STAT 8025

Lecture 11: Lattice Data (I)

Dr. Emily Lei Kang

Division of Statistics & Data Science
Department of Mathematical Sciences
University of Cincinnati

Today...

- Lattice Data
- Reading: SSD Sections 6.1, 6.4

Lattice Data

- Lattice data

  Given study area: D fixed
  Given small areas: $A_1, \ldots, A_n$ with centroid $\mathrm{s}_1, \ldots, \mathrm{s}_n$ and $\cup_{i=1}^{n} A_i = D$

Data: $\mathrm{Y} = (Y_1, \ldots, Y_n)'$ where $Y_i \equiv Y(\mathrm{s}_i)$

Spatial Proximity Measure

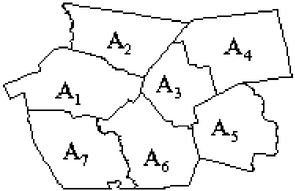
- Recall that for *geostatistical data*, we use
  - Distance between two points
- For lattice data, people often define:
  - Proximity matrix $W \equiv (w_{ij})$
  - There are different ways to define W

Spatial Proximity Measure, ctd.

- Adjacency

$$w_{ij} = \begin{cases} 1 & A_j \text{ shares a common boundary with } A_i \\ 0 & o.w. \end{cases}$$

$$W = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{array} \begin{array}{c} \begin{array}{ccccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$



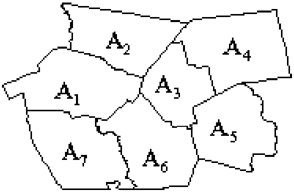
- W is symmetric

Spatial Proximity Measure, ctd.

- k nearest neighbors

$$w_{ij} = \begin{cases} 1 & \text{centroid of } A_j \text{ is one of the } k \text{ nearest centroids to that of } A_i \\ 0 & o.w. \end{cases}$$

$$W = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{array} \begin{array}{c} \begin{array}{ccccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$



- W is *not* symmetric

Spatial Proximity Measure, ctd.

- Other measures

$$w_{ij} = \begin{cases} 1 & \text{centroid of } A_j \text{ is within some specified distance of that of } A_i \\ 0 & o.w. \end{cases}$$

$$w_{ij} = \begin{cases} d_{ij}^{\gamma} & \text{if inter-centroid distance } d_{ij} < \delta (\delta > 0; \gamma < 0) \\ 0 & o.w. \end{cases}$$

$$w_{ij} = \frac{l_{ij}}{l_i}$$

where l_{ij} is the length of common boundary between A_i and A_j; l_i is the perimeter of A_i

Spatial Proximity Measure, ctd.

- Note that proximity matrices are user-defined.
- We can also define first-order neighbors $W^{(1)}$, second-order neighbors $W^{(2)}$, etc. For example: define distance intervals, $(0, d_1]$, $(d_1, d_2]$, and so on.
  - First order neighbors: all units within distance d_1.
  - First order proximity matrix $W^{(1)}$. Analogous to W, $w_{ij}^{(1)} = 1$ is A_i and A_j are first-order neighbors (distance within d_1); 0 otherwise
  - Second order neighbors: all units within distance d_2, but separated by more than d_1
  - And so on

Measures of Spatial Association

Moran's I

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

where

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

- Resembles the Pearson's correlation
- Numerator: Examine deviations from the mean relative to the immediate neighbors
- Denominator: Standardize with variance; trends within the data above or below the mean

Geary's c

Geary proposed a similar statistic: Interaction is not the cross-product of the deviations from the mean, but the deviations in intensities of each observation location with one another

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

- c is never negative, and has mean 1 with Y_i iid
- Low values (between 0 and 1) indicate positive spatial association
- Both I and c are asymptotically normal if Y_i are $i.i.d.$
- I seems to be more popular than c

Moran's I

- $E(I) = -\frac{1}{n-1}$ under independence
- $I > -\frac{1}{n-1} \implies$ positive autocorrelation
- $I < -\frac{1}{n-1} \implies$ negative autocorrelation
- For smaller sample sizes, can use Monte Carlo approach to evaluate significance

Models for Lattice Data

- Formulate one or more models that will allow responses to be correlated with responses at nearby locations. Generally, the most useful models are:
  - reasonably parsimonious
  - readily interpreted
  - computationally feasible to fit to the data
- Lattice data can be continuous or discrete. Initially, we will restrict attention to continuous data.
- Modeling binary and categorical lattice data is also known as *image analysis*.
- Is a geostatistical model based on distance between region centroids valid? Is this a good idea?

