that the model is well fitted and the coefficients of spatial autocovariate which involves the sum of caries statuses in neighboring teeth of a tooth, and the gingivitis around it, and the education and occupational statuses of the participant's father, are notably different from zero. In addition, as it is evident from the result, the presented model in comparison with the normal latent variable method has a better validity and fitting indicators. So the proposed methodology is a novel way for bypassing and overcoming the intensive computational burden and complications resulting from

Table 2. Results of parameter estimations of the normal latent variable model.

Covariate	Parameter	estimate	Standard error	95% Credible interval
Constant	B0	−0.786	0.36	(−1.106, −0.466)
Ginv	B1	0.151	0.068	(0.15, 0.287)
Gend	B2	0.0418	0.068	(−.0942, 0.1778)
Age	B3	0.018	0.016	(−0.014, 0.050)
Ocf	B4	−.0469	0.074	(−0.1949, 0.1011)
Edf	B5	−0.0255	0.072	(−0.1695, 0.1185)
Vzd6	B6	0.022	0.037	(−0.052, 0.096)
Tbr	B7	−0.004	0.093	(−0.19, 0.182)
Tfls	B8	0.0012	0.027	(−0.0528, 0.0552)
Spacov	γ	1.6	0.051	(1.57, 1.71)
–	θ (correlation)	0.54	0.046	(0.4, 0.48)

normalizing constant of the autologistic model in analyzing spatial binary data on multiple lattices.

In conclusion, the proposed methodology based on the augmenting skew-normal latent variables in analyzing spatially correlated binary data is an adequate and appropriate model. A carious tooth in nearest neighbors of a tooth is a cause of its decay. The higher education and the income of the head of the family are the lower caries occurrences of the individual. These can be considered in oral health care and tooth caries treatment programs in the surveyed sample [65].

Acknowledgements

References

[1] Allard, D. and Naveau, P. (2007) A New Spatial Skew-Normal Random Field Model. *Communications in Statistics—Theory and Methods*, **36**, 1821-1834. http://dx.doi.org/10.1080/03610920601126290

[2] Allard, D. and Soubeyrand, S. (2012) Skew-Normality for Climatic Data and Dispersal Models for Plant Epidemiology: When Application Fields Drive Spatial Statistics. *Spatial Statistics*, **1**, 50-64. http://dx.doi.org/10.1016/j.spasta.2012.03.001

[3] Hosseini, F., Eidsvik, J. and Mohammadzadeh, M. (2011) Approximate Bayesian Inference in Spatial GLMM with Skew Normal Latent Variables. *Computational Statistics and Data Analysis*, **55**, 1791-1806. http://dx.doi.org/10.1016/j.csda.2010.11.011

[4] Hughes, J., Haran, M. and Caragea, P.C. (2011) Autologistic Models for Binary Data on a Lattice. *Environmetrics*, **22**, 857-871. http://dx.doi.org/10.1002/env.1102

[5] Karimi, O. and Mohammadzadeh, M. (2012) Bayesian Spatial Regression Models with Closed Skew Normal Correlated Errors and Missing Observations. *Statistical Papers*, **53**, 205-218. http://dx.doi.org/10.1007/s00362-010-0329-2

[6] Kim, H.M. and Mallick, B.K. (2004) A Bayesian Prediction Using the Skew Gaussian Distribution. *Journal of Statitical Planning and Inference*, **120**, 85-101. http://dx.doi.org/10.1016/S0378-3758(02)00501-3

[7] Khormi, H.M. and Kumar, L. (2012) The Importance of Appropriate Temporal and Spatial Scales for Dengue Fever Control and Management. *Science of Total Environment*, **430**, 144-149. http://dx.doi.org/10.1016/j.scitotenv.2012.05.001

[8] March, D., Alós, J., Cabanellas-Reboredo, M., Infantes, E., Jordi, A. and Palmer, M. (2013) A Bayesian Spatial Approach for Predicting Seagrass Occurrence. *Estuarine*, *Coastal and Shelf Science*, **131**, 206-212.

http://dx.doi.org/10.1016/j.ecss.2013.08.009

[9]   BO, Y.-C. Song, C., Wang, J.-F. and Li, X.-W. (2014) Using an Autologistic Regression Model to Identify Spatial Risk Factors and Spatial Risk Patterns of Hand Foot and Mouth Disease (HFMD) in Mainland China. *BMC Public Health*, **14**, 358.

