A Hierarchical Model for Spatially Clustered Disease Rates

Ronald E. Gangnon and Murray K. Clayton

March 19, 2002

Department of Biostatistics and Medical Informatics
University of Wisconsin - Madison

http://www.biostat.wisc.edu/~ronald
- Inner neighbors, area, perimeter.

- Geographic centroid of cell from Census Bureau.

- Population from 1980 U.S. Census.


- Census tracts or census blocks.

- Black counties in Upstate New York.

New York Leukemia Data
New York Leukemia Data
• If so, what are the risks associated with the cluster?

• If so, where are the clusters located?

Other definitions possible:

Small areas with high (or low) leukemia rates.

• Is there evidence of clustering?

• Is there evidence of a nonconstant leukemia rate?

What is the leukemia rate for a cell?
Estimate \( d \) (disease mapping) 

Test \( H_0: d = \bar{d} \) (clustering) 

Goals of analysis:

- \( \bar{y} \) Poisson(\( d \cdot u \)).
- \( \bar{y} \) true rate in cell \( i \).
- \( u \) population at risk in cell \( i \).
- \( y \) number of cases in cell \( i \).
- \( i = 1, 2, \ldots, N \) cells.

Basic Statistical Model

Hierarchical Model for Spatial Clustering

Gangnon
Cangnan and Clayton (2001)
Waller et al. (1992), Lawson (1993)
Stone (1998)

Model-based statistics

Tango (2000)

Excess rate statistics

Bonetti (2000)

Whittlemore et al. (1987)

Distance-based statistics

Tests for Spatial Clustering

Hierarchical Model for Spatial Clustering

Point process models

Cressie, James & Clayton (2000)


Partitioning methods

Besag, York & Mollie (1991), Waller et al. (1997)

Local shrinkage models

Clayton & Kalder (1987)

Global shrinkage models

Bayes or Empirical Bayes Approaches

Hierarchical Model for Spatio-temporal Clustering

Cagnan
\( \epsilon_1, \ldots, \epsilon_N \text{ i.i.d. } \mathcal{N}(0, 1) \)

Exterior Poisson variation: \( \epsilon \)

\[
\theta \left\{ \left\{ \epsilon \in \mathcal{C} \right\} \right\} \frac{1}{\theta} \prod_{\gamma} \sum_{l=1}^{\gamma} \log(\gamma) \]

Note: \( \theta \) is the parameter of interest.

\( 0 \) is the log relative risk associated with cluster \( \mathcal{C} \).

\( \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_\gamma \) denote sets of cells belonging to each cluster.

\( \gamma \) is possibly overlapping clusters.

A factor prior for \( \eta \) (or \( \gamma \))

Fixed effect: \( \eta \) can be replaced with \( x \).

\[
\log(\gamma) = \gamma + \left\{ \left\{ \epsilon \in \mathcal{C} \right\} \right\} \frac{1}{\theta} \prod_{\gamma} \sum_{l=1}^{\gamma} \eta = (\gamma \theta) \]

Proposed Model for Spatial Clustering

Hierarchical Model for Spatial Clustering

Ganagnon
Proposed Model for Spatial Clustering
Discrete uniform prior for $\mathcal{K}$.

Finite number of potential clusters available.

$\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k \text{ i.i.d. } p(c) \cdot \theta_1, \theta_2, \ldots, \theta_k \text{ i.i.d. } N(0, \sigma^2) \cdot$

$\omega \geq \theta_\sigma \geq \omega = 0.99.$

We use $\theta_\sigma = 0.355$ so that $P(0.25 \leq \omega \mid \theta_\sigma) = 0.99$.

$\theta$ must be fixed.

Prior Distribution for $\theta_1, \theta_2, \ldots, \theta_k \cdot$

$\{f \in \mathcal{F} \mid \theta \subseteq \mathcal{F} \}$
Potential Clusters for New York Data

- Place prior on center/radius of circle adjusted to avoid empty circles.
- Cell belongs to cluster if centroid falls inside circle.
- Maximum geographic radius.
- Radius ranging from 0 km up to 20 km.
- Circular clusters centered at cell centroids.
Conjugate Gamma prior leads to (exact) Gibbs update.

- Parameter $T$.

Approximate Gibbs sampler serves as proposal distribution for Metropolis algorithm.

Posterior Simulation

Hierarchical Model for Spatial Clustering

Conjugate prior to normal approximation for

Normal priors conjugate to normal approximation for

Parameters $\theta_1, \theta_2, \ldots, \theta_k, \epsilon_1, \ldots, \epsilon_N$.

General approach to inference given in Gehman et al. (1995).

Hierarchical Poisson generalized linear model.

Suppose $c_1, \ldots, c_R$ are known.
Select one of the available transitions at random.

CHANGE the composition of a cluster.

DROP a cluster.

ADD a new cluster.

Possible transitions between models:

1997) to transition between models.

Use reversible jump MCMC (RJMCMC) algorithm (Green).

In reality, $k$ and $c_1, \ldots, c_r$ are unknown.

**Posterior Simulation**
ADD Cluster
DROP Cluster
CHANGE Cluster
New York Leukemia Data
Posterior Distribution for $k$

Uniform Prior

Geometric Prior
Cluster for $k = 3$

Posterior Probability Cell Belongs to a Hierarchical Model for Spatial Clustering
\[ \kappa = 3 \]

Posterior Mean for Clustering Effect for Hierarchical Model for Spatial Clustering of Cancer
Concluding Remarks

- Straightforward
- Inclusion of covariates effects and/or temporal effects is
- Formal methods available for identifying the number of clusters.
- Reversible jump MCMC algorithm for inference.
- Natural prior specification for clustering component of model.
- Extra-Poisson variation.
- Proposed spatial model for clustering effects, which includes

Hierarchical Model for Spatial Clustering

Cagnon