Models for Data in One Dimension

- Due to the discrete nature of the spatial locations of lattice data, the most popular models are similar to commonly used models for discrete time series
- Let us digress from spatial statistics for a moment and review one very important time series model: the autoregressive model of order one [AR(1)]:

$$Z_t = \rho Z_{t-1} + \varepsilon_t,\ \{\varepsilon_t\} \sim iid\ N(0, \sigma^2),$$

  where $\rho \in (-1, 1)$ is called the autoregressive coefficient.
  - $corr(Z_t, Z_{t-1}) = \rho$, $corr(Z_t, Z_{t-2}) = \rho^2$, and more generally, $corr(Z_t, Z_{t-k}) = \rho^k$

- Note that interaction is "one-sided" in these models due to the unidirectional flow of time
- Our consideration in lattice data can differ from the classical time series situation:
  - Interactions may be "two-sided"
  - The domain is usually bounded

For example, a model of the two-sided version in one dimension is as follows:

$$
\begin{aligned}
Z_1 &= \rho Z_2 + \varepsilon_1, \\
Z_s &= \rho Z_{s-1} + \rho Z_{s+1} + \varepsilon_s, \;\; s = 2, 3, \ldots, n-1 \\
Z_n &= \rho Z_{n-1} + \varepsilon_n, \\
& \{\varepsilon_s\} \sim iid \;\; N(0, \sigma^2)
\end{aligned}
$$

where we account for "two-side" effects and "edge effects."

An equivalent way to describe this model

$$Z_i|Z_j, j \neq i \sim N(\rho(Z_{i-1} + Z_{i+1}, \sigma^2), i = 2, \ldots, n-1$$

$$Z_1|Z_2, \ldots, Z_n \sim N(\rho Z_2, \sigma^2)$$

$$Z_n|Z1, \ldots, Z_{n-1} \sim N(\rho Z_{n-1}, \sigma^2)$$

described by these conditional distributions.

Markov Random Fields

- Consider statistical models for $\{Y(\mathrm{s}_i) : i = 1, 2, \ldots\}$. One way to define spatial dependence is to use conditional probabilities
- First, note that given $p(y_1, \ldots, y_n)$, the joint distribution, then $p(y_i | y_j, j \neq i)$ (full conditional) for $i = 1, \ldots, n$, are uniquely determined.
- ??? Does $\{p(y_i | y_j, j \neq i)\}$ determine $p(y_1, y_2, \ldots, y_n)$? If so, we call the joint distribution a Markov Random Field (MRF)
- That is, MRF is a type of model defined **"locally"** from all n conditional probabilities: $p(y_i | y_j, j \neq i), i = 1, 2, \ldots, n$

Local Specifications

- **Neighborhoods:** Neighborhood ∂_i of i-th site is defined by,

$$Y_i \perp Y(D - \{\mathrm{s}_i \cup \partial_i\}) | Y(\partial_i)$$

or

$$p(Y_i | \mathrm{Y}_{-i}) = p(Y(\mathrm{s}_i) | Y(\partial_i))$$

That is, the random variable Y_i is **conditionally** *independent* of those at all other sites in D, given the values in ∂_i

- We specify a set of full conditional distributions for the $\{Y_i\}$ such that

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

The notion of using local specification to determine a joint distribution is refereed to as an MRF.

Conditional vs. Joint Distributions

- **Local characteristic**: The local characteristic of an MRF at site s_i is given by the conditional distribution:

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

  The central question underlying the validity of a MRF is when this local specification of an MRF leads to a joint distribution.

- ??? Does $\{p(y_i|y_j, j \neq i)\}$ determine $p(y_1, y_2, \ldots, y_n)$?

