# Areal Unit Data

## Regular or Irregular Grids or Lattices
## Large Point-referenced Datasets

- Is there *spatial* pattern?

# Areal Unit Data

## Regular or Irregular Grids or Lattices
## Large Point-referenced Datasets

- Is there *spatial* pattern?
- Do we want to smooth the data?

# Areal Unit Data

## Regular or Irregular Grids or Lattices
## Large Point-referenced Datasets

- Is there *spatial* pattern?
- Do we want to smooth the data?
- Inference for new areal units?

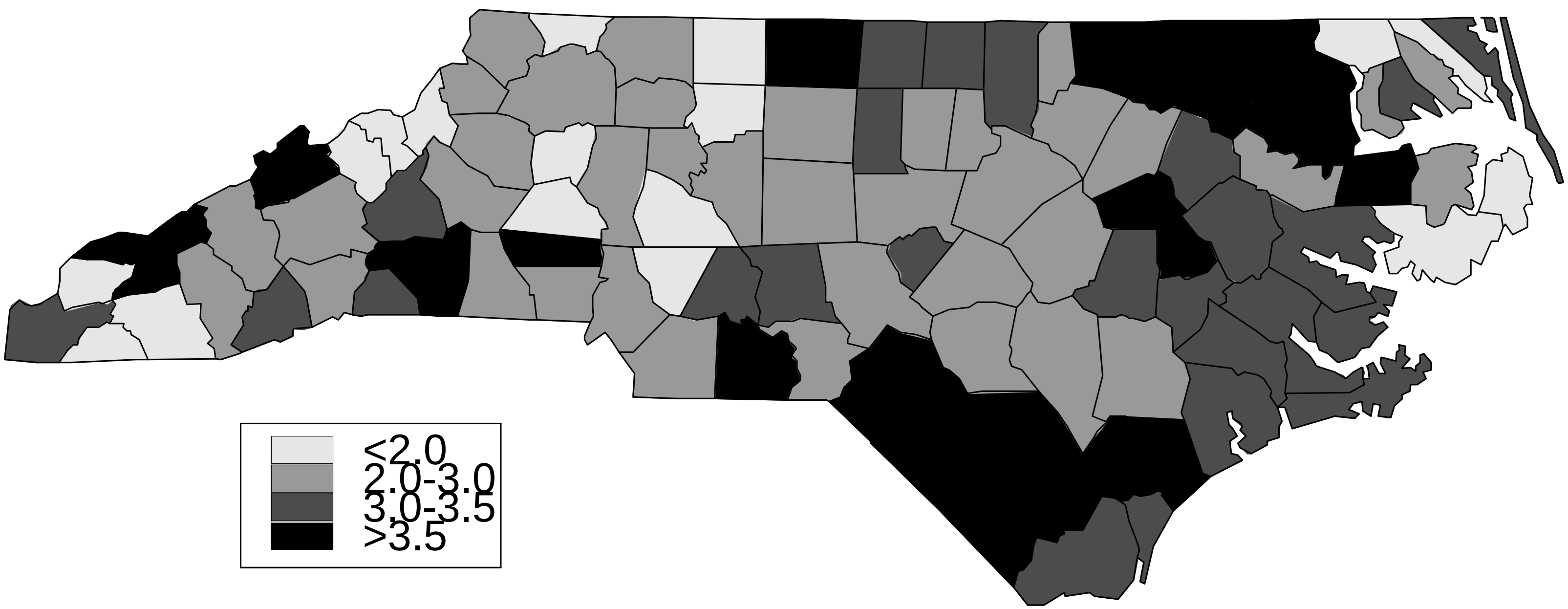
# Areal Unit Data

## Regular or Irregular Grids or Lattices
## Large Point-referenced Datasets

- Is there *spatial* pattern?
- Do we want to smooth the data?
- Inference for new areal units?
- Descriptive/algorithmic vs. Model-based

# Areal unit data

## Actual Transformed SIDS Rates



# Proximity matrices

- $W$, entries $w_{ij}$ (with $w_{ii} = 0$). Choices for $w_{ij}$:
  - $w_{ij} = 1$ if $i, j$ share a common boundary (possibly a common vertex)
  - $w_{ij}$ is an *inverse* distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for $m$ nearest neighbors.

# Proximity matrices

- $W$, entries $w_{ij}$ (with $w_{ii} = 0$). Choices for $w_{ij}$:
  - $w_{ij} = 1$ if $i, j$ share a common boundary (possibly a common vertex)
  - $w_{ij}$ is an *inverse* distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for $m$ nearest neighbors.
- $W$ is typically symmetric, but need not be

# Proximity matrices

- $W$, entries $w_{ij}$ (with $w_{ii} = 0$). Choices for $w_{ij}$:
  - $w_{ij} = 1$ if $i, j$ share a common boundary (possibly a common vertex)
  - $w_{ij}$ is an *inverse* distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for $m$ nearest neighbors.
- $W$ is typically symmetric, but need not be
- $\widetilde{W}$: standardize row $i$ by $w_{i+} = \sum_j w_{ij}$ (so matrix is now row stochastic, but probably no longer symmetric).

# Proximity matrices

- $W$, entries $w_{ij}$ (with $w_{ii} = 0$). Choices for $w_{ij}$:
  - $w_{ij} = 1$ if $i, j$ share a common boundary (possibly a common vertex)
  - $w_{ij}$ is an *inverse* distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for $m$ nearest neighbors.
- $W$ is typically symmetric, but need not be
- $\widetilde{W}$: standardize row $i$ by $w_{i+} = \sum_j w_{ij}$ (so matrix is now row stochastic, but probably no longer symmetric).
- $W$ elements often called "weights"; interpretation

# Proximity matrices

- $W$, entries $w_{ij}$ (with $w_{ii} = 0$). Choices for $w_{ij}$:
  - $w_{ij} = 1$ if $i, j$ share a common boundary (possibly a common vertex)
  - $w_{ij}$ is an *inverse* distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for $m$ nearest neighbors.
- $W$ is typically symmetric, but need not be
- $\widetilde{W}$: standardize row $i$ by $w_{i+} = \sum_j w_{ij}$ (so matrix is now row stochastic, but probably no longer symmetric).
- $W$ elements often called "weights"; interpretation
- Could also define first-order neighbors $W^{(1)}$, second-order neighbors $W^{(2)}$, etc.

