Lecture 18

Models for areal data

Colin Rundel
03/22/2017
areal / lattice data
Example - NC SIDS

SID79
If we have observations at \( n \) spatial locations \((s_1, \ldots s_n)\):

\[
I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y(s_i) - \bar{y})(y(s_j) - \bar{y})}{\sum_{i=1}^{n} (y(s_i) - \bar{y})}
\]

where \( w \) is a spatial weights matrix.
If we have observations at $n$ spatial locations $(s_1, \ldots s_n)$

$$I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y(s_i) - \bar{y}) (y(s_j) - \bar{y})}{\sum_{i=1}^{n} (y(s_i) - \bar{y})}$$

where $w$ is a spatial weights matrix.

Some properties of Moran’s I (when there is no spatial autocorrelation):

- $E(I) = -1/(n - 1)$

- $Var(I) = E(I^2) - E(I)^2 = $ Something ugly but closed form

- Asymptotically, $\frac{I - E(I)}{\sqrt{Var(I)}} \sim \mathcal{N}(0, 1)$
NC SIDS & Moran’s I

Lets start by using an adjacency matrix for \( w \) (shared county borders).

\[
\text{morans}_I = \text{function}(y, w) \\
\{ \\
\quad n = \text{length}(y) \\
\quad y\_bar = \text{mean}(y) \\
\quad \text{num} = \sum(w \times (y-y\_bar)^2) \\
\quad \text{denom} = \sum((y-y\_bar)^2) \\
\quad (n/\text{sum}(w)) \times (\text{num}/\text{denom}) \\
\}
\]

\[
\text{morans}_I(y = \text{nc}\$\text{SID74}, \ w = 1*\text{st\_touches}(\text{nc}, \ \text{sparse=}\text{FALSE}))
\]

## [1] 0.119089

\[
\text{library(ape)} \\
\text{Moran}\_I(\text{nc}\$\text{SID74}, \ \text{weight} = 1*\text{st\_touches}(\text{nc}, \ \text{sparse=}\text{FALSE})) \%>\% \text{str()}
\]

## List of 4
## $ observed: num 0.148
## $ expected: num -0.0101
## $ sd : num 0.0627
## $ p.value : num 0.0118
Like Moran’s I, if we have observations at \( n \) spatial locations \((s_1, \ldots s_n)\)

\[
C = \frac{n - 1}{2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left( y(s_i) - y(s_j) \right)^2} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left( y(s_i) - y(s_j) \right)^2}{\sum_{i=1}^{n} \left( y(s_i) - \bar{y} \right)}
\]

where \( w \) is a spatial weights matrix.

Some properties of Geary’s C:

- \( 0 < C < 2 \)
- If \( C \approx 1 \) then no spatial autocorrelation
- If \( C > 1 \) then negative spatial autocorrelation
- If \( C < 1 \) then positive spatial autocorrelation

Geary’s C is inversely related to Moran’s I.
EDA - Geary’s C

Like Moran’s I, if we have observations at \( n \) spatial locations \((s_1, \ldots s_n)\)

\[
C = \frac{n - 1}{2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y(s_i) - y(s_j))^2}{\sum_{i=1}^{n} (y(s_i) - \bar{y})}
\]

where \( w \) is a spatial weights matrix.

Some properties of Geary’s C:

• \( 0 < C < 2 \)
  • If \( C \approx 1 \) then no spatial autocorrelation
  • If \( C > 1 \) then negative spatial autocorrelation
  • If \( C < 1 \) then positive spatial autocorrelation

• Geary’s C is inversely related to Moran’s I
NC SIDS & Geary’s C

Again using an adjacency matrix for $w$ (shared county borders).

```r
gearys_C = function(y, w)
{
  n = length(y)
  y_bar = mean(y)
  y_i = y %*% t(rep(1,n))
  y_j = t(y_i)
  num = sum(w * (y_i-y_j)^2)
  denom = sum((y-y_bar)^2)
  ((n-1)/(2*sum(w))) * (num/denom)
}

gearys_C(y = nc$SID74, w = 1*st_touches(nc, sparse=FALSE))
## [1] 0.8898868
```
Spatial Correlogram

d = nc %>% st_centroid() %>% st_distance() %>% strip_class()
breaks = seq(0, max(d), length.out = 21)
d_cut = cut(d, breaks)

adj_mats = map(
  levels(d_cut),
  function(l)
  {
    (d_cut == l) %>%
      matrix(ncol=100) %>%
      'diag<-'(0)
  }
)

d = data_frame(
  dist = breaks[-1],
  morans = map_dbl(adj_mats, morans_I, y = nc$SID74),
  gearys = map_dbl(adj_mats, gearys_C, y = nc$SID74)
)
Autoregressive Models
Lets just focus on the simplest case, an AR(1) process

\[ y_t = \delta + \phi y_{t-1} + w_t \]

where \( w_t \sim \mathcal{N}(0, \sigma^2) \) and \(|\phi| < 1\), then

\[ E(y_t) = \frac{\delta}{1 - \phi} \]

\[ \text{Var}(y_t) = \frac{\sigma^2}{1 - \phi} \]
Previously we saw that an AR(1) model can be represented using a multivariate normal distribution

