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

Received 3 August 2014; accepted 16 April 2015; published 20 April 2015

Copyright © 2015 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

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-Shaarawi [29], Mohammazadeh 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 Mohammazadeh [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 \( y_i = (y_{i1}, \ldots, y_{i\gamma}, \ldots, y_{in}) \) 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 \( x_i = (x_{i1}, \ldots, x_{im}) \) be the \( m \)-vector of covariates related to response variable \( y_i \). Thus, in lattice \( i \), \( y_i \) is a random field and the observations \( (y_{i1}, \ldots, y_{i\gamma}, \ldots, y_{in}) \) 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 \( p_{\eta} \), where \( y_i = 1 \) given all other values \( y_j \):

\[
p_{\eta}(\eta) = p(y_{ij}|y_{ik}, j \neq k, k \in N_i(j)) = \frac{\exp(x_j^T\beta + \gamma \sum_{k \in N_i} y_{ik})}{1 + \exp(x_j^T\beta + \gamma \sum_{k \in N_i} y_{ik})} \tag{1}
\]

where \( x_j \) as defined before is the vector of covariates of \( y_j \) with its first element 1; \( \beta = (\beta_1, \ldots, \beta_m)^T \) is the vector of coefficients of the covariates \( x_j \); and \( \gamma \) is the coefficient of spatial auto covariate \( \sum_{k \in N_i} y_{ik} \), such that \( N_i(j) = \{k : y_{kj} \text{ is a neighbor of } y_{ij}\} \) 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 \( y_{ij} \) (vector \( y_i \)) is

\[
p(y_j | \beta, \gamma) = (1/\phi(\beta, \gamma)) \exp \left[ \sum_{j=1}^{n} y_{ij} \left( x_{ij} \beta + \frac{1}{2} \gamma w_{ij} \right) \right]
\]

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

\[
\phi(\beta, \gamma) = \sum_{\text{for all } y_{ij}} \exp \left[ \sum_{j=1}^{n} y_{ij} \left( x_{ij} \beta + \frac{1}{2} \gamma w_{ij} \right) \right],
\]

and for all observations \( y = (y_1, \cdots, y_n) \)'

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

In this stage the set of parameters is denoted by \( \{\beta, \gamma\} \).

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 \( z_i = (z_{i1}, \cdots, z_{im}) \) be the vector of latent variables corresponding to \( y_i \), so that for every observed binary variable \( y_{ij} \), \( p(y_{ij} = 1) = p(z_{ij} > d_{ij}) \) and \( p(y_{ij} = 0) = p(z_{ij} < d_{ij}) \). 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-dimensional location parameters vector \( x_i \), \( n \times n \) positive-definite variance-covariance matrix \( \Sigma \) and skewness parameters vector \( a_i \), written as

\[
z_i \sim SN_n(x_i, \Sigma, \alpha_i),
\]

with probability density function

\[
f(z_i | \beta, \gamma) = 2 \phi_i(x_i \beta + \gamma w_i, \Sigma) \Phi(a_i \omega^{-1} (z_i - (x_i \beta + \gamma w_i)))
\]

where \( \phi_i \) and \( \Phi \) are n-dimensional normal density and standard normal cumulative distribution functions, respectively. \( a_i = a_1 \) such that \( 1_1 \) is a n \( \times \) 1 vector of 1 and \( a \in R \), \( x_i = (x_{i1}, x_{i2}, \cdots, x_{im}) \) and \( \omega \) is a diagonal matrix such that \( \Sigma = \omega \Sigma \omega \), where \( \Sigma \) is the correlation (positive definite) matrix of the vector \( z_i \), the spatially correlated variables \( z_{ij} \). If the element of \( \Sigma \) is denoted by \( \sigma_{ij} \) and \( \sigma_{ij} = \sigma_{ji} = 1 \), then \( \omega = \text{diag}(\sigma_1, \cdots, \sigma_m)^{1/2} \). The element of \( \Sigma \) in lattice \( i \) is defined in the equation below, such that

\[
\text{corr}(z_{ij}, z_{ik}) = k_{ij} \left(d_{jk}\right) = \theta^{\gamma_{ik}}
\]