Conditional vs. Joint Distributions

- In general we cannot write down an arbitrary set of conditionals and assert that they determine the joint distribution.
- Example:

$$Y_1|Y_2 \sim N(\alpha_0 + \alpha_2 Y_2, \sigma_1^2)$$
$$Y_2|Y_1 \sim N(\beta_0 + \beta_1 Y_1^3, \sigma_1^2)$$

  - From the first equation: $E[Y_1] = \alpha_0 + \alpha_2 E[Y_2]$, i.e., $E[Y_1]$ is linear in $E[Y_2]$
  - From the second equation: $E[Y_2] = \beta_0 + \beta_1 E[Y_1^3]$, i.e. $E[Y_2]$ is linear in $E[Y_1^3]$.
  - Clearly this isn't true in general.

Conditional vs. Joint Distributions, cont'd

- Also $p(y_1, \ldots, y_n)$ may be improper even if all the full conditionals are proper
- Example:

$$p(y_1, y_2) \propto \exp\left\{-\frac{1}{2}(y_1 - y_2)^2\right\}$$

  - It can be derived that $p(Y_2|Y_1) \propto N(Y_1, 1)$ and $p(Y_1|Y_2) \propto N(Y_2, 1)$
  - However, such joint distribution $p(y_1, y_2)$ is *improper*.
- Full conditionals are said to be compatible if there is a valid joint distribution that leads to these full conditionals.

Brook's Lemma

- Brook's lemma enables us to retrieve the joint distribution determined by the conditionals

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

  here $\mathrm{y}_0 = (y_{10}, \ldots, y_{n0})$ is any fixed point in the support of p.
- From Brook's lemma, the joint distribution can be determined up to a proportionality constant by the conditionals, since if LHS is *proper*, the fact that it integrates to 1 determines the normalizing constant!

Proof of Brook's Lemma

The simplest way to prove this is to observe that:

$$p(y_1, \ldots, y_n) = \frac{p(y_n | y_1, \ldots, y_{n-1})}{p(y_{n0} | y_1, \ldots, y_{n-1})} p(y_1, \ldots, y_{n-1}, y_{n0}),$$

[Why?]

and then to proceed recursively by considering:

$$p(y_1, \ldots, y_{n-1}, y_{n0}) = \frac{p(y_{n-1} | y_2, \ldots, y_{n-2}, y_{n0})}{p(y_{n-1,0} | y_1, \ldots, y_{n-2}, y_{n0})} p(y_1, \ldots, y_{n-2}, y_{n-1,0}, y_{n0}),$$

and so on.

Clique and Potential

- **Clique:** A clique is a set of elements such that each element is a neighbor of every other element. We use notation $i \sim j$ if i is a neighbor of j and j is a neighbor of i.
  - Any singleton $\{s\}$ is also a clique
  - A clique is called *maximal* if inclusion of any other additional site prevents it from remaining a clique.
- **Potential:** A potential of order k is a function of k arguments that is exchangeable in these arguments. The arguments of the potential would be the values taken by variables associated with the elements for a clique of size k.

Clique and Potential

Example:

- Consider a clique of size say $k = 2$, $i \sim j$
- The function $\phi(y_i, y_j) = y_i y_j$ is a potential of order 2
- The function $\phi(y_i, y_j) = (y_i - y_j)^2$ is also a potential of order 2

Gibbs distribution

$p(y_1, \ldots, y_n)$ is a Gibbs distribution if it is a function of the Y_i only through potentials on cliques.

$$p(y_1, \ldots, y_n) \propto \exp\left(\gamma \sum_k \sum_{\boldsymbol{\alpha} \in M_k} \phi^{(k)}_{\alpha_1,\ldots,\alpha_k}(y_{\alpha_1}, \ldots, y_{\alpha_k})\right)$$

where $\phi^{(k)}_{\alpha_1,\ldots,\alpha_k}$ is a potential of order k, M_k is the collection of all cliques of size k indexed by $\boldsymbol{\alpha}$, and $\gamma > 0$ is a scale parameter.