[10] Thanapongtram, W., Linard, C., Pamaranon, N., Kawkalong, S., Noimoh, T., Chanachai, K., Parakgamawongsa, T. and Gilbert, M. (2014) Spatial Edidemiology of Porcine Reproductive and Respiratory Syndrome in Thailand. *BMC Veterinary Research*, **10**, 174. http://dx.doi.org/10.1186/s12917-014-0174-y

[11] Mattsson, B.J., Zipkin, E.F., Gardner, B., Blank, P.J., Sauer, J.R. and Role, J.A. (2013) Explaining Local-Scale Species Distributions: Relative Contributions of Spatial Autocorrelation and Landscape Heterogeneity for an Avian Assemblage. *PLoS ONE*, **8**, e55097. http://dx.doi.org/10.1371/journal.pone.0055097

[12] Afroughi, S., Faghihzadeh, S., Jafari Khaledi, M., Ghandehari Motlagh, M. and Hajizadeh, E. (2011) Analysis of Clustered Spatially Correlated Binary Data Using Autologistic Model and Bayesian Method with an Application to Dental Caries of 3 - 5-Year-Old Children. *Journal of Applied Statistics*, **38**, 2763-2774. http://dx.doi.org/10.1080/02664763.2011.570315

[13] Afroughi, S., Ghandehari Motlagh, M., Faghihzadeh, S. and Jafari Khaledi, M. (2013) A Model for Analyzing Spatially Correlated Binary Data Clustered in Uncorrelated Lattices. *Statistical Methodology*, **14**, 1-14. http://dx.doi.org/10.1016/j.stamet.2013.01.004

[14] Besag, J.E. (1972) Nearest-Neighbor Systems and the Auto-Logistic Model for Binary Data. *Journal of the Royal Statistical Society. Series B*, **34**, 75-83.

[15] Besag, J.E. (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion). *Journal of the Royal Statistical Society. Series B*, **36**, 192-236.

[16] Genton, M.G., He, L. and Liu, X. (2001) Moments of Skew Normal Random Vectors and Their Quadratic Forms. *Statistics & Probability Letters*, **51**, 319-325. http://dx.doi.org/10.1016/S0167-7152(00)00164-4

[17] Gumpertz, M.L., Graham, J.M. and Ristaino J.B. (1997) Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence. *Journal of Agricultural*, *Biological and Environmental Statistics*, **2**, 131-156. http://dx.doi.org/10.2307/1400400

[18] He, F.L., Zhou, J. and Zhu, H.T. (2003) Autologistic Regression Model for the Distribution of Vegetation. *Journal of Agricultural*, *Biological and Environmental Statistics*, **8**, 205-222. http://dx.doi.org/10.1198/1085711031508

[19] Huffer, F.W. and Wu, H.L. (1998) Markov Chain Monte Carlo for Autologistic Regression Models with Application to the Distribution of Plant Species. *Biometrics*, **54**, 509-524. http://dx.doi.org/10.2307/3109759

[20] Kaiser, M.S. and Caragea, P.C. (2009) Exploring Dependence with Data on Spatial Lattices. *Biometrics*, **65**, 857-865. http://dx.doi.org/10.1111/j.1541-0420.2008.01118.x

[21] Sherman, M., Apanasovich, T.V. and Carroll, R.J. (2006) On Estimation in Binary Autologistic Spatial Models. *Journal of Statistical Computation and Simulation*, **76**, 167-179. http://dx.doi.org/10.1080/00949650412331320873

[22] Caragea, P.C. and Kaiser, M.S. (2009) Autologistic Models with Interpretable Parameters. *Journal of Agricultural*, *Biological*, *and Environmental Statistics*, **14**, 281-300. http://dx.doi.org/10.1198/jabes.2009.07032

[23] Besag, J.E. (1975) Statistical Analysis of Non-Lattice Data. *The Statistician*, **24**, 179-195. http://dx.doi.org/10.2307/2987782

[24] Besag, J.E. (1993) Towards Bayesian Image Analysis. *Journal of Applied Statistics*, **20**, 107-119. http://dx.doi.org/10.1080/02664769300000061

[25] Besag, J.E., York, J. and Molli, A. (1991) Bayesian Image Restoration with Two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1-20. http://dx.doi.org/10.1007/BF00116466