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

\[
f(z_i | \eta) = 2 \phi_i(x_i \beta + \gamma w_i, \Sigma) \Phi(a \omega^{-1}(z_i - (x_i \beta + \gamma w_i)))
\]

and \( \eta = \{\beta, \gamma, \theta, a\} \), 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 $f(\eta, z | y)$ 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 done from $f(z | \eta, y)$.

a) In lattice $i$, the distribution of the skew-normal (latents) vector $z_i$ is

$$f(z_i | \eta, y) = 2\phi(z_i, \Sigma_y) \Phi(a'_i z_i) \prod_{j} \left[ I(z_i > 0) I(y_j = 1) + I(z_i < 0) I(y_j = 0) \right]$$

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:

$$f(z_j | z_{-j}, \eta, y) = \phi(z_j - \mu_{21}, 1 - \Sigma_{11} \Sigma_{21}^{-1} \Sigma_{22}) \times \Phi(a(z_j - \mu_{21}) + \varphi \sqrt{1 + a^2 (L_{21})}) \Phi(\varphi)$$

where,

$$z_{-j} = (z_{i1}, \ldots, z_{ij-1}, z_{j+1}, \ldots, z_{in})'$$

$$\mu_{21} = [x'_j \beta + w_j' + \Sigma_{12} \Sigma_{22}^{-1} (z_{-j} - x'_{-j} \beta_{-j})]$$

$$\varphi = a \times a' (z_{-j} - x'_{-j} \beta_{-j})$$

$$L_{21} = 1 - \Sigma_{12}^{-1} \Sigma_{21}^{-1} \Sigma_{22}, \quad \alpha'_j = a (I_{(m-1)} + \Sigma_{22}^{-1} \Sigma_{12}^{-1} I_{(n-1)}) \sqrt{1 + a^2 (1 - L_{21})}$$

$$x'_j = [x_{j1}, \ldots, x_{jk}, \ldots, x_{jq}]'$$

$$x_i = [x_{i1}, \ldots, x_{ij}, \ldots, x_{in}]'$$

$$x_{-j} = [x'_{i1}, \ldots, x'_{ij-1}, x'_j, \ldots, x_{in}]'$$

$$\beta = [\beta_1, \beta_2, \ldots, \beta_q]'$$

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

and $\Sigma_{12}$, $\Sigma_{22}$ and $\Sigma_{21}$ are the submatrixes of the $n \times n$ partitioned correlation matrix $\Sigma$ of the vector $(z_j, z_{-j})'$ 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):

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

$$\propto f(z | \beta, \theta, a, \gamma) f_0(\beta)$$
where $f(z|\beta,\theta,\alpha,\gamma) = \prod_{i=1}^{N} f(z_i|\eta)$ in which $z_i \sim SN_\alpha(x'_i\beta + \gamma w_i, \Sigma_\beta, a_\alpha)$

Such that

$$f(z|\beta,\gamma,\theta,\alpha) = 2\phi_\alpha(x'_i\beta + \gamma w_i, \Sigma_\beta)\Phi\left(a'_i, z_i\right)$$

or in other form

$$f(z|\beta,\gamma,\theta,\alpha) \propto 2\phi_\alpha(x'_i\beta + \gamma w_i, \Sigma_\beta)\Phi\left(a'_i, z_i - (x'_i\beta + w_i\gamma)\right)$$

$$= 2(2\pi)^{-N/2} |\Sigma_\beta|^{-1/2} \exp\left[\frac{-1}{2} (z_i - (x'_i\beta + w_i\gamma))' \Sigma_\beta^{-1} (z_i - (x'_i\beta + w_i\gamma))\right] \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \, dz_i,$$

since

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

$$\propto 2^N (2\pi)^{-N/2} |\Sigma_\beta|^{-N/2} \exp\left[\frac{-1}{2} \sum_{i=1}^{N} (z_i - (x'_i\beta + w_i\gamma))' \Sigma_\beta^{-1} (z_i - (x'_i\beta + w_i\gamma))\right] \prod_{i=1}^{N} \Phi\left(a\left(z_i - (x'_i\beta + w_i\gamma)\right)\right)$$

Then

$$f(\beta|z,\theta,\alpha,\gamma) \propto f(z|\eta) f_0(\beta)$$

where $f_0(\beta)$ is the prior distribution of the vector $\beta$ such that $\beta \sim f_\phi(\beta_0, V_0)$.