\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}
\sim \mathcal{N}
\begin{bmatrix}
1 \\
\delta \\
\frac{1}{1-\phi} \\
\frac{\sigma^2}{1-\phi}
\end{bmatrix},
\begin{bmatrix}
1 & \phi & \cdots & \phi^{n-1} \\
\phi & 1 & \cdots & \phi^{n-2} \\
\vdots & \vdots & \ddots & \vdots \\
\phi^{n-1} & \phi^{n-2} & \cdots & 1
\end{bmatrix}
\]
Previously we saw that an AR(1) model can be represented using a multivariate normal distribution

\[
\begin{pmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{pmatrix}
\sim \mathcal{N}
\begin{pmatrix}
  \delta \\
  1 \\
  \vdots \\
  1
\end{pmatrix},
\frac{\sigma^2}{1-\phi}
\begin{pmatrix}
  1 & \phi & \cdots & \phi^{n-1} \\
  \phi & 1 & \cdots & \phi^{n-2} \\
  \vdots & \vdots & \ddots & \vdots \\
  \phi^{n-1} & \phi^{n-2} & \cdots & 1
\end{pmatrix}
\]

In writing down the likelihood we also saw that an AR(1) is 1st order Markovian,

\[
f(y_1, \ldots, y_n) = f(y_1)f(y_2|y_1)f(y_3|y_2, y_1) \cdots f(y_n|y_{n-1}, y_{n-2}, \ldots, y_1)
\]
\[
= f(y_1)f(y_2|y_1)f(y_3|y_2) \cdots f(y_n|y_{n-1})
\]
Competing Definitions for $y_t$

\[ y_t = \delta + \phi y_{t-1} + w_t \]

vs.

\[ y_t | y_{t-1} \sim \mathcal{N}(\delta + \phi y_{t-1}, \sigma^2) \]
Competing Definitions for $y_t$

$$y_t = \delta + \phi y_{t-1} + w_t$$

vs.

$$y_t | y_{t-1} \sim \mathcal{N}(\delta + \phi y_{t-1}, \sigma^2)$$

In the case of time, both of these definitions result in the same multivariate distribution for $y$. 
Even in the simplest spatial case there is no clear / unique ordering,
\[ f(y(s_1), \ldots, y(s_{10})) = f(y(s_1)) f(y(s_2) | y(s_1)) \ldots = f(y(s_{10})) f(y(s_9) | y(s_{10})) \ldots \]

Instead we need to think about things in terms of their neighbors / neighborhood. We will define \( N(s_i) \) to be the set of neighbors of location \( s_i \).

- If we define the neighborhood based on “touching” then \( N(s_3) = \{ s_2, s_4 \} \)
- If we use distance within 2 units then \( N(s_3) = \{ s_1, s_2, s_3, s_4 \} \)
- etc.

<table>
<thead>
<tr>
<th>s1</th>
<th>s2</th>
<th>s3</th>
<th>s4</th>
<th>s5</th>
<th>s6</th>
<th>s7</th>
<th>s8</th>
<th>s9</th>
<th>s10</th>
</tr>
</thead>
</table>
Even in the simplest spatial case there is no clear / unique ordering,

\[ f(y(s_1), \ldots, y(s_{10})) = f(y(s_1)) f(y(s_2)|y(s_1)) \cdots f(y(s_{10}|y(s_9), y(s_8), \ldots, y(s_1)) \]

\[ = f(y(s_{10})) f(y(s_9)|y(s_{10})) \cdots f(y(s_1|y(s_2), y(s_3), \ldots, y(s_{10})) \]

\[ = ? \]
Even in the simplest spatial case there is no clear / unique ordering,

\[ f(y(s_1), \ldots, y(s_{10})) = f(y(s_1)) \cdot f(y(s_2)|y(s_1)) \cdot \ldots \cdot f(y(s_{10}|y(s_9), y(s_8), \ldots, y(s_1)) \]
\[ = f(y(s_{10})) \cdot f(y(s_9)|y(s_{10})) \cdot \ldots \cdot f(y(s_1)|y(s_2), y(s_3), \ldots, y(s_{10})) \]
\[ = ? \]

Instead we need to think about things in terms of their neighbors / neighborhoods. We will define \( N(s_i) \) to be the set of neighbors of location \( s_i \).

