Spatial Modeling for Areal Data: Spatial Autoregression

Sudipto Banerjee

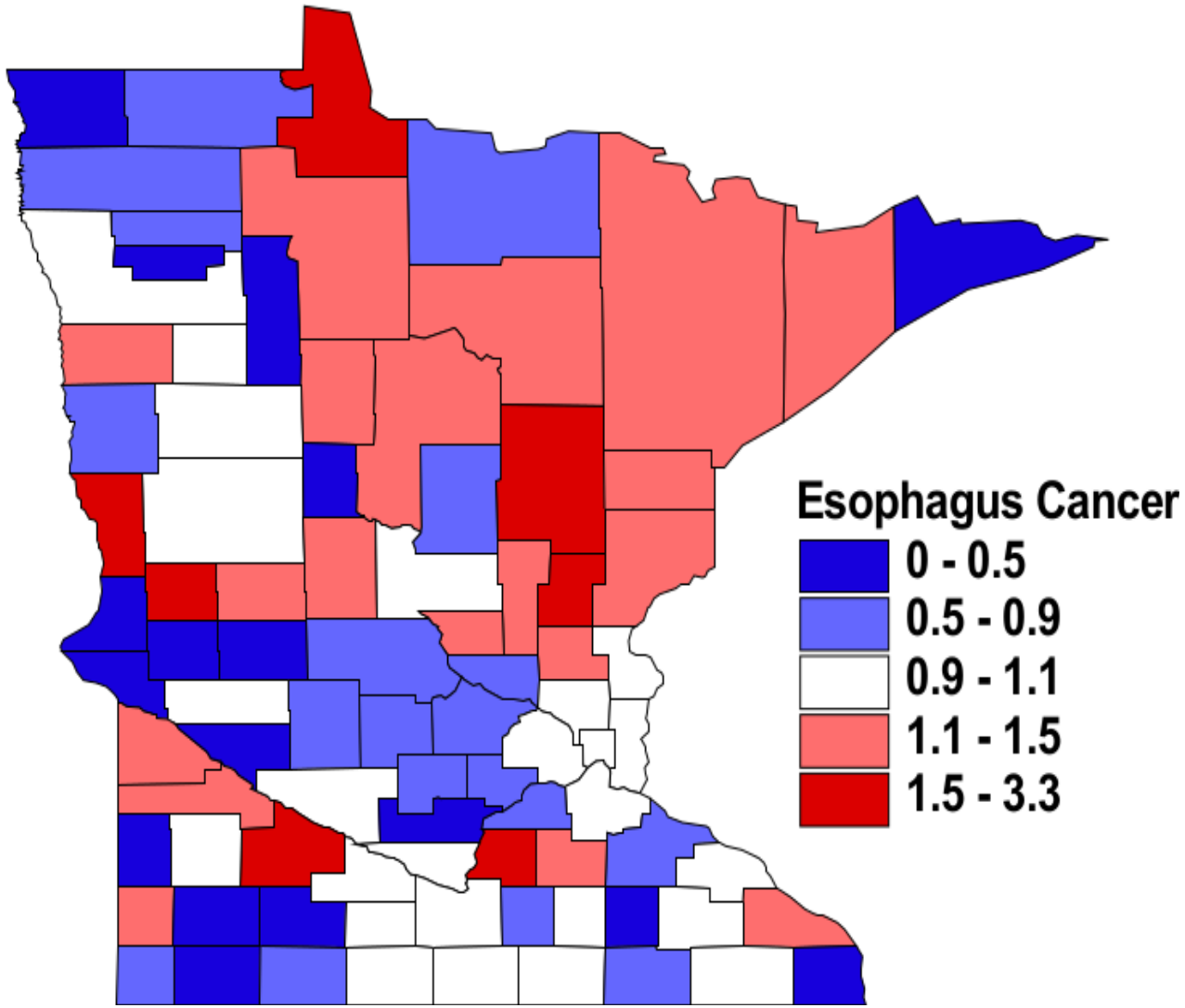
University of California, Los Angeles, USA