Therefore,

$$f(\beta|z,\theta,\gamma,\alpha) \\ \propto |\Sigma_\beta|^{-N/2} \exp\left[\frac{-1}{2} \sum_{i=1}^{N} (z_i - (x'_i\beta + w_i\gamma))' \Sigma_\beta^{-1} (z_i - (x'_i\beta + w_i\gamma))\right] \prod_{i=1}^{N} \Phi\left(a\left(z_i - (x'_i\beta + w_i\gamma)\right)\right)$$

$$\propto \phi_\alpha(\beta - u_\beta, v_\beta) \prod_{i=1}^{N} \Phi\left(a\left(z_i - (x'_i\beta + w_i\gamma)\right)\right),$$

where

$$u_\beta = \sum_{i=1}^{N} v_i \left(x'_i \Sigma_\beta^{-1} x_i + V_0^{-1} \beta_0 \right), \quad v_i = (x'_i \Sigma_\beta^{-1} x_i + V_0^{-1})^{-1}$$

$$w_\beta = \sum_{i=1}^{N} v_i \text{ and } \hat{\beta}_i = \left(x'_i \Sigma_\beta^{-1} x_i \right)^{-1} x'_i \Sigma_\beta^{-1} z_i$$

b-2):

$$f(\gamma|z,\theta,\beta,\alpha) \propto f(z|\gamma) f_0(\gamma) f(\gamma|\beta,\theta,\alpha) = f(z|\gamma, \beta, \theta, \alpha) g_0(\gamma)$$

(11).

where the above equations $g_0(\gamma)$ is the prior distribution of $\gamma$ and assumed Gamma distribution $G(a_\theta, b_\theta)$. So

$$f(\gamma|z,\beta,\theta,\alpha) \propto f(z|\gamma) g_0(\gamma) = \prod_{i=1}^{N} f(z_i|\eta) G(a_\theta, b_\theta)$$

$$= 2^N (2\pi)^{-N/2} |\Sigma_\beta|^{-N/2} \exp\left[\frac{-1}{2} \sum_{i=1}^{N} (z_i - (x'_i\beta + \gamma w_i))' \Sigma_\beta^{-1} (z_i - (x'_i\beta + \gamma w_i))\right] \prod_{i=1}^{N} \Phi\left(a\left(z_i - (x'_i\beta + \gamma w_i)\right)\right)$$

$$\times G(a_\theta, b_\theta)$$

(12)

$$= 2^N (2\pi)^{-N/2} |\Sigma_\beta|^{-N/2} \exp\left[\frac{-1}{2} \sum_{i=1}^{N} (z_i - (x'_i\beta + \gamma w_i))' \Sigma_\beta^{-1} (z_i - (x'_i\beta + \gamma w_i))\right] \prod_{i=1}^{N} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$

$$\times \frac{b_\theta^n}{\Gamma(a_\theta)} \gamma^{a_\theta-1} e^{-b_\theta \gamma}.$$
(Appendix A) are derived from $\partial \ln \left[ (z|\beta,\gamma,\theta,a) / \partial \gamma \right]$ and $-\partial^2 \ln \left[ (z|\beta,\gamma,\theta,a) / \partial^2 \gamma \right]$, respectively.

