Bayesian Modeling of Multivariate Spatial Binary Data with Application to Dental Caries

Dipankar Bandyopadhyay
dbandyop@umn.edu

Division of Biostatistics, School of Public Health
University of Minnesota
Minneapolis, MN 55455
http://www.biostat.umn.edu/~dipankar/
Joint work with

- Brian Reich (North Carolina State University)
- Elizabeth Slate (Medical University of South Carolina)
Bayesian Modeling of Multivariate Spatial Binary Data with Application to Dental Caries
Background

- Dental caries: Infectious disease damaging any tooth surface exposed to the oral cavity.

- Estimated 90% of schoolchildren worldwide and most adults (2003 WHO) have experienced caries, most prevalent in Asian and Latin American countries and least prevalent in African countries.

- Factors: (a) tooth surface (enamel or dentin), (b) cariogenic (or potentially caries-causing) bacteria, (c) fermentable carbohydrates (like sucrose) and (d) time.
Caries Index

- Tooth-level DMFS: Counts the total number of decayed, missing and filled surface within a particular tooth.

- Mouth level DMFS is also used, but there is evident loss (tooth level) information in doing that.
Clinical study (Fernandes et al., 2007) conducted at MUSC to determine dental caries status of Type-2 diabetic Gullah speaking African-Americans.

Population is highly understudied with the dental health status vastly unknown.
Data

Within-mouth quadrant has 8 teeth (3 molar, 2 pre-molar, 1 canine and 2 incisors).

Non-anterior tooth (Molars and Pre-molars) contributes 5 surfaces (occlusal, mesial, distal, facial and lingual) whereas anterior teeth (Incisors and Canines) contributes 4 surfaces (missing occlusal).

Response: \( \{ Y_{ij} : (i, j, s) \in D \} = \text{Binary (0/1) indicator whether the (i, j, s)th surface is either D/M/F, } i = 1, \ldots, 100, j = 1, \ldots, 32 \text{ and } s = 1, \ldots, 4 \text{ (I/C) and } s = 1, \ldots, 5 \text{ (PM/M)}. \)

Structure: Clustered data (subject \( \rightarrow \) tooth \( \rightarrow \) surface) \( \Rightarrow \) Random Effects Binary multilevel modeling.

Additionally, surfaces within a tooth, neighboring surfaces of adjacent teeth on the same jaw and touching surfaces of teeth on opposite jaws might experience spatial association.
Tooth surfaces

Mesial
All surfaces that are closest to midline of dental arch

Distal
All surfaces farthest from midline of dental arch

Facial
Surfaces toward the face. The term "facial" can be used in describing the tooth surface closest to the face for any tooth

Lingual
Surface of the maxillary and mandibular teeth nearest the tongue. This term may be applied to both maxillary and mandibular teeth

Occlusal
This term indicates the contacting, or biting, surfaces of all posterior teeth
Random sample of 100 subjects.

Potential co-variates are:

1. **Age**: Continuous with range from 27-73 years (mean = 53),
2. **Gender**: Binary female (1/0) indicator with about 75% females,
3. **Smoker**: Binary indicator with 36% smoker,
4. **Brush-floss**: Binary indicator whether brushed twice & flossed once every day (about 24%).
5. **Poverty**: Binary indicator with about 36% living below poverty line.
1. DMFS indicates tooth-level/mouth level cumulative caries severity and DOES NOT explore (i) surface-level caries experience or (ii) rate at which a surface possibly influences the decay of neighboring surfaces.

2. We propose to study effects of spatial association among different surfaces of tooth to determine dental caries status/progression.

3. We also propose to study the effect of key covariables, viz. age, gender, smoking habits, brush-floss habits and poverty indicator on dental caries after accounting for spatial associations.
Multivariate Binary Model

- Random effects logistic regression model + spatial association terms ⇒ Autologistic (Besag, 1972) model.

- We assume $Y_{ij} \sim \text{Bernoulli}(p_{ij})$ and

$$\text{logit}(p_{ij}) = \beta_0 + X_i^T \beta + G_{ij}$$

where $X$: subject-level co-variates,
$G_{ij}$: random effects corresponding to $(i, j, s)$th surface
$\beta$: fixed-effects parameter vector
$\beta_0$: Intercept
Spatial associations

Associations of **first-order** (surfaces within the same tooth), **second-order** (surfaces in between adjacent neighboring teeth on the same jaw) and **between-jaw** (touching surfaces in between two opposite jaws).