Spatial Data Analysis

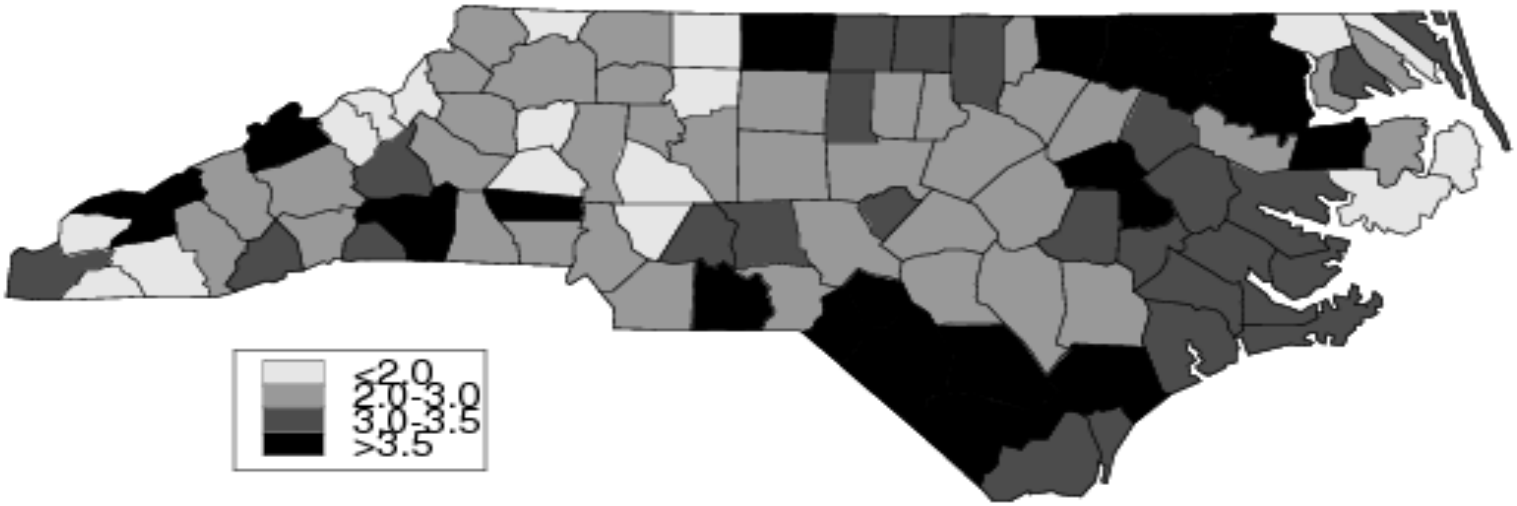
- Analyzing datasets that are referenced with respect to locations where they are observed or measured
- There are three types of spatial data based upon notions of "proximity" and underlying process:
  - Point-referenced data: Data are indexed by fixed coordinates (e.g., Lon-Lat; Easting-Northing). Example: Pollutant levels measured at fixed monitoring stations.
  - Point-process data: Locations themselves arise as realizations of a random process. Example: Locations emerging as disease cases in a region, as tumors on an organ etc.
  - Areal or regionally aggregated (lattice) data: Data are indexed by areas (e.g., geographic regions delineated by political or demographic boundaries, pixels on an image or grid etc.)
- Areal data are by far the most common in public health because point-level information is usually unavailable due to data confidentiality protocols involving human subjects.