$$f(\theta|z,\beta,\alpha,\gamma) \propto f(z|\beta,\gamma,\alpha)f(\theta|\beta,\alpha,\gamma) \propto f(z|\beta,\gamma,\alpha)f_0(\theta)$$

$$= 2^{-N/2} \left| \Sigma_0 \right|^{-N/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} (z_i - (x_i^t \beta + w_i \gamma))^t \Sigma_0^{-1} (z_i - (x_i^t \beta + w_i \gamma)) \right] \prod_{i=1}^{N} \Phi \left( a \left( z_i - (x_i^t \beta + w_i \gamma) \right) \right) \frac{1}{\sqrt{2\pi}} e^{-r^2/2} dr \quad (13)$$

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

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

The value

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

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

$$\alpha = \min \left\{ \left( f(\theta|\beta,\gamma,\alpha,z) / f(\theta|\beta,\gamma,\alpha,z) \right) \times \left( \theta'(1 - \theta') / \theta(1 - \theta') \right) \right\}. \quad (14)$$

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

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

$$\propto f_0(a) \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} (z_i - (x_i^t \beta + w_i \gamma))^t \Sigma_0^{-1} (z_i - (x_i^t \beta + w_i \gamma)) \right] \prod_{i=1}^{N} \Phi \left( a \left( z_i - (x_i^t \beta + w_i \gamma) \right) \right) \quad (15)$$

where $f_0$ is the prior distribution of $a$ and since we assumed priori $\alpha \sim N(\alpha_0 = 2, \sigma^2 = 1)$, then

$$f(a|z,\beta,\theta,\gamma) \propto \exp \left[ -\frac{1}{2} (a - \alpha_0)^2 \right] \prod_{i=1}^{N} \Phi \left( a \left( z_i - (x_i^t \beta + w_i \gamma) \right) \right)$$

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

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 a dentist 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 education, 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 autocovariate (spacov) ($w_{ij}$). The coefficients of explanatory covariates and spatial autocovariate are denoted by $\beta = (\beta_0, \beta_1, \cdots, \beta_k)$ and $\gamma$ respectively. As was explained in previous work [54], the teeth caries statuses ($y_{ij}$) 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.

$$\logit(p_{ij}) = \beta_0 + \beta_1\text{ging} + \beta_2\text{gend} + \beta_3\text{age} + \beta_4\text{ocf} + \beta_5\text{edf} + \beta_6\text{vizd6} + \beta_7\text{tbr} + \beta_8\text{tfls} + \gamma\text{spacov} \quad (16)$$

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 $N_9(\mu_0 = (0, 0, 0, 0, 0, 0, 0, 0, 0), \sigma_0 = 2I_9)$, Gamma$(1, 2)$, Uniform$(0, 1)$ and $N(0,10)$ for the parameters $\beta, \gamma, \theta, a$, respectively, a sample of $z_{ij}$ was generated [51] from skew-normal distribution (8). Then, implementing these $z_{ij}$ in each of 50 iterations a sample of size 10000 of parameters $\eta = \{\beta, \gamma, \theta, a\}$ 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 1. 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.

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

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 $BIC = -2\ln(L(\tilde{\eta})) + q \log n$, where $L, q, m$ and $\tilde{\eta}$ 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.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Parameter</th>
<th>estimate</th>
<th>Standard error</th>
<th>95% Credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant B0</td>
<td>-0.786</td>
<td>0.36</td>
<td>(-1.106, -0.466)</td>
<td></td>
</tr>
<tr>
<td>Ginv B1</td>
<td>0.151</td>
<td>0.068</td>
<td>(0.15, 0.287)</td>
<td></td>
</tr>
<tr>
<td>Gend B2</td>
<td>0.0418</td>
<td>0.068</td>
<td>(-0.0942, 0.1778)</td>
<td></td>
</tr>
<tr>
<td>Age B3</td>
<td>0.018</td>
<td>0.016</td>
<td>(-0.014, 0.050)</td>
<td></td>
</tr>
<tr>
<td>Ocf B4</td>
<td>-0.0469</td>
<td>0.074</td>
<td>(-0.1949, 0.1011)</td>
<td></td>
</tr>
<tr>
<td>Edf B5</td>
<td>-0.0255</td>
<td>0.072</td>
<td>(-0.1695, 0.1185)</td>
<td></td>
</tr>
<tr>
<td>Vzd6 B6</td>
<td>0.022</td>
<td>0.037</td>
<td>(-0.052, 0.096)</td>
<td></td>
</tr>
<tr>
<td>Tbr B7</td>
<td>-0.004</td>
<td>0.093</td>
<td>(-0.19, 0.182)</td>
<td></td>
</tr>
<tr>
<td>Tfls B8</td>
<td>0.0012</td>
<td>0.027</td>
<td>(-0.0528, 0.0552)</td>
<td></td>
</tr>
<tr>
<td>Spacov $\gamma$</td>
<td>1.6</td>
<td>0.051</td>
<td>(1.57, 1.71)</td>
<td></td>
</tr>
<tr>
<td>$\theta$ (correlation)</td>
<td>0.54</td>
<td>0.046</td>
<td>(0.4, 0.48)</td>
<td></td>
</tr>
</tbody>
</table>