**Assumption**: Markov Random Field (MRF) model: A surfaces’ full conditional depends only on its neighbors.

**Random effects** $G_{ijs}$ is decomposed as:

$$G_{ijs} = U_i + E_{ijs}$$

$E_{ijs}$: Spatial RE; $U_i$: Uncorrelated RE
Spatial associations

\[ E_{ijs} = \sum_{l \neq s} b_{ls} Y_{ijl} + \sum_{m \sim j} \sum_{l} c_{ls} Y_{iml} + \sum_{p \leftrightarrow j} d_s Y_{ips} \]

- \( b = \{b_{ls} : l < s; l, s = 1, \ldots, 5\} \): first order,
- \( m \sim j \): teeth \( m \) and \( j \) are adjacent on the same jaw,
- \( c = \{c_{ls} : l, s = 1, \ldots, 5\} \): second-order,
- \( p \leftrightarrow j \): teeth \( p \) and \( j \) are contacting on opposite jaws,
- \( d = \{d_k : k = 1, \ldots, 5\} \): between jaw.

Under the autologistic specification:

\[
\Pr(Y_{ijs} = y_{ijs} | U, X, Y_{i'j's'}, (i, j, s) \neq (i', j', s')) = \frac{\exp\{y_{ijs}[\beta_0 + X^T_i \beta + U_i + E_{ijs}]\}}{1 + \exp\{\beta_0 + X^T_i \beta + U_i + E_{ijs}\}}
\]
Let \( \Omega = (\beta_0, \beta, b, c, d) \) : parameter vector.

Joint likelihood of the observed data is:

\[
L(\Omega, U; X, y) = c(\Omega)^{-1} \exp \left\{ \sum_{(i,j,s) \in D} y_{ijs} (\beta_0 + X_i^T \beta + U_i + \frac{1}{2} E_{ijs}) \right\}
\]

where

\[
c(\Omega) = \sum_{y} \exp \left\{ \sum_{(i,j,s) \in D} y_{ijs} (\beta_0 + X_i^T \beta + U_i + \frac{1}{2} E_{ijs}) \right\}
\]

and \( \sum_y \) denotes the sum over all possible \( 2^{|D|} \) realizations of \( Y \).
The normalizing constant $c(\Omega)$ doesn’t have a closed form, hence direct maximization of the likelihood is somewhat infeasible.

Estimation of the normalizing constant is a wide area of research, though not very exciting for the applied/clinical audience.

Often, $c(\Omega)$ is estimated by MCMC based methods like, path sampling (Gelman and Meng, 1998), MCMC MLE (Hu and Wuffer, 1997; Zheng and Zhu, 2008) etc; however all are computationally extensive.

Using a Bayesian framework, we pursue pseudo-likelihood (PL) approximation of the full likelihood.
Using Brook’s Lemma and HC-Theorem (Banerjee et al., 2004), the joint distribution of the spatial process (specified by the product of full conditionals) is well defined, and is given by:

$$\text{PL}(\Omega; U, X, y) = \prod_{(i,j,s) \in D} \Pr(Y_{ijs} = y_{ijs}|U, X, Y_{i'j's'}, (i, j, s) \neq (i', j', s'))$$

PL-based estimation is NO FREE LUNCH (Besag and Tantrum, 2003); might be unstable in presence of high intrinsic auto-correlation. (Wu and Huffer, 1997)

Recent simulation studies (Wintle, 2003 unpublished Ph.D. dissertation) shows robust behavior of PL.
Our approach is Bayesian: (a) ability to incorporate prior information, (b) posterior parameter inference and any arbitrary functionals, (c) no need to consider maximization of PL function, etc.

Related Bayesian methods (Heikkinen and Hogmander, 1994; Hoeting et. al, 2000;) rely on normal approximation of PL and related non-linear maximization steps.

Joint posterior density is given as:

$$p(\Omega, U, \tau, \sigma_u^2 | X, y) \propto PL(\Omega, U; X, y) \times \pi_0(\beta_0) \times \pi_1(\beta) \times \pi_2(\theta_1 | \tau) \times \pi_3(U | \sigma_u^2) \times \pi_4(\tau^2) \times \pi_5(\sigma_u^2)$$

where $\pi_j(\cdot), j = 0, \ldots, 5$ denote the prior/hyperpriors.

Our method uses WinBUGS with automated Gibbs sampling (Gilks et.al, 1994) to sample from the pseudo-likelihood.
Priors on $\beta$ are weakly informative
\[ N(0, \text{Precision} = 0.25) \Rightarrow \text{OR} \in (-4, 4). \]

$\beta_0$ is flat uniform (using the dflat() option in WinBUGS).

$\mathbf{b}, \mathbf{c}, \mathbf{d}$ are skeptical, sparsity inducing, 0-peaked $\text{DE}(0, \tau)$ (Bayesian Lasso). Prior on $\tau^2$ is Gamma(1,1) density ($\equiv \text{Exp}(1)$) to reflect a diffuse hyperprior.