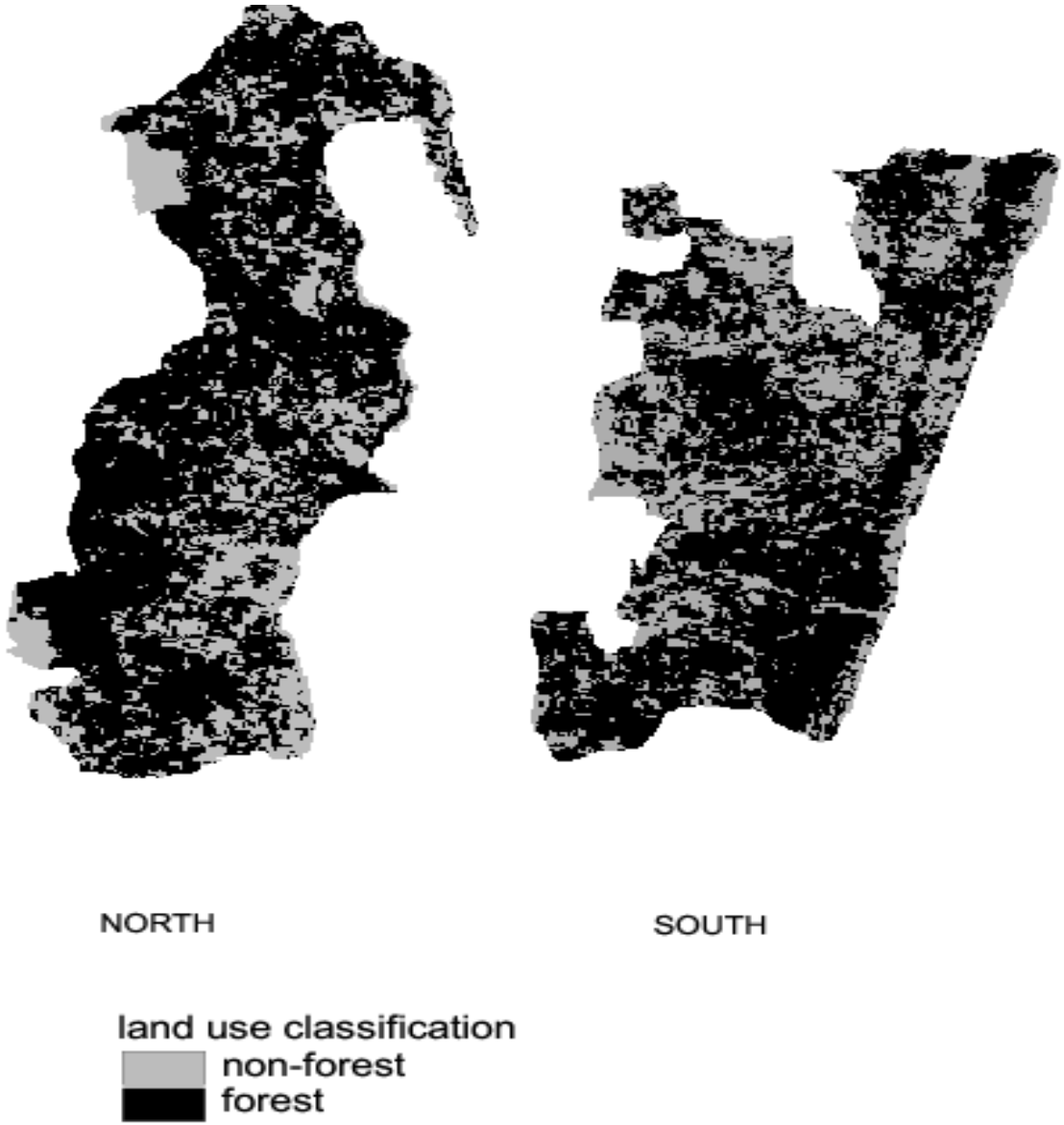
Disease Mapping: Mapping raw data

Standardized Mortality Rates (SMR) for esophagus cancer across 87 counties in Minnesota.



Actual Transformed SIDS Rates




NORTH
SOUTH
land use classification
non-forest
forest

Areal data generated from point processes

- Areal data can often be envisioned as arising from point-processes
- Let A be an area of interest and $Y(s) = 1$ or 0 is a binary process indicating whether an outcome has occurred or not at location s
- One can conceptualize a "rate" for that region:

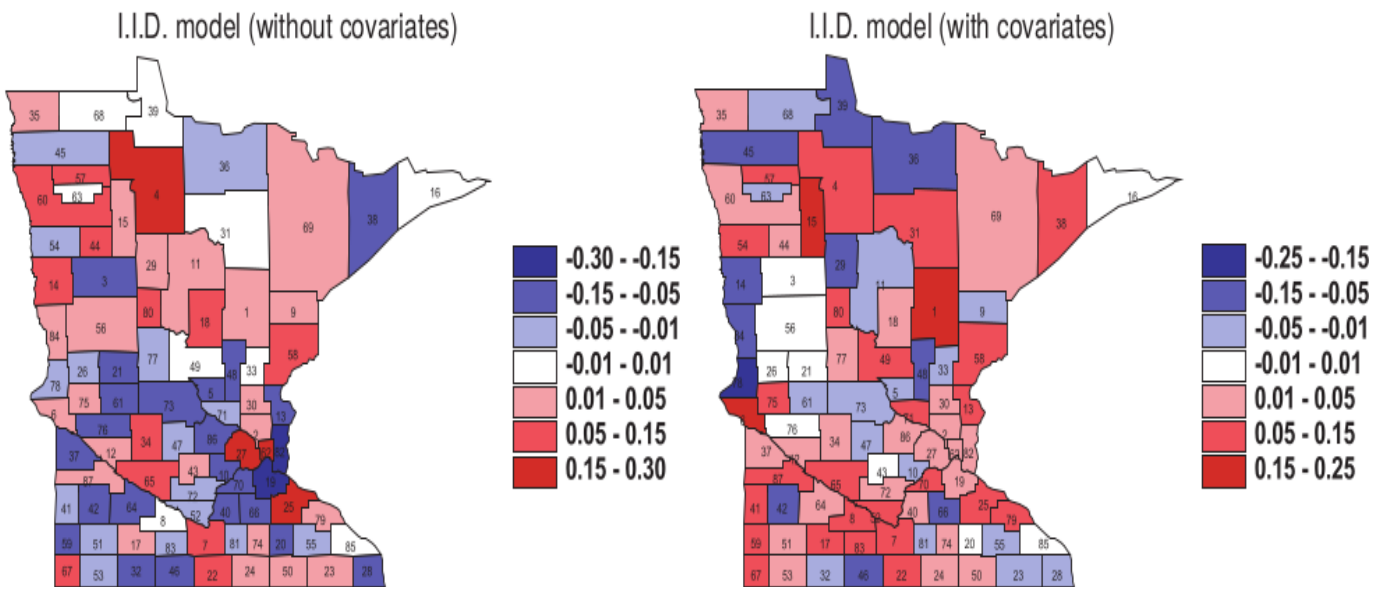
$$N(A) \approx r(A)\mu(A) \approx \int_{s \in A} Y(s)ds \ ,$$

  where $\mu(A)$ is some measure of region A (e.g., geographic area, population etc.) and $N(A)$ may represent an approximation of the "number" of cases in region A.
- So why not model the point process $\{(s, Y(s)) : s \in D\}$ and use that to model $N(A)$?
- Attractive but computationally much more difficult; also, at least in public health it is rare to have access to point-level data.

Disease Mapping: Mapping Random Effects

Model-based estimates of random effects across 87 counties in Minnesota

$$[\text{Infant Mortality Rates}] = [\text{Intercept}] + [\text{Fixed Effects}] + [\text{County-wise Random Effects}]$$



Key Issues

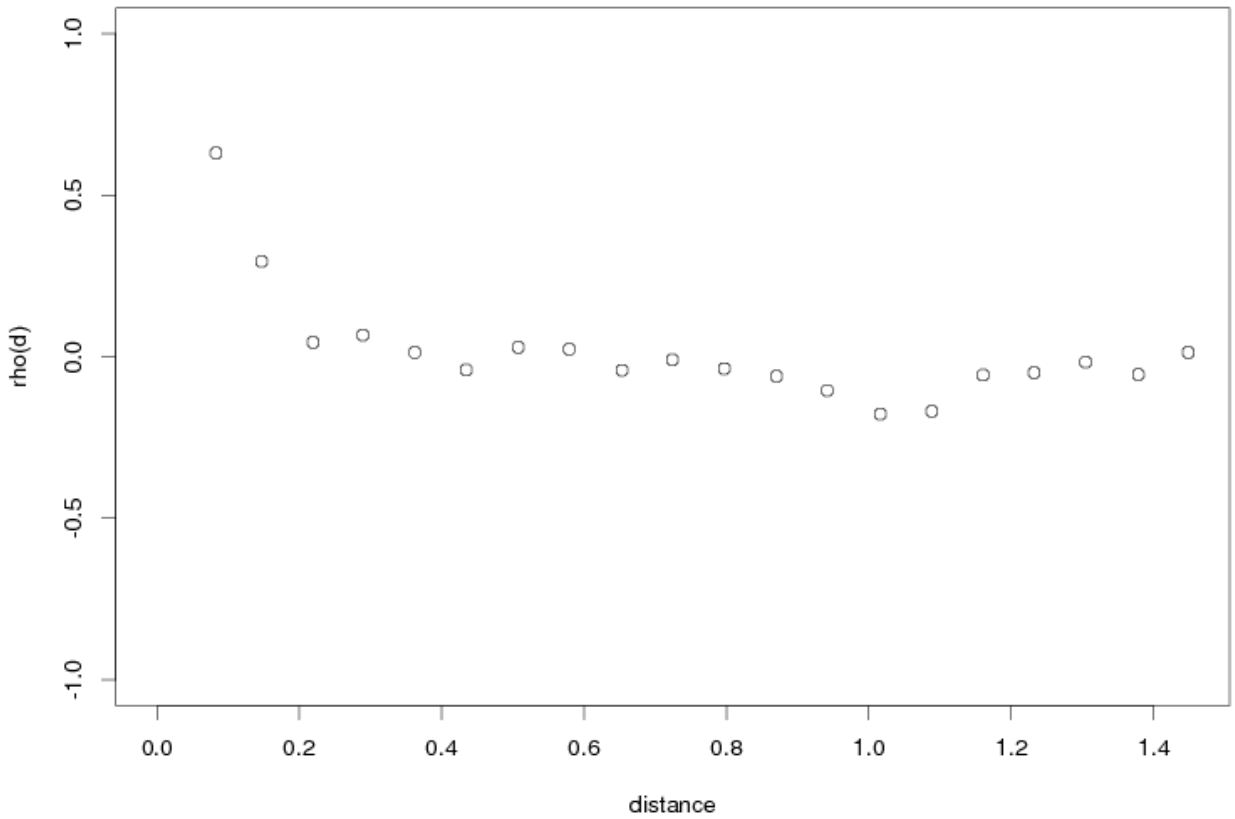
- Is there *spatial* pattern? *Spatial pattern* implies that observations from units closer to each other are more similar than those recorded in units farther away.
- Do we want to *smooth* the data? Perhaps to adjust for low population sizes (or sample sizes) in certain units? How much do we want to smooth?
- Simple IID random effects may not be enough.
- Inferential aims are usually explanatory rather than predictive. Often a hypothesis-generating exercise for epidemiologists.
- Is prediction meaningful here? Inference for *new* areal units? Modifiable Areal Unit Problem (MAUP) or Misalignment.