# Measures of spatial association

- Moran's $I$: essentially an "areal covariogram"

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

# Measures of spatial association

- Moran's $I$: essentially an “areal covariogram"

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

- Geary's $C$: essentially an “areal variogram"

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

# Measures of spatial association

- Moran's $I$: essentially an "areal covariogram"

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

- Geary's $C$: essentially an "areal variogram"

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

- Both are asymptotically normal if $Y_i$ are i.i.d.;
  Moran has mean $-1/(n-1) \approx 0$, Geary has mean 1

# Measures of spatial association

- Moran's $I$: essentially an "areal covariogram"

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

- Geary's $C$: essentially an "areal variogram"

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

- Both are asymptotically normal if $Y_i$ are i.i.d.; Moran has mean $-1/(n-1) \approx 0$, Geary has mean 1
- Significance testing by comparing to a collection of say 1000 random permutations of the $Y_i$

# Measures of spatial association (cont'd)

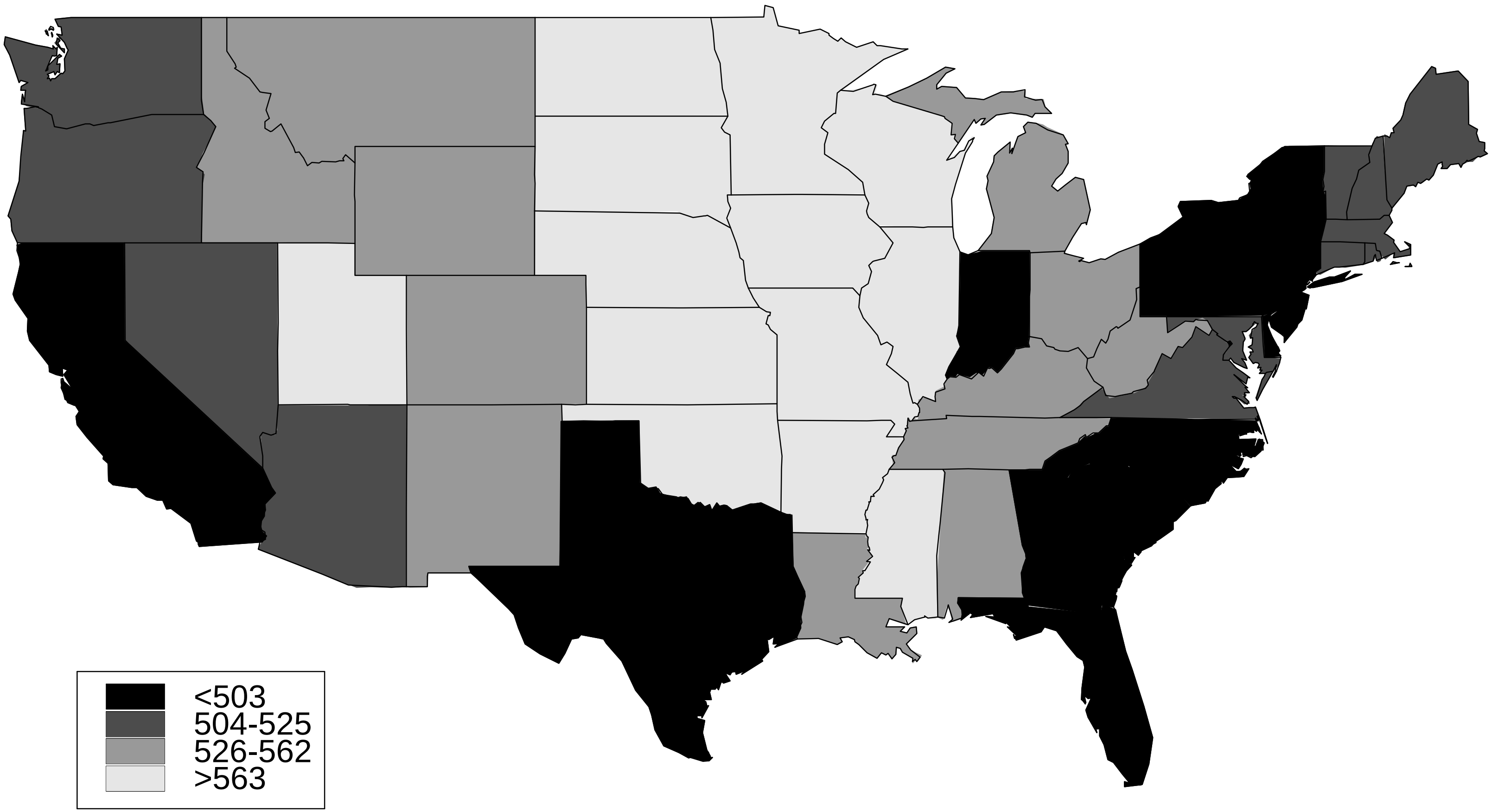



Figure 1: Choropleth map of 1999 average verbal SAT scores, lower 48 U.S. states.

# Measures of spatial association (cont'd)

- For these data, the `spatial.cor` function in `S+SpatialStats` gives a Moran's $I$ of 0.5833, with associated standard error estimate 0.0920 $\Rightarrow$ very strong evidence against $H_0$ : no spatial correlation

# Measures of spatial association (cont'd)

- For these data, the `spatial.cor` function in `S+SpatialStats` gives a Moran's $I$ of 0.5833, with associated standard error estimate 0.0920 $\Rightarrow$ very strong evidence against $H_0$ : no spatial correlation
- `spatial.cor` also gives a Geary's $C$ of 0.3775, with associated standard error estimate 0.1008 $\Rightarrow$ again, very strong evidence against $H_0$ (departure from 1)