Random effects $U \sim \text{Normal}(0, \text{Precision} = \frac{1}{\sigma_u^2})$ with
\[ \sigma_u \sim \text{Uniform}(0, 100) \]
following Gelman (2006).

WinBUGS(1.4.3) and R2WinBUGS package, 2 MCMC chain of length 50000 with a ‘very-high’ burn-in of 45000. Model convergence: Trace plots, ACF plots and Gelman-Rubin (R-Hat) statistics.
We compare 4 models:

1. Model 1: Simple random effects logistic regression with no spatial association

2. Model 2: Random effects autologistic regression with same autologistic parameters, i.e., $b = c = d$

3. Model 3: Random effects autologistic regression with only the first-order autologistic parameters, i.e., $c = d = 0$

4. Model 4: Random effects autologistic regression with varying autologistic parameters, i.e., $b \neq c \neq d$. 

Bayesian Modeling of Multivariate Spatial Binary Data with Application to Dental Caries
Model Comparison

Θ: entire parameter space & $y_{pr}$ denotes the predictive data vector, the posterior predictive distribution is given by:

$$p(y_{pr} \mid y) = \int p(y_{pr} \mid \Theta)p(\Theta \mid y)d\Theta$$

Cross-validation. Average posterior misclassification probabilities for deleted zeros and ones respectively.

$$CV_1 = N_1^{-1} \sum_{(i,j,s)} \{y_{ijs} \ast (1 - y_{ijs,pr})\}$$

$$CV_0 = N_0^{-1} \sum_{(i,j,s)} \{(1 - y_{ijs}) \ast y_{ijs,pr}\}$$

where $N_0$: total number of 0’s deleted and $N_1$: total number of 1’s deleted from our study.
Posterior predictive L-measure (Chen, M-H., et al., 2003)

\[ L = \sum_{i=1}^{n} \sigma_{i(M)}^2 + \frac{k}{k+1} \sum_{i=1}^{n} (\mu_{i(M)} - M_i)^2 \]

\[ = P + \nu G \left( \frac{k}{k+1} \right) \]

\[ \mu_{i(M)} = E(M_{i,pr}|y) \quad \text{&} \quad \sigma_{i(M)}^2 = \text{Var}(M_{i,pr}|y) \quad \text{where} \]

\[ M_{i,pr} = \sum_{j,s} w_{js} y_{ijs,pr} \quad \text{and} \quad w_{js} = 1. \]

G: Goodness of fit; P: Penalty; \( \nu = 1. \)
Both measures are computed using MCMC samples within WinBUGS and smaller values are preferred.

Table 1: Model comparison using cross-validation and L-measure

<table>
<thead>
<tr>
<th>Model</th>
<th>$CV_0$</th>
<th>$CV_1$</th>
<th>$G$</th>
<th>$P$</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.338</td>
<td>0.515</td>
<td>3577.0</td>
<td>2364.0</td>
<td>5941.0</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.223</td>
<td>0.259</td>
<td>2474.0</td>
<td>500.5</td>
<td>2974.5</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.111</td>
<td>0.108</td>
<td>823.1</td>
<td>813.7</td>
<td>1636.8</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.067</td>
<td>0.096</td>
<td>557.0</td>
<td>493.7</td>
<td>1050.7</td>
</tr>
</tbody>
</table>

Model 4, the full autologistic model reigns supreme!
Spatial association parameter estimates of 5 surfaces within a tooth. Surfaces are 1 = Occlusal, 2 = Mesial, 3 = Distal, 4 = Facial, 5 = Lingual.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{12}$</td>
<td>1.922</td>
<td>0.195</td>
<td>(1.572, 2.225)</td>
</tr>
<tr>
<td>$b_{13}$</td>
<td>2.806</td>
<td>0.104</td>
<td>(2.593, 2.969)</td>
</tr>
<tr>
<td>$b_{14}$</td>
<td>0.631</td>
<td>0.251</td>
<td>(0.238, 1.049)</td>
</tr>
<tr>
<td>$b_{15}$</td>
<td>1.551</td>
<td>0.204</td>
<td>(1.144, 1.874)</td>
</tr>
<tr>
<td>$b_{23}$</td>
<td>1.636</td>
<td>0.157</td>
<td>(1.28, 1.897)</td>
</tr>
<tr>
<td>$b_{24}$</td>
<td>3.386</td>
<td>0.142</td>
<td>(3.052, 3.592)</td>
</tr>
<tr>
<td>$b_{25}$</td>
<td>2.152</td>
<td>0.155</td>
<td>(1.869, 2.51)</td>
</tr>
<tr>
<td>$b_{34}$</td>
<td>1.961</td>
<td>0.164</td>
<td>(1.625, 2.259)</td>
</tr>
<tr>
<td>$b_{35}$</td>
<td>2.697</td>
<td>0.214</td>
<td>(2.21, 3.017)</td>
</tr>
<tr>
<td>$b_{45}$</td>
<td>2.713</td>
<td>0.138</td>
<td>(2.418, 2.926)</td>
</tr>
</tbody>
</table>