[26] De Oliveira, V. (2000) Bayesian Prediction of Clipped Gaussian Random Fields. *Computational Statistics & Data Analysis*, **34**, 299-314. http://dx.doi.org/10.1016/S0167-9473(99)00103-6

[27] Flecher, C., Allard, D. and Naveau, P. (2010) Truncated Skew-Normal Distribution: Moments, Estimation by Weighted Moments and Application to Climatic Data. *METRON*, **68**, 331-345. http://dx.doi.org/10.1007/BF03263543

[28] Jara, A., Quintana, F. and Martin, E.S. (2008) Linear Mixed Models with Skew-Elliptical Distributions: A Bayesian Approach. *Computational Statistics and Data Analysis*, **52**, 5033-5045. http://dx.doi.org/10.1016/j.csda.2008.04.027

[29] Zhang, H. and El-Shaarawi, A. (2010) On Spatial Skew-Gaussian Processes and Applications. *Environmetrics*, **21**, 33-47.

[30] Arellano-Valle, R.B., Ozan, S., Bolfarine, H. and Lachos, V.H. (2005) Skew Normal Measurement Error Models. *Journal of Multivariate Analysis*, **96**, 265-281. http://dx.doi.org/10.1016/j.jmva.2004.11.002

[31] Mohammadzadeh, M. and Hosseini, F. (2011) Maximum Likelihood Estimation for Spatial GLM Models. *Procedia*

Environmental Sciences, **3**, 63-68. http://dx.doi.org/10.1016/j.proenv.2011.02.012

[32] Arellano-Valle, R.B., Bolfarine, H. and Lochos, V.H. (2005) Skew-Normal Linear Mixed Models. *Journal of Data Science*, **3**, 415-438.

[33] Kim, H.M., Ha, E. and Mallick, B.K. (2004) Spatial Prediction of Rainfall Using Skew-Normal Processes. In: Genton, M.G., Ed., *Skew-Elliptical Distributions and Their Applications*: *A Journey beyond Normality*, Chapter 16, Chapman & Hall/CRC, London, 279-289. http://dx.doi.org/10.1201/9780203492000.ch16

[34] Kim, H.M. and Mallick, B.K. (2002) Analyzing Spatial Data Using Skew-Gaussian Processes. In: Lawson, A. and Deninson, D., Eds., *Spatial Cluster Modelling*, Chapman & Hall/CRC, London, 164. http://dx.doi.org/10.1201/9781420035414.ch9

[35] Lin, T.I. (2009) Maximum Likelihood Estimation for Multivariate Skew-Normal Mixture Models. *Journal of Multivariate Analysis*, **100**, 257-265. http://dx.doi.org/10.1016/j.jmva.2008.04.010

[36] Asili, S., Rezaei, S. and Najjar, L. (2014) Using Skew-Logistic Probability Density Function as a Model for Age-Specific Fertility Rate Pattern. *BioMed Research International*, **2014**, Article ID: 790294. http://dx.doi.org/10.1155/2014/790294

[37] Azzalini, A. (1985) A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

[38] Azzalini, A. (1986) Further Results on a Class of Distributions Which Includes the Normal Ones. *Statistica*, **46**, 199-208.

[39] Flecher, C., Naveau, P. and Allard, D. (2009) Estimating the Closed Skew-Normal Distributions Parameters Using Weighted Moments. *Statistics & Probability Letters*, **79**, 1977-1984. http://dx.doi.org/10.1016/j.spl.2009.06.004

[40] Gupta, A.K. and Chen, J.T. (2004) A Class of Multivariate Skew-Normal Models. *Annals of the Institute of Statistical Mathematics*, **56**, 305-315. http://dx.doi.org/10.1007/BF02530547

[41] Gupta, A.K., Nguyen, T.T. and Sanqui, J.A.T. (2004) Characterization of the Skew-Normal Distribution. *Annals of the Institute of Statistical Mathematics*, **56**, 351-360. http://dx.doi.org/10.1007/BF02530549

[42] Lin, T.I., Lee, J.C. and Yen, S.Y. (2007) Finite Mixture Modeling Using the Skew-Normal Distribution. *Statistica Sinica*, **17**, 909-927.