- **Hammersley-Clifford Theorem:** If we have an MRF, i.e. if the conditional defines a joint distribution, then this joint distribution is a Gibbs distribution.
- **Geman and Geman (1984):** If we have a joint Gibbs distribution, then we have an MRF (i.e., full conditionals define the joint distribution)
  - Example: Consider the following joint distribution

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

    - Gibbs distribution? Yes
    - It is an MRF, and the full conditional is:

$$p(y_i|y_j, j \neq i) = N(\sum_{j\in\partial_i} y_j/m_i, \tau^2/m_i)$$

where m_i is the number of neighbors of i.

From Joint Distribution to Conditionals

- For an MRF, define the Q-function:

$$Q(\mathrm{y}) \equiv ln\{p(\mathrm{y})/p(\mathrm{y}_0)\}$$

- Using Hammersley-Clifford Theorem, we know that we have a Gibbs distribution. Then

$$Q(\mathrm{y}) = \sum_k \sum_{\boldsymbol{\alpha} \in M_k} \phi^{(k)}_{\alpha_1,\ldots,\alpha_k}(y_{\alpha_1},\ldots,y_{\alpha_k})$$

- Note that the conditional:

$$p(y_i|\mathrm{y}_{-i}) \propto \exp\left(\sum_k \sum_{\boldsymbol{\alpha} \in M_k \& i \in \boldsymbol{\alpha}} \phi^{(k)}_{\alpha_1,\ldots,\alpha_k}(y_{\alpha_1},\ldots,y_{\alpha_k})\right)$$

Besag (1974) shows:

$$\begin{aligned} Q(\mathrm{y}) = \; & \sum_{i=1}^{n} y_i G_i(y_i) + \sum_{i=1}^{n} \sum_{j>i}^{n} y_i y_j G_{ij}(y_i, y_j) \\ & + \cdots + y_1 \cdots y_n G_{1\ldots n}(y_1, \ldots, y_n) \end{aligned}$$

where $G_{i,j,\ldots,s}$ is non-null if and only if $i, j, \ldots, s$ form a clique

From Conditionals to the Joint Distribution

- To obtain the joint distribution, we use Brook's lemma:

$$p(y_1, \ldots, y_n) = \frac{p(y_1|y_2,\ldots,y_n)}{p(y_{10}|y_2,\ldots,y_n)} \frac{p(y_2|y_{10},y_3,\ldots,y_n)}{p(y_{20}|y_{10},y_3,\ldots,y_n)}$$

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

here $\mathrm{y}_0 = (y_{10}, \ldots, y_{n0})$ is any fixed point in the support of p.

- We will study MRFs that are Gibbs distributions with exponential-family conditional distribution and non-null potential of order 1 and 2 only
- Assume

$$p(y_i|\mathrm{y}_{-i}) = \exp[A_i(\mathrm{y}_{\partial_i})B_i(y_i) + C_i(y_i) + D_i(\mathrm{y}_{\partial_i})]$$

- **Proposition (Besag, 1974)**: If the potential functions $\phi^{(k)} = 0$ for $k \geq 3$, then

$$A_i(\mathrm{y}_{\partial_i}) = \alpha_i + \sum_{j=1}^{n} \theta_{ij} B_j(y_j),$$

where $\theta_{ji} = \theta_{ij}$, $\theta_{ii} = 0$, and $\theta_{ik} = 0$ for $k \notin \partial_i$. Further, if $\theta_{ij} = 0$ for all i, j, then $Y_1, \ldots, Y_n$ are independent

Auto Gaussian

$$p(y_i|\mathrm{y}_{\partial_i}) = (2\pi\tau_i^2)^{-1/2}\exp[(-1/2)(y_i - \nu_i(\mathrm{y}_{\partial_i}))^2/\tau_i^2]$$