Strongest: (Mesial, Facial) and Weakest: (Occlusal, Facial).
### Results: Posterior Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_{11}$</td>
<td>1.95</td>
<td>0.0931</td>
<td>(1.781, 2.124)</td>
</tr>
<tr>
<td>$c_{12}$</td>
<td>-0.708</td>
<td>0.0941</td>
<td>(-0.881, -0.534)</td>
</tr>
<tr>
<td>$c_{13}$</td>
<td>-0.512</td>
<td>0.1056</td>
<td>(-0.6968, -0.275)</td>
</tr>
<tr>
<td>$c_{14}$</td>
<td>0.016</td>
<td>0.1392</td>
<td>(-0.292, 0.245)</td>
</tr>
<tr>
<td>$c_{15}$</td>
<td>-0.218</td>
<td>0.114</td>
<td>(-0.422, -0.023)</td>
</tr>
<tr>
<td>$c_{22}$</td>
<td>1.219</td>
<td>0.253</td>
<td>(0.816, 1.775)</td>
</tr>
<tr>
<td>$c_{23}$</td>
<td>2.048</td>
<td>0.193</td>
<td>(1.724, 2.501)</td>
</tr>
<tr>
<td>$c_{24}$</td>
<td>-1.538</td>
<td>0.123</td>
<td>(-1.758, -1.323)</td>
</tr>
<tr>
<td>$c_{25}$</td>
<td>-0.822</td>
<td>0.116</td>
<td>(-1.044, -0.603)</td>
</tr>
<tr>
<td>$c_{33}$</td>
<td>-0.091</td>
<td>0.262</td>
<td>(-0.59, 0.278)</td>
</tr>
<tr>
<td>$c_{34}$</td>
<td>-0.869</td>
<td>0.183</td>
<td>(-1.186, -0.381)</td>
</tr>
<tr>
<td>$c_{35}$</td>
<td>-0.274</td>
<td>0.197</td>
<td>(-0.582, 0.063)</td>
</tr>
<tr>
<td>$c_{44}$</td>
<td>3.934</td>
<td>0.162</td>
<td>(3.665, 4.249)</td>
</tr>
<tr>
<td>$c_{45}$</td>
<td>-0.908</td>
<td>0.196</td>
<td>(-1.371, -0.636)</td>
</tr>
<tr>
<td>$c_{55}$</td>
<td>2.29</td>
<td>0.205</td>
<td>(1.937, 2.667)</td>
</tr>
<tr>
<td>$d_{11}$</td>
<td>2.41</td>
<td>0.211</td>
<td>(2.123, 2.883)</td>
</tr>
</tbody>
</table>
### Results: Fixed/random effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev</td>
</tr>
<tr>
<td>intercept</td>
<td>-1.3</td>
<td>0.44</td>
</tr>
<tr>
<td>age</td>
<td>2.07</td>
<td>0.68</td>
</tr>
<tr>
<td>gender</td>
<td>0.43</td>
<td>1.74</td>
</tr>
<tr>
<td>smoking status</td>
<td>-1.72</td>
<td>3.31</td>
</tr>
<tr>
<td>brush-floss</td>
<td>0.72</td>
<td>0.85</td>
</tr>
<tr>
<td>poverty</td>
<td>-0.51</td>
<td>1.23</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>3.23</td>
<td>0.41</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.53</td>
<td>0.13</td>
</tr>
</tbody>
</table>

**warning**! Fixed-effects parameters for Models 1 and 4 cannot be directly compared (high spatial associations).

In Model 4, there is no significant fixed-effects! All has been sucked up by spatial components. Same for Model 3.

95% CI are however tighter for Models 3 and 4.
Conclusions

- A multivariate extension of the autologistic model with convenient odds-ratio interpretation.

- Can be estimated within WinBUGS, using PL-approximation of the full likelihood.

- Our full autologistic model provides marked improvement in prediction. High level of spatial associations observed among the 5 surfaces for (a) surfaces within tooth, (b) surfaces in neighboring tooth on same jaw, and (c) touching surfaces in neighboring tooth on opposite jaw.

- Despite its criticism by a ‘full-likelihood’ minded audience, this research underscores the potential of the pseudo-likelihood autologistic framework in analyzing multivariate spatial binary dental data.
Acknowledgements

This research is supported by South Carolina COBRE for Oral Health (NIH/NCRR Grant: P20 RR-017696-06) / Center for Oral Health Research at MUSC and NIH 1R01LM009153.

To Andrew Lawson for interesting discussions.