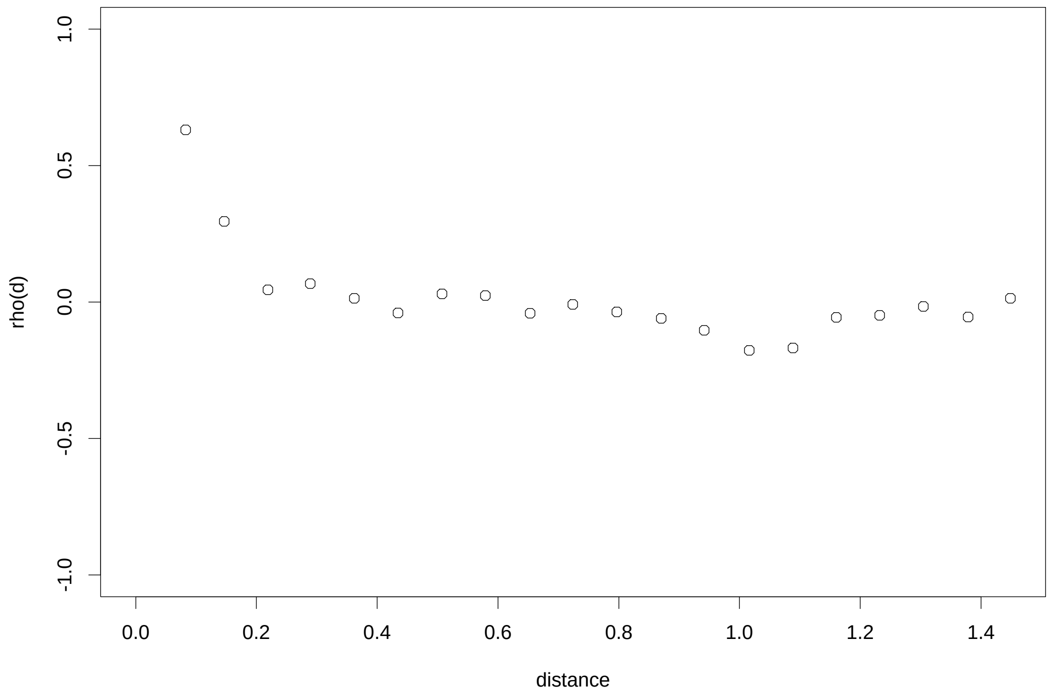
# Measures of spatial association (cont'd)

- For these data, the `spatial.cor` function in `S+SpatialStats` gives a Moran's $I$ of 0.5833, with associated standard error estimate 0.0920 $\Rightarrow$ very strong evidence against $H_0$ : no spatial correlation
- `spatial.cor` also gives a Geary's $C$ of 0.3775, with associated standard error estimate 0.1008 $\Rightarrow$ again, very strong evidence against $H_0$ (departure from 1)
- Warning: These data have not been adjusted for covariates, such as the proportion of students who take the exam (Midwestern colleges have historically relied on the ACT, not the SAT; only the best and brightest students in these states would bother taking the SAT)

# Measures of spatial association (cont'd)

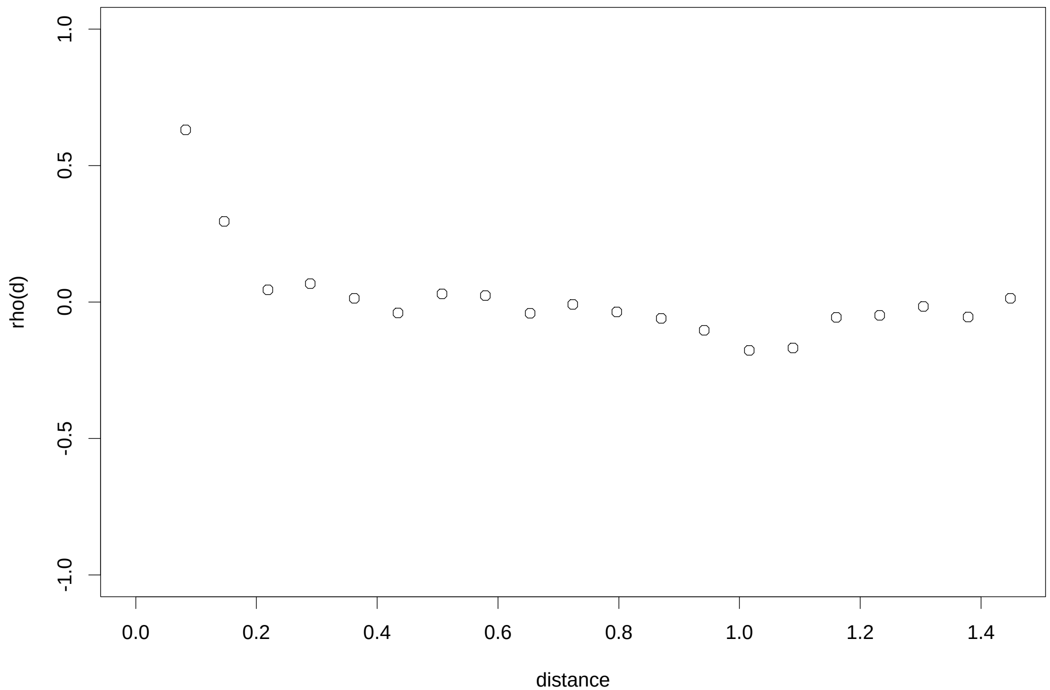
- For these data, the `spatial.cor` function in `S+SpatialStats` gives a Moran's $I$ of 0.5833, with associated standard error estimate 0.0920 $\Rightarrow$ very strong evidence against $H_0$ : no spatial correlation
- `spatial.cor` also gives a Geary's $C$ of 0.3775, with associated standard error estimate 0.1008 $\Rightarrow$ again, very strong evidence against $H_0$ (departure from 1)
- Warning: These data have not been adjusted for covariates, such as the proportion of students who take the exam (Midwestern colleges have historically relied on the ACT, not the SAT; only the best and brightest students in these states would bother taking the SAT)
- $\Rightarrow$ the map, $I$, and $C$ all motivate the search for spatial covariates!