That is:

$$Y_i|\mathrm{y}_{\partial_i} \sim N(\nu_i(\mathrm{y}_{\partial_i}), \tau_i^2)$$

Now,

$$\nu_i(\mathrm{y}_{\partial_i}) \quad = \alpha_i\tau_i^2 + \sum_{j=1}^n \theta_{ij}\tau_i^2 y_j$$

where the RHS comes from the proposition. Notice that $\theta_{ji} = \theta_{ij}$, $\theta_{ii} = 0$, and $\theta_{ik} = 0$ for $k \notin \partial_i$.
Define $c_{ij} = \theta_{ij}\tau_i^2$; $i, j = 1, \ldots, n$

Auto Gaussian, cont'd

Now further *model*:

$$E(Y_i|\mathrm{y}_{\partial_i}) = \mu_i + \sum_{j=1}^{n} c_{ij}(y_j - \mu_j)$$

Then

$$\frac{c_{ij}}{\tau_i^2} = \frac{c_{ji}}{\tau_j^2}$$

$$c_{ii} = 0$$

$$c_{ik} = 0 \text{ for } k \notin \partial_i$$

The Auto Gaussian model is often called the Conditional Autoregressive (CAR) model; in SSD, it is called as the Conditionally specified Gaussian (CG) model.

Auto Gaussian, cont'd

Proposition (Besag, 1974):
Define $M \equiv diag(\tau_1^2, \ldots, \tau_n^2)$ and $C \equiv (c_{ij})$. We have the full conditionals assumed previously. If $M^{-1}(I - C)$ is positive-definite, then these full conditionals are compatible and the joint distribution is:

$$\mathrm{Y} \sim Gau(\boldsymbol{\mu}, (I - C)^{-1} M),$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$.
Proof: Using Brook's lemma

Let $y_0 = \boldsymbol{\mu}$ in Brook's lemma.

$$
\begin{aligned}
& ln\{p(y_1,\ldots,y_n)/p(\mu_1,\ldots,\mu_n)\} \\
= \; & \textstyle\sum_{i=1}^{n} \left(ln\{p(y_i|y_1,\ldots,y_{i-1},\mu_{i+1},\ldots,\mu_n)\} \right. \\
& \left. -ln\{p(\mu_i|y_1,\ldots,y_{i-1},\mu_{i+1},\ldots,\mu_n)\} \right) \\
= \; & \textstyle\sum_{i=1}^{n} -\frac{1}{2\tau_i^2}\left[y_i-\mu_i-\sum_{j=1}^{i-1} c_{ij}(y_j-\mu_j)\right]^2 \\
& +\textstyle\sum_{i=1}^{n} \frac{1}{2\tau_i^2}\left[\sum_{j=1}^{i-1} c_{ij}(y_j-\mu_j)\right]^2 \\
= \; & \textstyle\sum_{i=1}^{n} \frac{1}{2\tau_i^2}(y_i-\mu_i)^2+\sum_{i=1}^{n}\sum_{j=1}^{i-1}\frac{1}{\tau_i^2} c_{ij}(y_i-\mu_i)(y_j-\mu_j)
\end{aligned}
$$

Since

$$\sum_{i=1}^{n} \frac{1}{2\tau_i^2}(y_i - \mu_i)^2 = -(1/2)(\mathrm{y} - \boldsymbol{\mu})' M^{-1}(\mathrm{y} - \boldsymbol{\mu}),$$

$$\sum_{i=1}^{n}\sum_{j=1}^{i-1} \frac{1}{\tau_i^2} c_{ij}(y_i - \mu_i)(y_j - \mu_j) = (1/2)(\mathrm{y} - \boldsymbol{\mu})' M^{-1} C(\mathrm{y} - \boldsymbol{\mu}),$$

based on

$$\sum_{i=1}^{n}\sum_{j\neq i} a_i d_{ij} a_j = 2\sum_{i=1}^{n}\sum_{j<i} a_i d_{ij} a_j = \mathrm{a}' D^0 \mathrm{a}$$

where $D^0 = (d_{ij})$ is symmetric and $d_{ii} = 0$
Therefore,

$$ln\{p(y_1, \ldots, y_n)/p(\mu_1, \ldots, \mu_n)\} = -(1/2)(\mathrm{y}-\boldsymbol{\mu})' M^{-1}(I-C)(\mathrm{y}-\boldsymbol{\mu})$$