Toward spatial dependence: Adjacency or proximity matrices

- A, entries a_{ij}, ($a_{ii} = 0$); choices for a_{ij}:
  - $a_{ij} = 1$ if i, j share a common boundary (possibly a common vertex)
  - a_{ij} is an *inverse* distance between units
  - $a_{ij} = 1$ if distance between units is $\leq K$
  - $a_{ij} = 1$ for m nearest neighbors.
- A need not be symmetric.
- $\widetilde{A}$: standardize row i by $a_{i+} = \sum_j a_{ij}$ (row stochastic but need not be symmetric).
- A elements often called "weights"; perhaps nicer interpretation?

- Note that proximity matrices are user-defined.
- We can define distance intervals, $(0, d_1]$, $(d_1, d_2]$, and so on.
  - First order neighbours: all units within distance d_1.
  - First order proximity matrix $A^{(1)}$. Analogous to A, $a_{ij}^{(1)} = 1$ if i and j are first order neighbors; 0 otherwise.
  - Second order neighbors: all units within distance d_2, but separated by more than d_1.
  - Second order proximity matrix $A^{(2)}$; $a_{ij}^{(2)} = 1$ if i and j are second order neighbors; 0 otherwise
  - And so on...

- The *areal correlogram* is a useful tool to study spatial association with areal data.
- Working with I, we can replace a_{ij} with $a_{ij}^{(1)}$ taken from $A^{(1)}$ and compute $\rightarrow I^{(1)}$
- Next replace a_{ij} with $a_{ij}^{(2)}$ taken from $A^{(2)}$ and compute $\rightarrow I^{(2)}$, etc.
- Plot $I^{(r)}$ vs. r
- If there is spatial pattern, we expect $I^{(r)}$ to decline in r initially and then vary about 0.


rho(d)
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
distance

Simple hypothesis tests for spatial autocorrelation

- Y_i is the outcome associated with region i over a map with adjacency matrix A.
- Moran's I: analogue of lagged autocorrelation

$$I = \frac{n \sum_i \sum_j a_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i \neq j} a_{ij})(\sum_i (Y_i - \bar{Y})^2}$$

  I is not supported on $[-1, 1]$.