# Correlogram (via Moran's $I$)



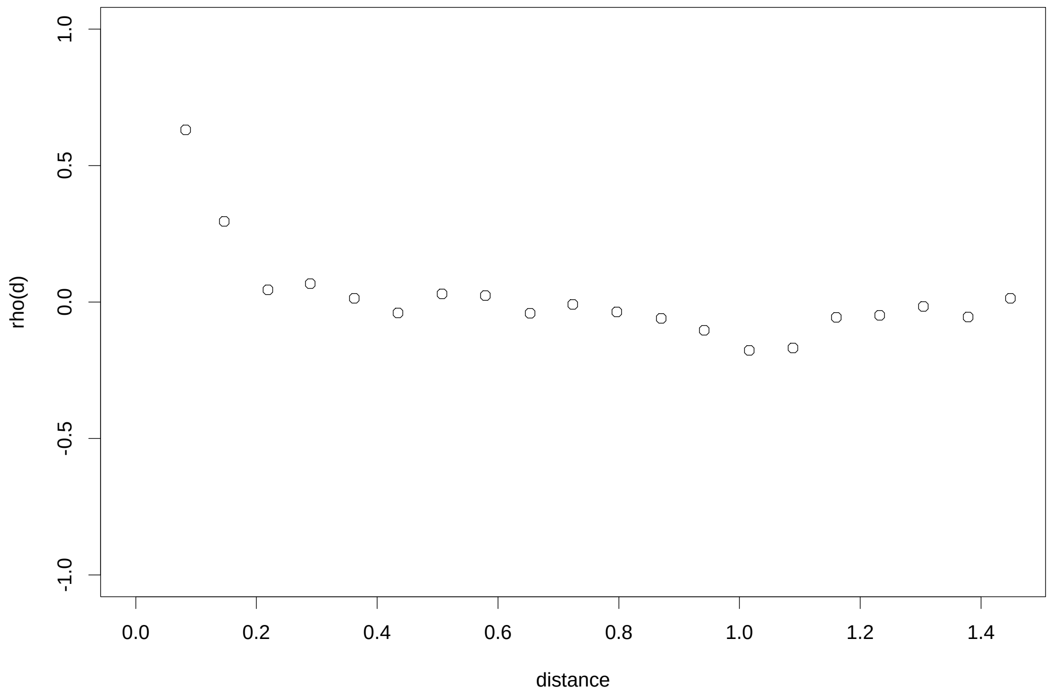

- Replace $w_{ij}$ with $w_{ij}^{(1)}$ taken from $W^{(1)} \Rightarrow I^{(1)}$

# Correlogram (via Moran's $I$)



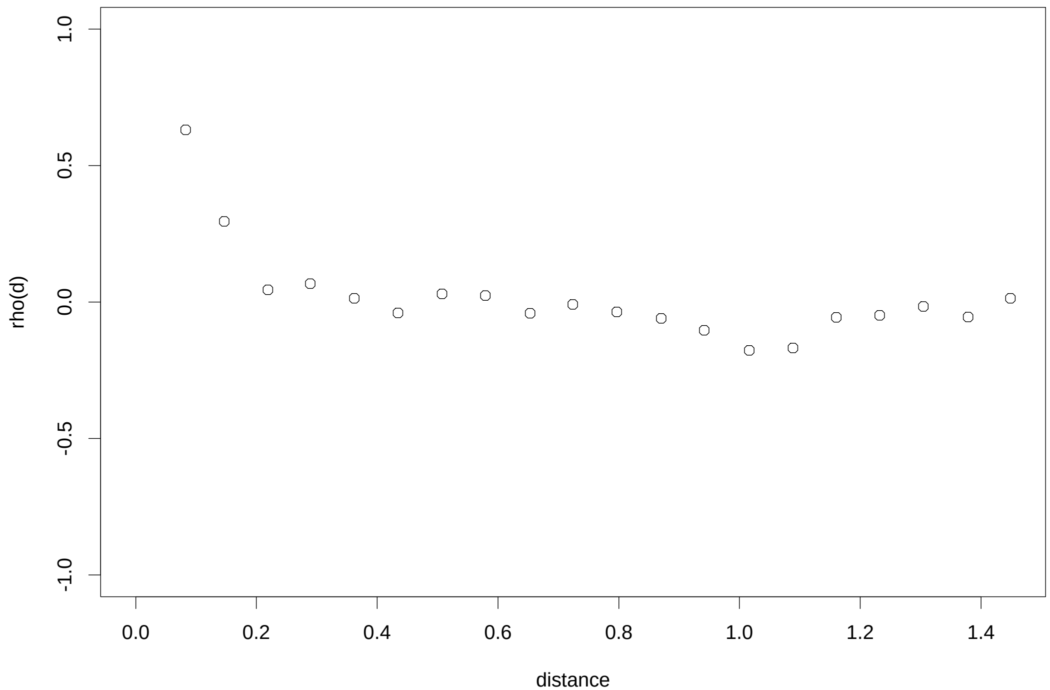

- Replace $w_{ij}$ with $w_{ij}^{(1)}$ taken from $W^{(1)} \Rightarrow I^{(1)}$
- Replace $w_{ij}$ with $w_{ij}^{(2)}$ taken from $W^{(2)} \Rightarrow I^{(2)}$, etc.