[43] Azzalini, A. and Dalla Valle, A. (1996) The Multivariate Skew-Normal Distribution. *Biometrika*, **83**, 715-726. http://dx.doi.org/10.1093/biomet/83.4.715

[44] Azzalini, A. and Capitanio, A. (1999) Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society*: *Series B*, **61**, 579-602. http://dx.doi.org/10.1111/1467-9868.00194

[45] Azzalini, A. (2005) The Skew-Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, **32**, 159-188. http://dx.doi.org/10.1111/j.1467-9469.2005.00426.x

[46] Liseo, B. and Loperfido, N. (2003) A Bayesian Interpretation of the Multivariate Skew-Normal Distribution. *Statistics & Probability Letters*, **61**, 395-401. http://dx.doi.org/10.1016/S0167-7152(02)00398-X

[47] Minozzo, M. and Ferracuti, L. (2012) On the Existence of Some Skew-Normal Stationary Processes. *Chilean Journal of Statistics*, **3**, 157-170.

[48] Kim, H.M. and Mallick, B.K. (2005) A Bayesian Prediction Using the Elliptical and the Skew-Gaussian Processes. Technical Report. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.5432&rep=rep1&type=ps

[49] Bandyopadhyay, D., Lochos, V.H., Abanto-Valle, C.A. and Ghosh, P. (2010) Linear Mixed Models for Skew-Normal/ Independent Bivariate Responses with an Application to Periodontal Disease. *Statistics in Medicine*, **29**, 2643-2655. http://dx.doi.org/10.1002/sim.4031

[50] Lu, J., Gong, D., Shen, Y., Liu, M. and Chen, D. (2013) An Inversed Bayesian Modeling Approach for Estimating Nitrogen Export Coefficients and Uncertainty Assessment in an Agricultural Watershed in Eastern China. *Agricultural Water Management*, **116**, 79-88. http://dx.doi.org/10.1016/j.agwat.2012.10.015

[51] Wang, Z. and Zheng, Y. (2013) Analysis of Binary Data via a Centered Spatial-Temporal Autologistic Regression Model. *Environmental and Ecological Statistics*, **20**, 37-57. http://dx.doi.org/10.1007/s10651-012-0206-3

[52] Cressie, N. (1993) Statistics for Spatial Data. Revised Edition, Wiley, New York.

[53] Haining, R. (2003) Spatial Data Analysis: Theory and Practice. Cambridge University Press, London. http://dx.doi.org/10.1017/CBO9780511754944

[54] Afroughi, S., Faghihzadeh, S., Jafari Khaledi, M. and Ghandehari Motlagh, M. (2010) Dental Caries Analysis in 3 - 5 Years Old Children: A Spatial Modeling. *Archives of Oral Biology*, **55**, 374-378. http://dx.doi.org/10.1016/j.archoralbio.2010.03.008

[55] Reich, B.J. and Bandyopadhyay, D. (2010) A Latent Factor Model for Spatial Data with Informative Missingness. *An-*

nals of Applied Statistics, **4**, 439-459. http://dx.doi.org/10.1214/09-AOAS278

[56] Ashour, S.K. and Abdel-Hameed, M.A. (2010) Approximate Skew-Normal Distribution. *Journal of Advanced Research*, **1**, 341-350. http://dx.doi.org/10.1016/j.jare.2010.06.004

[57] Liseo, B. and Parisi, A. (2013) Bayesian Inference for the Multivariate Skew-Normal Model: A Population Monte Carlo Approach. *Computational Statistics & Data Analysis*, **63**, 125-138. http://dx.doi.org/10.1016/j.csda.2013.02.007

[58] Figueiredo, F. and Gomes, I. (2013) The Skew-Normal Distribution in SPC. *Statistical Journal*, **11**, 83-104.

[59] Robert, C.P. and Casella, G. (2004) Monte Carlo Statistical Methods. Springer, New York.
http://dx.doi.org/10.1007/978-1-4757-4145-2

[60] Horrace, W.C. (2005) Some Results on the Multivariate Truncated Normal Distribution. *Journal of Multivariate Analysis*, **94**, 209-221. http://dx.doi.org/10.1016/j.jmva.2004.10.007

[61] Robertson, H.T. and Allison, D.B. (2012) A Novel Generalized Normal Distribution for Human Longevity and Other Negatively Skewed Data. *PLoS ONE*, **7**, e37025. http://dx.doi.org/10.1371/journal.pone.0037025

[62] Mcdonald, R.E., Avery, D.R. and Dean, J.A. (2004) Dentistry for the Child and Adolescent. 8th Edition, Mosby, New York.