- Geary's C: analogue of Durbin-Watson statistic

$$C = \frac{(n-1) \sum_i \sum_j a_{ij}(Y_i - Y_j)^2}{\sum_{i \neq j} a_{ij}) \sum_i (Y_i - \bar{Y})^2}$$

- Both are asymptotically normal if Y_i are i.i.d., the first with mean $-1/(n-1)$ and the second with mean 1.
- `spdep (CRAN)`: Significance testing using Monte Carlo or permutation tests

Spatial smoothers

- To smooth Y_i, replace with $\hat{Y}_i = \dfrac{\sum_i a_{ij} Y_j}{a_{i+}}$ Note: K-nearest neighbours (KNN) regression falls within this framework.
- More generally,
$$(1-\alpha)Y_i + \alpha\hat{Y}_i$$
Linear (convex) combination, shrinkage
- Model-based smoothing, e.g.,
$E(Y_i|\{Y_j,\ j=1,2,...,n\})$
- Let $Y = (y_1, y_2, ..., y_n)$ and consider the collection $\{p(y_i \mid y_j,\ j \neq i)\}$
- Does $\{p(y_i \mid y_j, j \neq i)\}$ determine $p(y_1, y_2, ...y_n)$?

CAR models

- CAR model:

$$w_i \,|\, w_{-i} \sim N\left(\frac{\rho}{n_i}\sum_{j \,|\, i\sim j} w_j, \tau_w n_i\right)$$

- Apply Brook's Lemma to obtain joint density for w
- $w = (w_1, w_2, \ldots, w_k)^\top \sim N(0, \tau_w(D - \rho A))$ where $D = \text{diag}(n_1, n_2, \ldots, n_k)$
- $\rho = 1 \Rightarrow$ Improper distribution as $(D - A)1 = 0$ (ICAR)
  - Can be still used as a prior for random effects
  - Cannot be used directly as a data generating model

CAR models

- CAR model:

$$w_i \,|\, w_{-i} \sim N\left(\frac{\rho}{n_i}\sum_{j \,|\, i\sim j} w_j, \tau_w n_i\right)$$

- Apply Brook's Lemma to obtain joint density for w
- $w = (w_1, w_2, \ldots, w_k)^\top \sim N(0, \tau_w(D - \rho A))$ where $D = \text{diag}(n_1, n_2, \ldots, n_k)$
- $\rho = 1 \Rightarrow$ Improper distribution as $(D - A)1 = 0$ (ICAR)
  - Can be still used as a prior for random effects
  - Cannot be used directly as a data generating model
- $\rho < 1 \Rightarrow$ Proper distribution with added parameter flexibility

Hierarchical GLM for Spatial disease mapping

- At unit (region) i, we observe response y_i and covariate x_i
- $g(E(y_i)) = x_i^\top \beta + w_i$ where $g(\cdot)$ denotes a suitable link function

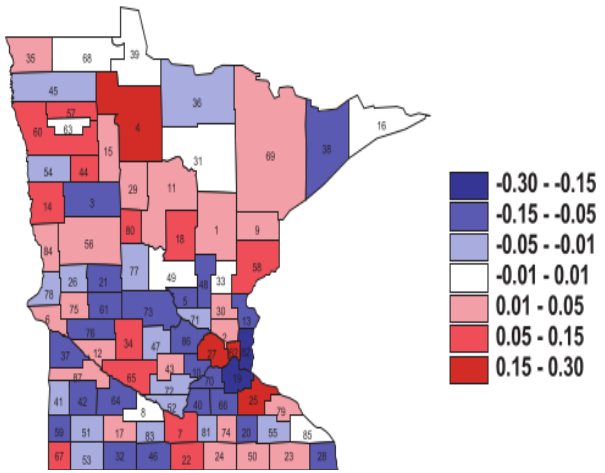
$$p_2(\beta, \tau_w, \rho) \times N(w \,|\, 0, \tau_w(D - \rho A)) \times \prod_{i=1}^{n} p_1(y_i \,|\, x_i^\top \beta + w_i)$$

- p_1 denotes the density corresponding to the link $g(\cdot)$

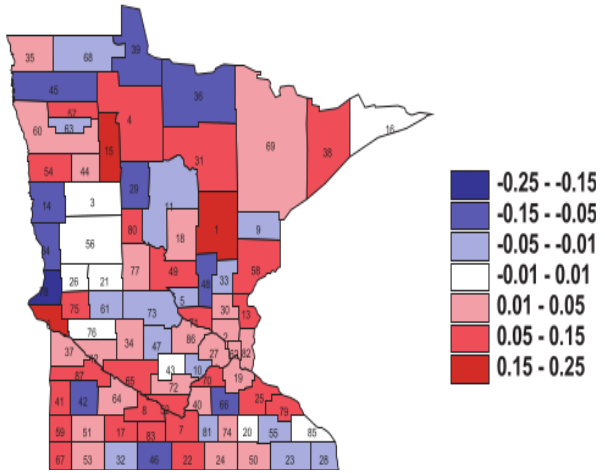
Disease Mapping: Mapping Random Effects

[Infant Mortality Rates] = [Intercept] + [Fixed Effects] + [County-wise Random Effects]