# Correlogram (via Moran's $I$)



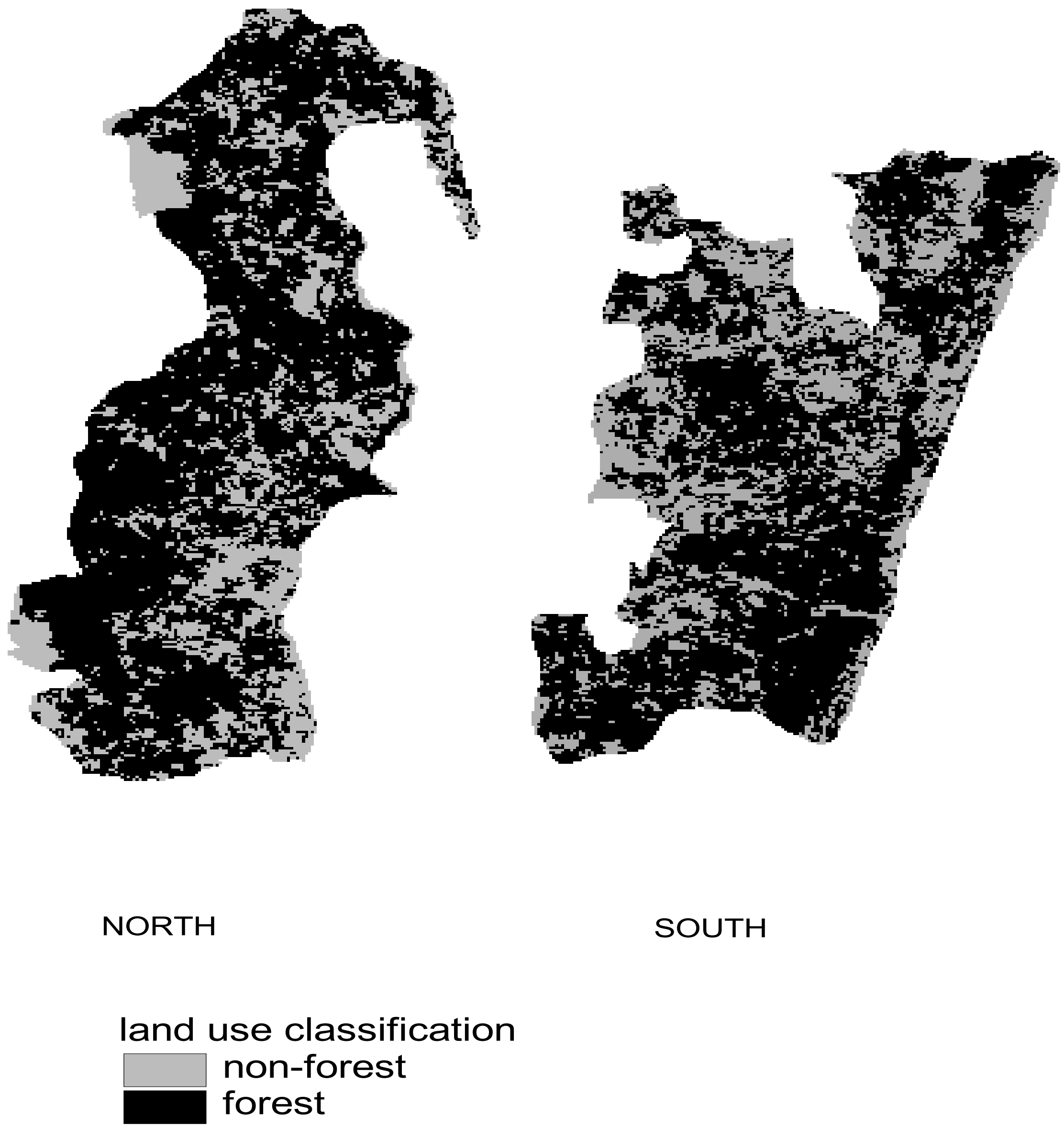

- Replace $w_{ij}$ with $w_{ij}^{(1)}$ taken from $W^{(1)} \Rightarrow I^{(1)}$
- Replace $w_{ij}$ with $w_{ij}^{(2)}$ taken from $W^{(2)} \Rightarrow I^{(2)}$, etc.
- Plot $I^{(r)}$ vs. $r$; an initial decline in $r$ followed by variation around 0 $\Rightarrow$ spatial pattern!

# Correlogram (via Moran's $I$)

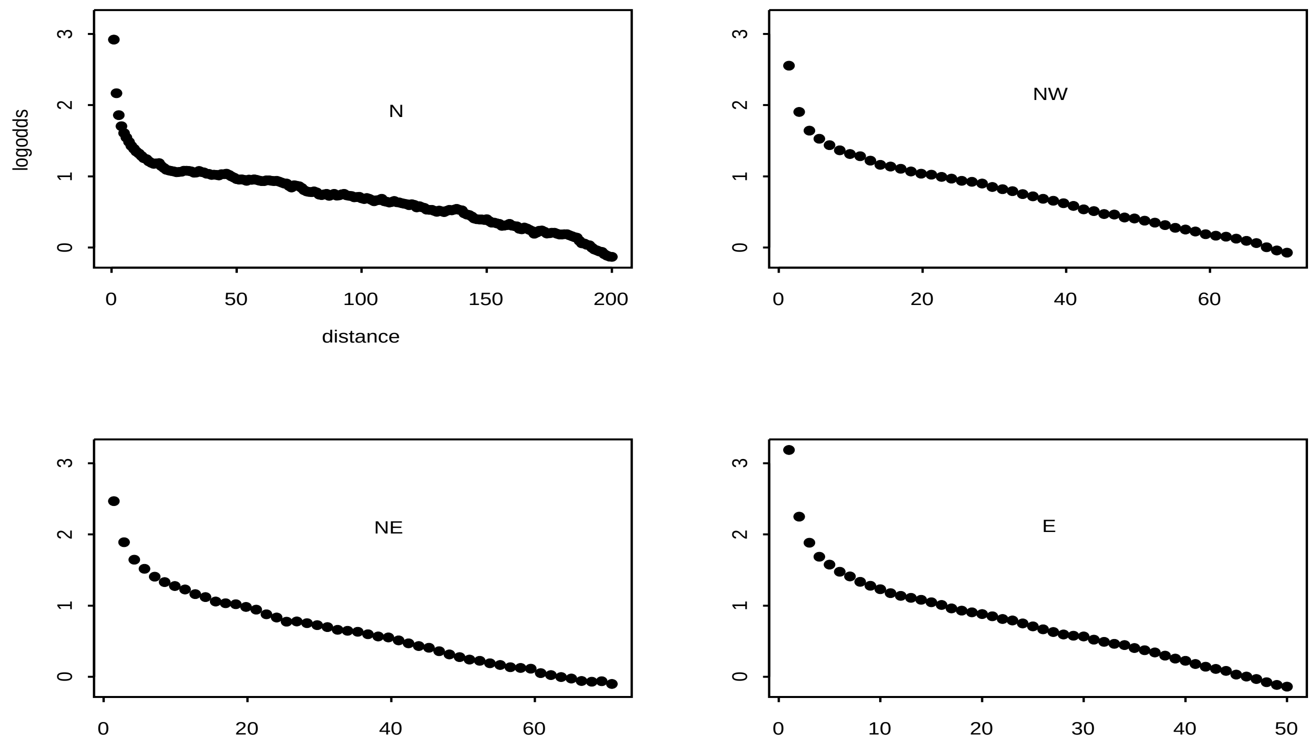



- Replace $w_{ij}$ with $w_{ij}^{(1)}$ taken from $W^{(1)} \Rightarrow I^{(1)}$
- Replace $w_{ij}$ with $w_{ij}^{(2)}$ taken from $W^{(2)} \Rightarrow I^{(2)}$, etc.
- Plot $I^{(r)}$ vs. $r$; an initial decline in $r$ followed by variation around 0 $\Rightarrow$ spatial pattern!
- spatial analogue of the temporal lag autocorrelation plot