and

$$\mathrm{Y} \sim Gau(\boldsymbol{\mu}, (I - C)^{-1} M),$$

Auto Gaussian, con't

- Therefore, conditional Gaussian distributions define joint Gaussian distributions (with certain conditions satisfied)
- The converse is obviously true
- Note that the spatial dependence is expressed in terms of the inverse variance-covariance matrix

$$M\mathrm{var}(Y)^{-1} = (I - C)$$

- Suppose $\Sigma \equiv \mathrm{var}(Y)$ is known. Is there an auto Gaussian (or CAR) model defined by M and C such that $\Sigma = (I - C)^{-1}M$? YES! Write $\Sigma^{-1} = (\sigma^{(ij)})$. Then let $M = diag(\{\sigma^{(ii)}\}^{-1} : i = 1, \ldots, n)$ and $C = I - M\Sigma^{-1}$, satisfying
$$\Sigma = (I - C)^{-1}M,$$
where M is diagonal and C has zeros down its diagonal

Special Case of Auto Gaussian

Let $\tau_i = \tau$ for all i, and $C = \gamma H$, where H is symmetric and $h_{ii} = 0$. Then

$$Y \sim Gau(\boldsymbol{\mu}, \tau^2(I - \gamma H)^{-1})$$

if $I - \gamma H$ is positive-definite

- When $I - \gamma H$ is positive-definite?
  Write $H = P\Phi P'$, where $PP' = I$ and $\Phi = diag(\phi_1, \ldots, \phi_n)$ are the eigenvalues of H with $\phi_1 \leq \phi_2 \leq \cdots \leq \phi_n$.
  Then $I - \gamma H = P(I - \gamma\Phi))P'$, and so the eigenvalues of $(I - \gamma H)$ are $\{1 - \gamma\phi_1, \ldots, 1 - \gamma\phi_n\}$.
  So to guarantee that $I - \gamma H$ is positive-definite, we need

$$1 - \gamma\phi_i > 0 \text{ for all } i = 1, \ldots, n$$

Special Case of Auto Gaussian

$$Y \sim Gau(\boldsymbol{\mu}, \tau^2(I - \gamma H)^{-1})$$

where $I - \gamma H$ is positive-definite

To make $I - \gamma H$ is positive-definite

- we need $\phi_1^{-1} < \gamma < \phi_n^{-1}$, where $\phi_1 < 0$ and $\phi_n > 0$ [Why?]
- This is used to specify the prior on γ. For example: $Unif(\phi_1^{-1}, \phi_n^{-1})$
- Note that when $\gamma = 0$,

$$Y \sim Gau(\boldsymbol{\mu}, \tau^2 I)$$

That is, $Y_i \sim Gau(\mu_i, \tau^2)$ *independently*

Application: Model for Rates

- Suppose

$$\mathrm{r} = (r_1, \ldots, r_n)'$$

are raw rates, where

$$r_i = \frac{Y_i}{N_i}$$

- We can model r as

$$\mathrm{r} \sim Gau(\boldsymbol{\mu}, (I - C)^{-1} M)$$

where

$$M = diag(\frac{\tau^2}{N_i})$$

and C need to satisfy the associated conditions

Application: A Hierarchical Model with CAR Effects

- Consider the areal data disease mapping model: Poisson-Lognormal (Spatial)

$$Y_i|\mu_i \overset{ind}{\sim} Po(E_i e^{\mu_i})$$

  - Y_i is the observed disease count
  - E_i is expected count (known)
  - $\mu_i = \mathrm{x}_i'\boldsymbol{\beta} + \phi_i$ with explanatory variables x_i
- ϕ_i capture regional clustering via a CAR model

$$\phi_i|\phi_{j\neq i} \sim N(\bar{\phi}_i, \frac{\tau^2}{m_i})$$

$$\bar{\phi}_i = \frac{1}{m_i}\sum_{j\in\partial_i}\phi_j$$

  - Intrinsic autoregressive (IAR) model: Improper; requiring a constraint $\sum_i \phi_i = 0$ (imposed numerically after each MCMC iteration)

Summary

- Lattice data

Preview:

- Other models for lattice data