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

I much thank Dr Mohammad Ali Usofi, the head and associated Professor of Dentistry Faculty and miss Zainab Kazemi, the proficient and expert dentist in Immam Ali Dentistry Centre in Yasuj University of Medical Sciences, who prepared the tooth caries data.

### References


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


http://dx.doi.org/10.1016/j.archoralbio.2010.03.008

Years Old Children: A Spatial Modeling.


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


Appendix A

We have

\[ L = f(\mathbf{z} | \beta, \gamma, \mathbf{z}) \times f_0(\beta) f(\mathbf{z} | \theta, \gamma, \mathbf{z}) \]

\[ f(\mathbf{z} | \beta, \gamma, \mathbf{z}) = \prod_{i=1}^{N} f(z_i | \beta, \gamma, \mathbf{z}) = \prod_{i=1}^{N} N_o(x_i', \beta + \gamma w_i, \Sigma_o) \]

by assuming

\[ f_0(\beta) = N_p(\beta_o, V_0) \]

then

\[ \ln(L) = \ln\left[ f(\mathbf{z} | \beta, \gamma, \mathbf{z}) \right] = \left[ -\frac{p}{2} \ln(\pi_o | V_o) - \frac{Nn}{2} \ln(\pi_o | \Sigma_o) \right] \]

\[ + \left( -\frac{1}{2} \right) \left[ (\beta - \beta_o)' V_0^{-1} (\beta - \beta_o) + \sum_{i=1}^{N} (z_i - x_i' \beta)' \Sigma_o^{-1} (z_i - x_i' \beta) \right] \]

\[ \frac{d \ln(L)}{d \beta} = 2V_0^{-1} (\beta - \beta_o) - 2 \sum_{i=1}^{N} x_i' \Sigma_o^{-1} (z_i - x_i' \beta) \]

\[ = \left( V_0^{-1} + \sum_{i=1}^{N} x_i' \Sigma_o^{-1} x_i \right) (\beta - \beta_o) - 2 \left( \sum_{i=1}^{N} x_i' \Sigma_o^{-1} z_i + V_0^{-1} \beta_o \right) = 0 \]

so

\[ \beta_i = \hat{\beta} = \left( V_0^{-1} + \sum_{i=1}^{N} x_i' \Sigma_o^{-1} x_i \right)^{-1} \left( \beta_o V_0^{-1} + \sum_{i=1}^{N} x_i' \Sigma_o^{-1} z_i \right) \]

Following that, we have

\[ \frac{d^2 \ln(L)}{(d\beta)^2} = -\left( V_0^{-1} + \sum_{i=1}^{N} x_i' \Sigma_o^{-1} x_i \right), \]

since

\[ V(\hat{\beta}) = \left( -\frac{d^2 \ln(L)}{(d\beta)^2} \right)^{-1} \]

so

\[ V_i = V(\hat{\beta}) = \left( V_0^{-1} + \sum_{i=1}^{N} x_i' \Sigma_o^{-1} x_i \right)^{-1}. \]