# Rasterized binary data map

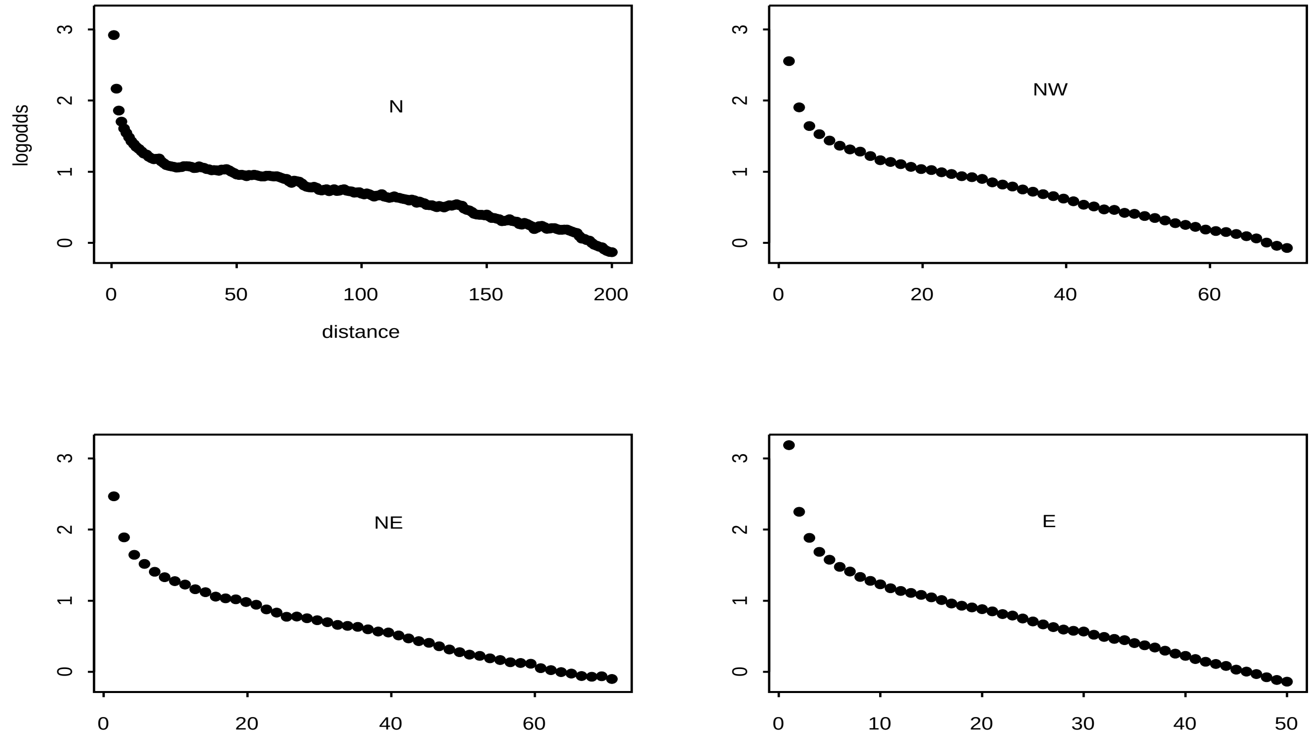


# Binary data correlogram




- A version for a binary map, using two-way tables and log odds ratios at the pixel level

# Binary data correlogram




- A version for a binary map, using two-way tables and log odds ratios at the pixel level
- Note strongest pattern is to the north (N), but in no direction are the values $\approx$ 0 even at 40 km

# Spatial smoothers

- To smooth $Y_i$, replace with $\hat{Y}_i = \frac{\sum_i w_{ij} Y_j}{w_{i+}}$

# Spatial smoothers

- To smooth $Y_i$, replace with $\hat{Y}_i = \frac{\sum_i w_{ij} Y_j}{w_{i+}}$
- More generally, we could include the value actually observed for unit $i$, and revise our smoother to

$$(1-\alpha)Y_i + \alpha\hat{Y}_i$$

For $0 < \alpha < 1$, this is a linear (convex) combination in “shrinkage" form

# Spatial smoothers

- To smooth $Y_i$, replace with $\hat{Y}_i = \frac{\sum_i w_{ij} Y_j}{w_{i+}}$
- More generally, we could include the value actually observed for unit $i$, and revise our smoother to

$$(1-\alpha)Y_i + \alpha\hat{Y}_i$$

For $0 < \alpha < 1$, this is a linear (convex) combination in “shrinkage" form

- Finally, we could try model-based smoothing, i.e., based on $E(Y_i|Data)$, i.e., the mean of the predictive distribution. Smoothers then emerge as byproducts of the hierarchical spatial models we use to explain the $Y_i$'s

# Markov random fields

- Consider $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ and the set $\{p(y_i|y_j,\ j \neq i)\}$

# Markov random fields

- Consider $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ and the set $\{p(y_i|y_j,\ j \neq i)\}$
- We know $p(y_1, y_2, ...y_n)$ determines $\{p(y_i|y_j, j \neq i)\}$ (the set of full conditional distributions)

# Markov random fields

- Consider $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ and the set $\{p(y_i|y_j,\ j \neq i)\}$
- We know $p(y_1, y_2, ...y_n)$ determines $\{p(y_i|y_j, j \neq i)\}$ (the set of full conditional distributions)
- Does $\{p(y_i|y_j, j \neq i)\}$ determine $p(y_1, y_2, ...y_n)$???

# Markov random fields

- Consider $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ and the set $\{p(y_i|y_j,\ j \neq i)\}$
- We know $p(y_1, y_2, ...y_n)$ determines $\{p(y_i|y_j, j \neq i)\}$ (the set of full conditional distributions)
- Does $\{p(y_i|y_j, j \neq i)\}$ determine $p(y_1, y_2, ...y_n)$???
- If the full conditionals are compatible, then by Brook's Lemma,

$$p(y_1, \ldots, y_n) = \frac{p(y_1|y_2, \ldots, y_n)}{p(y_{10}|y_2, \ldots, y_n)} \frac{p(y_2|y_{10}, y_3, \ldots, y_n)}{p(y_{20}|y_{10}, y_3, \ldots, y_n)} \cdots \frac{p(y_n|y_{10}, \ldots, y_{n-1,0})}{p(y_{n0}|y_{10}, \ldots, y_{n-1,0})} p(y_{10}, \ldots, y_{n0})$$

If left side is proper, the fact that it integrates to 1 determines the normalizing constant!

# “Local” modeling

- Suppose we specify the full conditionals such that

$$p(y_i|y_j, j \neq i) = p(y_i|y_j \in \partial_i) \ ,$$

where $\partial_i$ is the set of neighbors of cell (region) $i$.
When does $\{p(y_i|y_j \in \partial_i)\}$ determine $p(y_1, y_2, ...y_n)$?

# “Local” modeling

- Suppose we specify the full conditionals such that

$$p(y_i|y_j, j \neq i) = p(y_i|y_j \in \partial_i) \ ,$$

  where $\partial_i$ is the set of neighbors of cell (region) $i$.
  When does $\{p(y_i|y_j \in \partial_i)\}$ determine $p(y_1, y_2, ...y_n)$?

- Def’n: a clique is a set of cells such that each element is a neighbor of every other element

# "Local" modeling

- Suppose we specify the full conditionals such that

$$p(y_i | y_j, j \neq i) = p(y_i | y_j \in \partial_i) \; ,$$

where $\partial_i$ is the set of neighbors of cell (region) $i$.
When does $\{p(y_i | y_j \in \partial_i)\}$ determine $p(y_1, y_2, ...y_n)$?