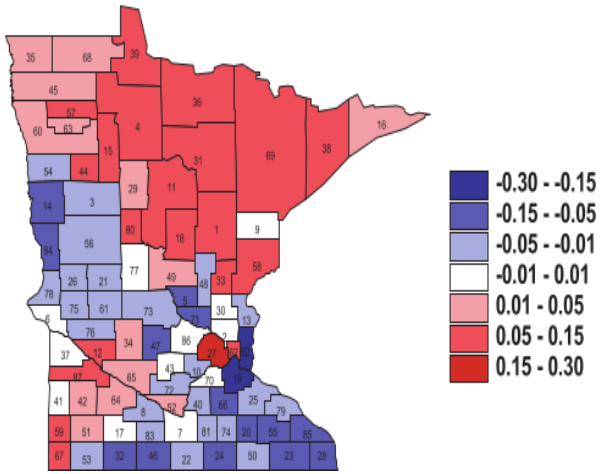
I.I.D. model (without covariates)



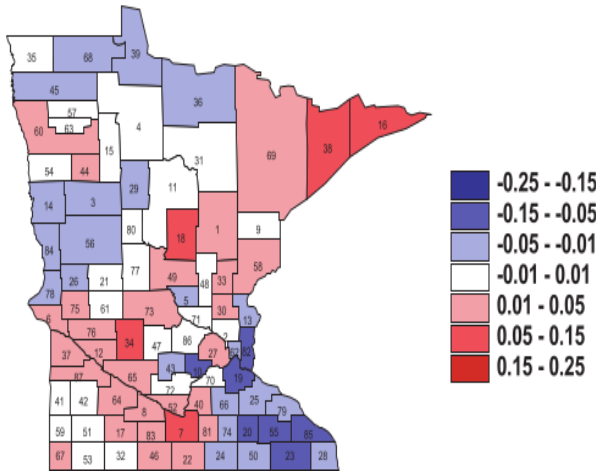
I.I.D. model (with covariates)



CAR model (without covariates)



CAR model (with covariates)



Simultaneous Auto-Regressive (SAR) models

- We may write the CAR model as:

$$y = By + \epsilon \Rightarrow (I - B)y = \epsilon;$$

Since $y \sim N(0, (I - B)^{-1}D)$, we have

$$\epsilon \sim N(0, D(I - B)^{\top}).$$

- Instead of letting y induce the distribution of ϵ, let ϵ induce a distribution for y. Letting $\epsilon \sim N(0, \tilde{D})$, where $\tilde{D}$ is diagonal, $\tilde{D}_{ii} = \sigma_i^2$ and let:

$$y_i = \sum_{j=1}^{n} b_{ij} y_j + \epsilon_i.$$

Assuming $(I - B)^{-1}$ exists, we obtain:

$$y \sim N\left(0, (I - B)^{-1}\tilde{D}(I - B)^{\top -1}\right).$$

SAR models (contd.)

- Often we take $B = \rho A$. If $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)}$ and $\lambda_{(n)}$ are the minimum and maximum eigenvalues of A. This ensures $(I - \rho A)^{-1}$ exists.
- Alternatively, we can replace A with $\tilde{A} = \{a_{ij}/a_{i+}\}$ where a_{i+} is the sum of the elements in the i-th row of A. Then $|\rho| < 1$ ensures existence of $(I - \rho\tilde{A})^{-1}$.
- Often SAR models are also applied to point-referenced data where A is taken to be the inter-point distance.

Variants of SAR models

- Two variants:
  - The SAR "lag model":

$$y = By + X\beta + \epsilon.$$

  - The SAR "residual" or "error model":

$$(I - B)(y - X\beta) = \epsilon; \Rightarrow y = By + (I - B)X\beta + \epsilon.$$

- SAR models are well suited to maximum likelihood estimation but not for MCMC fitting of Bayesian models. Because it is difficult to introduce SAR random effects (in the CAR framework this is easy because of the hierarchical conditional representation).