Non-parametric approaches to detecting "Wombling Boundaries" on Disease Maps

Pei Li and Sudipto Banerjee

Dept. of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, U.S.A.

June 11, 2008

Outline:

- Essentials of statistical disease mapping
- What is “wombling”?
- Hierarchical models for wombling on disease maps
- Non-parametric approaches
- Some illustrations
- Summary

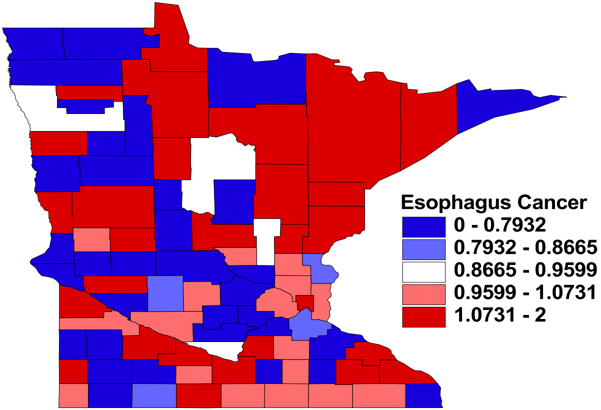
- Disease mapping refers to the methods for:
  - describing the geographic variation of disease
  - generating hypotheses about the possible causes for differences in risk of disease.
- Related spatial databases:
  - National Center for Health Statistics (NCHS)
  - Surveillance, Epidemiology, and End Results (SEER) from National Cancer Institute
- Standard mortality ratio: $SMR = Y_i/E_i$
  - Y_i is the observed number of deaths in region i.
  - E_i is the expected number of deaths in region i.

Disease maps are used to:

- Highlight geographic areas with high and low incidence
- Detect spatial clusters which may be due to common environmental, demographical effects shared by neighboring regions.
- Generate hypotheses about the possible cause for apparent difference in risk.

All these are achieved through statistical models that regress and smooth.

Map of raw standard mortality ratios (SMR) of esophagus cancer between 1991 and 1998 in Minnesota counties



- Recently spatial analysts have shown a growing interest in detecting boundaries that reveal sharp changes in the values of the variables.
- The general problem of identifying zones of abrupt change is known as “wombling”.
- Named after a foundational paper by William Womble (*Science*, 1953).
- For spatial (Gaussian) process surfaces one seeks distribution theory for directional (Banerjee et al., 2003) and curvilinear (Banerjee and Gelfand, 2006) gradients.
- Adaptation to point-process data (Cox process intensity surfaces) (Liang et al., 2008).
- What about wombling on a map?

- We would like a formal procedure to seek out "boundaries" that delineate/separate regions with vastly differing outcomes.
- But we should do so after accounting for regressors/predictors in the model.
- Consider a random effects generalized linear model for data from K counties:

$$g(E[Y_k]) = \eta_k = \mathbf{x}_k'\boldsymbol{\beta} + \phi_k; \quad k = 1, \ldots, K$$

- We could detect wombling boundaries by detecting pairwise clusters of η_k's. Or perhaps ϕ_k's.

- One approach: compute the posterior probs of all pairwise ties and retain only those edges whose incident regional effects have smaller probs of being equal.
- Let $Z_{ij} = 1$ if $P(\phi_i = \phi_j \,|\, i \sim j;\, Y) < c$, where c is some threshold. So Z_{ij}'s will be indicators for wombling boundaries.
- We need ϕ's to be almost sure discrete. So use DP (Muller and Quintana, 2004):

$$Y_k \overset{iid}{\sim} Po(e^{\eta_k});\ \eta_k = \mathbf{x}_k'\boldsymbol{\beta} + \phi_k;\quad i = 1, \ldots, k$$
$$\phi_k \sim DP(\alpha G_0);\ G_0 \equiv N(0, \sigma^2);\ \sigma^2 \sim IG(a_\sigma, b_\sigma).$$

  Note: α controls degree of clustering. Values close to 0 induce many ties; larger values induce fewer ties.
- Criticism: Model does not accommodate spatial associations.

- Build dependence over adjacency relationships ("graphs") using conditional distributions:

$$\phi_i \,|\, \phi_j,\, j \neq i \sim N\left(\rho \sum_{j=1}^{K} \frac{w_{ij}}{w_{i+}} \phi_j, \frac{\tau^2}{w_{i+}}\right)$$

- $W = w_{ij}$ is "proximity matrix" connecting regions i and j through a binary relation $\sim$: "is neighbor of"
- The joint distribution can be derived as

$$p(\phi_1, \ldots, \phi_K) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\phi}'(D - \rho W)\boldsymbol{\phi}\right\};\, D = Diag\{w_{i+}\}.$$

- Caution: The above distribution may be improper! $\rho \in \left(\frac{1}{\lambda_{(1)}}, 1\right)$ ensures propriety, where $\lambda_{(1)}$ is the smallest eigen-value of $D^{-1/2} W D^{-1/2}$.