- Def'n: a clique is a set of cells such that each element is a neighbor of every other element
- Def'n: a potential function of order $k$ is a function of $k$ arguments that is exchangeable in these arguments

# “Local” modeling

- Suppose we specify the full conditionals such that

$$p(y_i|y_j, j \neq i) = p(y_i|y_j \in \partial_i) \ ,$$

where $\partial_i$ is the set of neighbors of cell (region) $i$.
When does $\{p(y_i|y_j \in \partial_i)\}$ determine $p(y_1, y_2, ...y_n)$?

- Def’n: a clique is a set of cells such that each element is a neighbor of every other element
- Def’n: a potential function of order $k$ is a function of $k$ arguments that is exchangeable in these arguments
- Def’n: $p(y_1, \ldots, y_n)$ is a Gibbs distribution if, as a function of the $y_i$, it is a product of potentials on cliques

# "Local" modeling

- Suppose we specify the full conditionals such that

$$p(y_i|y_j, j \neq i) = p(y_i|y_j \in \partial_i) \ ,$$

  where $\partial_i$ is the set of neighbors of cell (region) $i$.
  When does $\{p(y_i|y_j \in \partial_i)\}$ determine $p(y_1, y_2, ...y_n)$?
- Def'n: a clique is a set of cells such that each element is a neighbor of every other element
- Def'n: a potential function of order $k$ is a function of $k$ arguments that is exchangeable in these arguments
- Def'n: $p(y_1, \ldots, y_n)$ is a Gibbs distribution if, as a function of the $y_i$, it is a product of potentials on cliques
- Example: For binary data, $k = 2$, we might take $Q(y_i, y_j) = I(y_i = y_j) = y_i y_j + (1 - y_i)(1 - y_j)$

# “Local” modeling

- Cliques of size 1 $\Leftrightarrow$ independence

# “Local” modeling

- Cliques of size 1 $\Leftrightarrow$ independence
- Cliques of size 2 $\Leftrightarrow$ pairwise difference form

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i,j}(y_i - y_j)^2 I(i \sim j)\right\}$$

and therefore $p(y_i|y_j, j \neq i) = N(\sum_{j\in\partial_i} y_j/m_i, \tau^2/m_i)$, where $m_i$ is the number of neighbors of $i$

# “Local” modeling

- Cliques of size 1 $\Leftrightarrow$ independence
- Cliques of size 2 $\Leftrightarrow$ pairwise difference form

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i,j}(y_i - y_j)^2 I(i \sim j)\right\}$$

and therefore $p(y_i|y_j, j \neq i) = N(\sum_{j\in\partial_i} y_j/m_i, \tau^2/m_i)$, where $m_i$ is the number of neighbors of $i$

- Hammersley-Clifford Theorem: If we have a Markov Random Field (i.e., $\{p(y_i|y_j \in \partial_i)\}$ uniquely determine $p(y_1, y_2, ...y_n)$), then the latter is a Gibbs distribution

# “Local” modeling

- Cliques of size 1 $\Leftrightarrow$ independence
- Cliques of size 2 $\Leftrightarrow$ pairwise difference form

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i,j}(y_i - y_j)^2 I(i \sim j)\right\}$$

  and therefore $p(y_i|y_j, j \neq i) = N(\sum_{j\in\partial_i} y_j/m_i, \tau^2/m_i)$, where $m_i$ is the number of neighbors of $i$

- Hammersley-Clifford Theorem: If we have a Markov Random Field (i.e., $\{p(y_i|y_j \in \partial_i)\}$ uniquely determine $p(y_1, y_2, ...y_n)$), then the latter is a Gibbs distribution
- Geman and Geman result : If we have a joint Gibbs distribution, then we have a Markov Random Field

# Conditional autoregressive (CAR) model

- Gaussian (autonormal) case

$$p(y_i|y_j, j \neq i) = N\left(\sum_j b_{ij} y_j, \tau_i^2\right)$$

Using Brook's Lemma we can obtain

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2}\mathbf{y}' D^{-1}(I - B)\mathbf{y}\right\}$$

where $B = \{b_{ij}\}$ and $D$ is diagonal with $D_{ii} = \tau_i^2$.

# Conditional autoregressive (CAR) model

- Gaussian (autonormal) case

$$p(y_i|y_j, j \neq i) = N\left(\sum_j b_{ij} y_j, \tau_i^2\right)$$

  Using Brook's Lemma we can obtain

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2}\mathbf{y}' D^{-1}(I - B)\mathbf{y}\right\}$$

  where $B = \{b_{ij}\}$ and $D$ is diagonal with $D_{ii} = \tau_i^2$.

- $\Rightarrow$ suggests a multivariate normal distribution with $\mu_Y = 0$ and $\Sigma_Y = (I - B)^{-1} D$

# Conditional autoregressive (CAR) model

- Gaussian (autonormal) case

$$p(y_i|y_j, j \neq i) = N\left(\sum_j b_{ij}y_j, \tau_i^2\right)$$

Using Brook's Lemma we can obtain

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2}\mathbf{y}'D^{-1}(I - B)\mathbf{y}\right\}$$

where $B = \{b_{ij}\}$ and $D$ is diagonal with $D_{ii} = \tau_i^2$.

- $\Rightarrow$ suggests a multivariate normal distribution with $\mu_Y = 0$ and $\Sigma_Y = (I - B)^{-1}D$
- $D^{-1}(I - B)$ symmetric requires $\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}$ for all $i, j$

# CAR Model (cont'd)

- Returning to $W$, let $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, so

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - W)\mathbf{y}\right\}$$

where $D_w$ is diagonal with $(D_w)_{ii} = w_{i+}$ and thus

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i\neq j} w_{ij}(y_i - y_j)^2\right\}$$

Intrinsic autoregressive (IAR) model!

# CAR Model (cont'd)

- Returning to $W$, let $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, so

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - W)\mathbf{y}\right\}$$

where $D_w$ is diagonal with $(D_w)_{ii} = w_{i+}$ and thus

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i\neq j} w_{ij}(y_i - y_j)^2\right\}$$

Intrinsic autoregressive (IAR) model!

- Joint distribution is improper since $(D_w - W)\mathbf{1} = 0$, so requires a constraint – say, $\sum_i y_i = 0$

# CAR Model (cont'd)

- Returning to $W$, let $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, so

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - W)\mathbf{y}\right\}$$

  where $D_w$ is diagonal with $(D_w)_{ii} = w_{i+}$ and thus

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i \neq j} w_{ij}(y_i - y_j)^2\right\}$$

  Intrinsic autoregressive (IAR) model!