- If we define the neighborhood based on “touching” then 
  \[ N(s_3) = \{ s_2, s_4 \} \]
- If we use distance within 2 units then 
  \[ N(s_3) = \{ s_1, s_2, s_3, s_4 \} \]
- etc.
Defining the Spatial AR model

Here we will consider a simple average of neighboring observations, just like with the temporal AR model we have two options in terms of defining the autoregressive process,

- Simultaneous Autoregressive (SAR)
  \[ y(s) = \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2) \]

- Conditional Autoregressive (CAR)
  \[ y(s)|y_{-s} \sim \mathcal{N} \left( \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s'), \sigma^2 \right) \]
Simultaneous Autoregressive (SAR)

Using

$$y(s) = \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2)$$

we want to find the distribution of $y = \left( y(s_1), y(s_2), \ldots, y(s_n) \right)^t$. 
Simultaneous Autoregressive (SAR)

Using

\[ y(s) = \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2) \]

we want to find the distribution of \( y = (y(s_1), y(s_2), \ldots, y(s_n))^t \).

First we need to define a weight matrix \( \mathbf{W} \) where

\[
\{\mathbf{W}\}_{ij} = \begin{cases} 
1/|N(s_i)| & \text{if } j \in N(s_i) \\
0 & \text{otherwise}
\end{cases}
\]
Simultaneous Autoregressive (SAR)

Using

\[ y(s) = \delta + \phi \frac{1}{|N(s)|} \sum_{s' \in N(s)} y(s') + \mathcal{N}(0, \sigma^2) \]

we want to find the distribution of \( y = (y(s_1), y(s_2), \ldots, y(s_n))^t \).

First we need to define a weight matrix \( W \) where

\[
\{W\}_{ij} = \begin{cases} 
1/|N(s_i)| & \text{if } j \in N(s_i) \\
0 & \text{otherwise}
\end{cases}
\]

then we can write \( y \) as follows,

\[ y = \delta + \phi Wy + \epsilon \]

where

\[ \epsilon \sim \mathcal{N}(0, \sigma^2 I) \]
A toy example

\[ W = \begin{bmatrix}
0 & 1/2 & 1/2 \\
1/2 & 0 & 1/2 \\
1/2 & 1/2 & 0
\end{bmatrix} \]
\[ y = \delta + \phi W y + \epsilon \]
This is a bit trickier, in the case of the temporal AR process we actually went from joint distribution $\rightarrow$ conditional distributions (which we were then able to simplify).

Since we don’t have a natural ordering we can’t get away with this (at least not easily).

Going the other way, conditional distributions $\rightarrow$ joint distribution is difficult because it is possible to specify conditional distributions that lead to an improper joint distribution.
Brook’s Lemma

For sets of observations $x$ and $y$ where $p(x) > 0 \ \forall x \in x$ and $p(y) > 0 \ \forall y \in y$ then

$$
\frac{p(y)}{p(x)} = \prod_{i=1}^{n} \frac{p(y_i | y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n)}{p(x_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}
$$

$$
= \prod_{i=1}^{n} \frac{p(y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{p(x_i | y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n)}
$$
A simplified example

Let $y = (y_1, y_2)$ and $x = (x_1, x_2)$ then we can derive Brook’s Lemma for this case,

\[
p(y_1, y_2) = p(y_1 | y_2) p(y_2)
\]

\[
= p(y_1 | y_2) \frac{p(y_2 | x_1) p(x_1)}{p(x_1 | y_2)} = \frac{p(y_1 | y_2)}{p(x_1 | y_2)} p(y_2 | x_1) p(x_1)
\]

\[
= \frac{p(y_1 | y_2)}{p(x_1 | y_2)} p(y_2 | x_1) p(x_1) \left( \frac{p(x_2 | x_1)}{p(x_2 | x_1)} \right)
\]

\[
= \frac{p(y_1 | y_2)}{p(x_1 | y_2)} \frac{p(y_2 | x_1)}{p(x_2 | x_1)} p(x_1, x_2)
\]

\[
\frac{p(y_1, y_2)}{p(x_1, x_2)} = \frac{p(y_1 | y_2) p(y_2 | x_1)}{p(x_1 | y_2) p(x_2 | x_1)}
\]
Utility?