[63] Albert, J. (2007) Bayesian Computation with R. Springer, New York. http://dx.doi.org/10.1007/978-0-387-71385-4

[64] Ward, E.J. (2008) A Review and Comparison of Four Commonly Used Bayesian and Maximum Likelihood Model Selection. *Ecological Modelling*, **211**, 1-10. http://dx.doi.org/10.1016/j.ecolmodel.2007.10.030

[65] Yiengprugsawan, V., Somkotra, T., Seubsman, S. and Sieigh, A.C. (2013) Longitudinal Associations between Oral Health Impacts and Quality of Life among a National Cohort of Thai Adults. *Health and Quality of Life Outcomes*, **11**, 172.
http://dx.doi.org/10.1186/1477-7525-11-172

Appendix A

We have

$$L = f\left(\boldsymbol{\beta}\,|\,\gamma,\theta,z\right) \propto f_0\left(\boldsymbol{\beta}\right) f\left(z\,|\,\theta,\gamma,\boldsymbol{\beta}\right)$$

$$f\left(z\,|\,\boldsymbol{\beta},\gamma,\theta\right) = \prod_{i=1}^{N} f\left(z_i\,|\,\boldsymbol{\beta},\gamma,\theta\right) = \prod_{i=1}^{N} N_n\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \gamma\boldsymbol{w}_i, \Sigma_\theta\right)$$

by assuming

$$f_0\left(\boldsymbol{\beta}\right) = N_p\left(\boldsymbol{\beta}_0, V_0\right)$$

then

$$\ln\left(L\right) = \ln\left[f\left(z\,|\,\boldsymbol{\beta},\gamma,\theta\right)\right] = \left[-\frac{p}{2}\ln\left(\pi_0\,|V_0|\right) - \frac{Nn}{2}\ln\left(\pi_0\,|\Sigma_\theta|\right)\right]$$

$$+ \left(-\frac{1}{2}\right)\left\{\left[\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)' V_0^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)\right] + \left[\sum_{i=1}^{N}\left(z_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)' \Sigma_\theta^{-1}\left(z_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right]\right\}$$

$$\frac{\mathrm{d}\ln\left(L\right)}{\mathrm{d}\boldsymbol{\beta}} = 2V_0^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) - 2\sum_{i=1}^{N}\boldsymbol{x}_i\Sigma_\theta^{-1}\left(z_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)$$

$$= \left(V_0^{-1} + \sum_{i=1}^{N}\boldsymbol{x}_i\Sigma_\theta^{-1}\boldsymbol{x}_i'\right)2\boldsymbol{\beta} - 2\left(\sum\boldsymbol{x}_i'\Sigma_\theta^{-1}z_i + V_0^{-1}\boldsymbol{\beta}_0\right) = 0$$

so

$$B_1 = \hat{\boldsymbol{\beta}} = \left(V_0^{-1} + \sum_{i=1}^{n}\boldsymbol{x}_i\Sigma_\theta^{-1}\boldsymbol{x}_i'\right)^{-1}\left(\boldsymbol{\beta}_0 V_0^{-1} + \sum_{i=1}^{N}\boldsymbol{x}_i'\Sigma_\theta^{-1}z_i\right)$$

Following that, we have

$$\frac{\mathrm{d}^2\ln\left(L\right)}{\left(\mathrm{d}\boldsymbol{\beta}\right)^2} = -\left(V_0^{-1} + \sum_{i=1}^{N}\boldsymbol{x}_i'\Sigma_\theta^{-1}\boldsymbol{x}_i\right),$$

since

$$V\left(\hat{\boldsymbol{\beta}}\right) = \left(-\frac{\mathrm{d}^2\ln\left(L\right)}{\left(\mathrm{d}\boldsymbol{\beta}\right)^2}\right)^{-1}$$

so

$$V_1 = V\left(\hat{\boldsymbol{\beta}}\right) = \left(V_0^{-1} + \sum_{i=1}^{N}\boldsymbol{x}_i'\Sigma_\theta^{-1}\boldsymbol{x}_i\right)^{-1}.$$