- Joint distribution is improper since $(D_w - W)\mathbf{1} = 0$, so requires a constraint – say, $\sum_i y_i = 0$
- Not a data model – a random effects model!

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of “islands”

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of “islands”
- “Proper version:” replace $D_w - W$ by $D_w - \rho W$, and choose $\rho$ so that $\Sigma_{\mathbf{y}} = (D_w - \rho W)^{-1}$ exists!
  This in turn implies $Y_i | Y_{j \neq i} \sim N(\rho \sum_j w_{ij} Y_j, \tau^2 / m_i)$

---

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of "islands"
- "Proper version:" replace $D_w - W$ by $D_w - \rho W$, and choose $\rho$ so that $\Sigma_{\mathbf{y}} = (D_w - \rho W)^{-1}$ exists!
  This in turn implies $Y_i | Y_{j \neq i} \sim N(\rho \sum_j w_{ij} Y_j, \tau^2 / m_i)$

---

"To $\rho$ or not to $\rho$?" (that is the question)

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of "islands"

- "Proper version:" replace $D_w - W$ by $D_w - \rho W$, and choose $\rho$ so that $\Sigma_{\mathbf{y}} = (D_w - \rho W)^{-1}$ exists!
  This in turn implies $Y_i | Y_{j \neq i} \sim N(\rho \sum_j w_{ij} Y_j, \tau^2 / m_i)$

---

"To $\rho$ or not to $\rho$?" (that is the question)

Advantages:

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of “islands”
- “Proper version:” replace $D_w - W$ by $D_w - \rho W$, and choose $\rho$ so that $\Sigma_{\mathbf{y}} = (D_w - \rho W)^{-1}$ exists! This in turn implies $Y_i | Y_{j \neq i} \sim N(\rho \sum_j w_{ij} Y_j, \tau^2 / m_i)$

---

“To $\rho$ or not to $\rho$?” (that is the question)

Advantages:

- makes distribution proper

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of "islands"
- "Proper version:" replace $D_w - W$ by $D_w - \rho W$, and choose $\rho$ so that $\Sigma_{\mathbf{y}} = (D_w - \rho W)^{-1}$ exists! This in turn implies $Y_i | Y_{j \neq i} \sim N(\rho \sum_j w_{ij} Y_j, \tau^2 / m_i)$

---

"To $\rho$ or not to $\rho$?" (that is the question)

Advantages:

- makes distribution proper
- adds parametric flexibility

# CAR Model Issues

- With $\tau^2$ unknown, appropriate exponent for $\tau^2$? Hodges et al.: $n - I$, where $I$ is the # of "islands"
- "Proper version:" replace $D_w - W$ by $D_w - \rho W$, and choose $\rho$ so that $\Sigma_{\mathbf{y}} = (D_w - \rho W)^{-1}$ exists!
  This in turn implies $Y_i | Y_{j \neq i} \sim N(\rho \sum_j w_{ij} Y_j, \tau^2 / m_i)$

---

"To $\rho$ or not to $\rho$?" (that is the question)

Advantages:

- makes distribution proper
- adds parametric flexibility
- $\rho = 0$ interpretable as independence

# CAR Models with $\rho$ parameter

Disadvantages:

- why should we expect $Y_i$ to be a proportion of average of neighbors – a sensible spatial interpretation?

# CAR Models with $\rho$ parameter

Disadvantages:

- why should we expect $Y_i$ to be a proportion of average of neighbors – a sensible spatial interpretation?
- calibration of $\rho$ as a correlation, e.g.,

$$\rho = 0.80 \text{ yields } 0.1 \leq \text{ Moran's } I \leq 0.15,$$
$$\rho = 0.90 \text{ yields } 0.2 \leq \text{ Moran's } I \leq 0.25,$$
$$\rho = 0.99 \text{ yields Moran's } I \leq 0.5$$

# CAR Models with $\rho$ parameter

Disadvantages:

- why should we expect $Y_i$ to be a proportion of average of neighbors – a sensible spatial interpretation?
- calibration of $\rho$ as a correlation, e.g.,

$$\rho = 0.80 \text{ yields } 0.1 \leq \text{ Moran's } I \leq 0.15,$$
$$\rho = 0.90 \text{ yields } 0.2 \leq \text{ Moran's } I \leq 0.25,$$
$$\rho = 0.99 \text{ yields } \text{ Moran's } I \leq 0.5$$

- So, used with random effects, scope of spatial pattern may be limited

# Comments

- The CAR specifies $\Sigma_Y^{-1}$, not $\Sigma_Y$ (as in the point-level modeling case), so does not directly model association

# Comments

- The CAR specifies $\Sigma_Y^{-1}$, not $\Sigma_Y$ (as in the point-level modeling case), so does not directly model association
- $(\Sigma_Y^{-1})_{ii} = 1/\tau_i^2$; $(\Sigma_Y^{-1})_{ij} = 0 \Leftrightarrow$ cond'l independence

# Comments

- The CAR specifies $\Sigma_Y^{-1}$, not $\Sigma_Y$ (as in the point-level modeling case), so does not directly model association
- $(\Sigma_Y^{-1})_{ii} = 1/\tau_i^2$; $(\Sigma_Y^{-1})_{ij} = 0 \Leftrightarrow$ cond'l independence
- Prediction at new sites is ad hoc, in that if

$$p(y_0|y_1, y_2, ...y_n) = N(\sum_j w_{0j}y_j/w_{0+}, \tau^2/w_{0+})$$

then $p(y_0, y_1, ...y_n)$ well-defined but not CAR

# Comments

- The CAR specifies $\Sigma_Y^{-1}$, not $\Sigma_Y$ (as in the point-level modeling case), so does not directly model association
- $(\Sigma_Y^{-1})_{ii} = 1/\tau_i^2$; $(\Sigma_Y^{-1})_{ij} = 0 \Leftrightarrow$ cond'l independence
- Prediction at new sites is ad hoc, in that if

$$p(y_0|y_1, y_2, ...y_n) = N(\sum_j w_{0j}y_j/w_{0+}, \tau^2/w_{0+})$$

then $p(y_0, y_1, ...y_n)$ well-defined but not CAR

- Non-Gaussian case: For binary data, the autologistic:

$$p(y_i|y_j, j \neq i) \propto \exp\left\{\phi \sum_j w_{ij} I(y_i = y_j)\right\}$$