- The parametric approach to wombling attempts to model the adjacency matrix itself. The W matrix becomes a random binary matrix with some distribution attached to it.
- Lu et al. (2007) modelled w_{ij} using a logistic regression in the hierarchy.
- Ma, Carlin and Banerjee (2008) employ compound Gaussian Markov Random fields to hierarchically model the sites and edges.
- Methods work but sensitive to prior assumptions and parameter restrictions.
- Several issues arise concerning impropriety of posteriors, identifiability of parameters and MCMC convergence issues.

How can we embed a Markov Random Field within a non-parametric framework?

- Consider the *stick-breaking* prior Sethuraman (1994):

$$F \stackrel{d}{=} \sum_{i=1}^{N} p_i \delta_{\theta_i},$$

where δ_θ is a Dirac distribution with point mass at θ.

- $\theta_1, \ldots, \theta_N \stackrel{iid}{\sim} G_0$, where G_0 is a fixed distribution.
- The mixtures yield

$$p_i = V_i \prod_{j<i}(1 - V_j);\ V_1, \ldots, V_{N-1} \stackrel{indep}{\sim} Beta(a_i, b_i)$$

The stick-breaking distribution is characterized by "locations" (atoms) θ_i and "masses" V_i.

- The mixture probabilities "break the stick" into N pieces so that $\sum_{i=1}^{N} p_i = 1$:

$$p_1 = V_1 \sim Beta(a,b);\; p_i = \left(1 - \sum_{j=1}^{i-1} p_j\right) V_i,\; i = 1,\ldots,N,$$

where $V_1,\ldots,V_{N-1} \overset{iid}{\sim} Beta(a,b)$. For finite N,

$$V_N = 1 \Rightarrow \sum_{i=1}^{N} p_i = 1 \text{ and } p_N = \prod_{j<N} (1 - V_j)$$

- A special case of this prior is the Dirichlet Process prior:

$$N = \infty;\; V_i \overset{iid}{\sim} Beta(1,\alpha) \Rightarrow F \overset{d}{=} DP(\alpha G_0).$$

- Dependence structures introduced by modelling the locations and/or the masses.
  - MacEachern (1999); Gelfand, Kottas and MacEachern (2004): let the locations vary spatially (a Gaussian process).
  - De Iorio et al. (2004): similar modelling to achieved ANOVA-type structures.
- Idea: index locations by space, say $\theta(\mathbf{s}) \sim GP(0, C(\cdot, \cdot))$. Problem: Resulting distribution is a mixture of DP's and posterior process has updated mass parameter $\alpha + n$ *for all* ***s***
- Griffin and Steel (2006): remedy this problem with a "Dependent DP" using "ordering".
- Reich and Fuentes (2008): Model locations using a multivariate point-referenced process.

- A hierarchical model for spatial count data:

$$Y_k \,|\, \eta_k \stackrel{iid}{\sim} Po(e^{\eta_k}); \; \eta_k = \mathbf{x}_k' \boldsymbol{\beta} + \phi_k; \quad k = 1, \ldots, K;$$

$$\phi_k \sim F_k = \sum_{i=1}^{N} p_{ik} \delta_{\theta_i}; \; \theta_i \stackrel{iid}{\sim} N(0, \sigma^2);$$

$$p_{ik} = V_{ik} \prod_{j<i} (1 - V_{jk}); \; V_{ik} = w_{ik} V_i; \; V_i \stackrel{iid}{\sim} Beta(a, b);$$

$$w_{ik} \in [0, 1]; \; w_{ik} = g^{-1}(u_{ik}); \; \mathbf{u}_i = \{u_{ik}\}_{k=1}^{N};$$

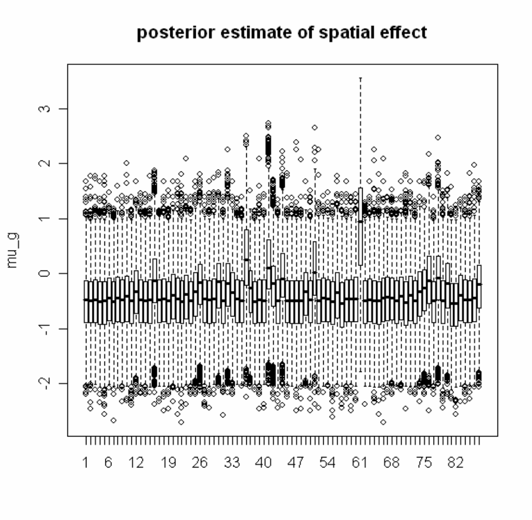
$$\mathbf{u}_i \stackrel{iid}{\sim} N(\mathbf{0}; \tau^2 (D - \rho W)^{-1}); \; \rho \in (0, 1)$$

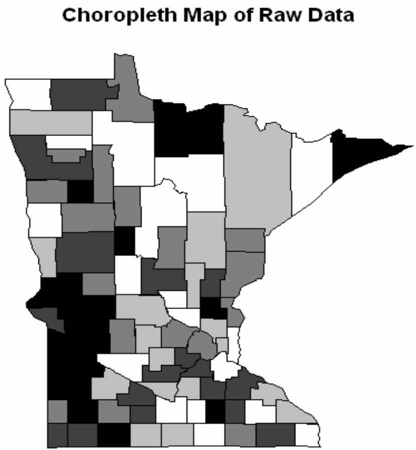
$$\sigma^2 \sim IG(a_\sigma, b_\sigma); \; \boldsymbol{\beta} \sim flat; \; \rho \sim Beta(a_\rho, b_\rho); a, \, b \sim U(0, 10).$$

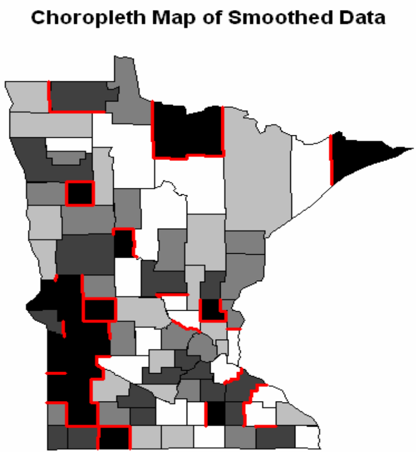
- The prior for ρ should center ρ close to its upper bound.

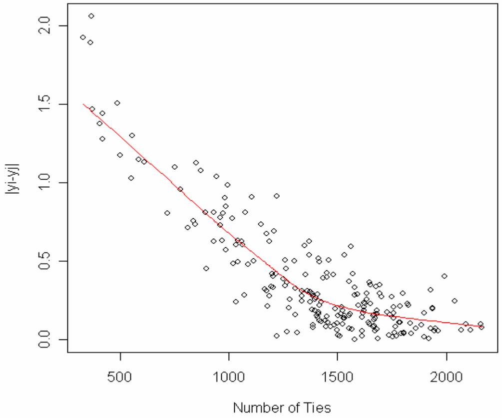
- We illustrate our approach in the context of a Minnesota Pneumonia and Influenza diagnosis dataset
- This dataset provides the cases of pneumonia and influenza older than 65 years who are enrolled in both medicare in December 2001.
- Covariate to adjust for: average income per person in each county.
- Boundary analysis for this map might help identify barriers separating counties with different diagnosis rates - help study the causes behind healthcare barriers.

Plot of posterior distributions of the spatial effects




Choropleth Map of Raw Data


Choropleth Map of Smoothed Data


|yi-yj|
0.0
0.5
1.0
1.5
2.0
500
1000
1500
2000
Number of Ties

- Griffin & Steel (2006): Introduced a highly flexible class of order-based dependent stick-breaking prior.
- Based upon a stochastic process $\{\boldsymbol{\pi}(\mathbf{s})\}_{\mathbf{s}\in D}$ (an *order on* ***s***), such that:

  $\{\pi_1(\mathbf{s}), \ldots \pi_{n(\mathbf{s})}\} \subseteq \{1, \ldots, N\}$ for some $n(\mathbf{s}) \leq N$;

  $\pi_i(\mathbf{s}) = \pi_j(\mathbf{s})$ if and only if $i = j$.
- A Dependent Stick-Breaking Prior $DSBP(\boldsymbol{\pi}, G_0, \{a_i, b_i\}_{i-1}^N, N)$:

$$F_{\mathbf{s}} \stackrel{d}{=} \sum_{i=1}^{N} p_i(\mathbf{s}) \delta_{\theta_{\pi_i(\mathbf{s})}},$$

$$p_i(\mathbf{s}) = V_{\pi_i(\mathbf{s})} \prod_{j<i} (1 - V_{\pi_j(\mathbf{s})}),$$

$$V_i \stackrel{ind}{\sim} Beta(a_i, b_i);\ \theta_i \stackrel{iid}{\sim} G_0.$$

- The $DSBP(\boldsymbol{\pi}, G_0, 1, \alpha, \infty)$ leads to a $DDP(\pi, G_0, \alpha)$: a very flexible class of models.
- The framework includes Bayesian Partition Models:
  - Denison, Holmes, Mallick and Smith (2002);
  - Ferreira, Dennison and Holmes (2002);
  - Knorrheld and Raβer (2000).
- This work was supported by:

NSF Grant DMS 0706870;
NIH Grant R01-CA-95995.

- Thank you!