Let’s repeat that last example but consider the case where \( y = (y_1, y_2) \) but now we let \( x = (y_1 = 0, y_2 = 0) \)

\[
\frac{p(y_1, y_2)}{p(x_1, x_2)} = \frac{p(y_1, y_2)}{p(y_1 = 0, y_2 = 0)}
\]

\[
p(y_1, y_2) = \frac{p(y_1|y_2)}{p(y_1 = 0|y_2)} \frac{p(y_2|y_1 = 0)}{p(y_2 = 0|y_1 = 0)} p(y_1 = 0, y_2 = 0)
\]

\[
p(y_1, y_2) \propto \frac{p(y_1|y_2) p(y_2|y_1 = 0)}{p(y_1 = 0|y_2)} \frac{p(y_2|y_1 = 0)}{p(y_2 = 0|y_1)} p(y_1 = 0, y_2 = 0)
\]
As applied to a simple CAR

\[ y(s_1) | y(s_2) \sim \mathcal{N}(\phi_{W_{12}} y(s_2), \sigma^2) \]
\[ y(s_2) | y(s_1) \sim \mathcal{N}(\phi_{W_{21}} y(s_1), \sigma^2) \]
As applied to a simple CAR

\[ y(s_1) | y(s_2) \sim \mathcal{N} (\phi W_{12} y(s_2), \sigma^2) \]

\[ y(s_2) | y(s_1) \sim \mathcal{N} (\phi W_{21} y(s_1), \sigma^2) \]

\[
p(y(s_1), y(s_2)) \propto \frac{p(y(s_1)|y(s_2)) \ p(y(s_2)|y(s_1) = 0)}{p(y(s_1) = 0|y(s_2))} \]

\[
\propto \exp \left( -\frac{1}{2\sigma^2} (y(s_1) - \phi W_{12} y(s_2))^2 \right) \ exp \left( -\frac{1}{2\sigma^2} (y(s_2) - \phi W_{21} 0)^2 \right) \]

\[
\propto \exp \left( -\frac{1}{2\sigma^2} \ (y(s_1) - \phi W_{12} y(s_2))^2 + y(s_2)^2 - (\phi W_{12} y(s_2))^2 \right) \]

\[
\propto \exp \left( -\frac{1}{2\sigma^2} \ (y(s_1)^2 - 2\phi W_{12} y(s_1) y(s_2) + y(s_2)^2) \right) \]

\[
\propto \exp \left( -\frac{1}{2\sigma^2} \ (y - 0) \begin{pmatrix} 1 & -\phi W_{12} \\ -\phi W_{12} & 1 \end{pmatrix} (y - 0)^t \right) \]
Implications for $y$

\[ \mathbf{\mu} = 0 \]

\[ \mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 1 & -\phi W_12 \\ -\phi W_12 & 1 \end{pmatrix} \]

\[ = \frac{1}{\sigma^2} (I - \phi W) \]

\[ \mathbf{\Sigma} = \sigma^2 (I - \phi W)^{-1} \]
Implications for $y$

\[ \mu = 0 \]

\[ \Sigma^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 1 & -\phi W_{12} \\ -\phi W_{12} & 1 \end{pmatrix} \]

\[ = \frac{1}{\sigma^2} (I - \phi W) \]

\[ \Sigma = \sigma^2 (I - \phi W)^{-1} \]

we can then conclude that for $y = (y(s_1), y(s_2))^t$,

\[ y \sim \mathcal{N} \left(0, \sigma^2 (I - \phi W)^{-1} \right) \]

which generalizes for all mean 0 CAR models.
Let $\mathbf{y} = (y(s_1), \ldots, y(s_n))$ and $\mathbf{0} = (y(s_1) = 0, \ldots, y(s_n) = 0)$ then by Brook’s lemma,

$$
\frac{p(\mathbf{y})}{p(\mathbf{0})} = \prod_{i=1}^{n} \frac{p(y_i|y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}{p(0_i|y_1, \ldots, y_{i-1}, 0_{i+1}, \ldots, 0_n)}
$$

$$
= \prod_{i=1}^{n} \frac{\exp \left( -\frac{1}{2\sigma^2} \left( y_i - \phi \sum_{j<i} W_{ij} y_j - \phi \sum_{j>i} 0_j \right)^2 \right)}{\exp \left( -\frac{1}{2\sigma^2} \left( 0_i - \phi \sum_{j<i} W_{ij} y_j - \phi \sum_{j>i} 0_j \right)^2 \right)}
$$

$$
= \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( y_i - \phi \sum_{j<i} W_{ij} y_j \right)^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( \phi \sum_{j<i} W_{ij} y_j \right)^2 \right)
$$

$$
= \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 - 2\phi y_i \sum_{j<i} W_{ij} y_j \right)
$$

$$
= \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 - \phi \sum_{i=1}^{n} \sum_{j=1}^{n} y_i W_{ij} y_j \right) \quad \text{(if } W_{ij} = W_{ji} \text{)}
$$

$$
= \exp \left( -\frac{1}{2\sigma^2} \mathbf{y}^t (I - \phi \mathbf{W}) \mathbf{y} \right)
$$
• Simultaneous Autoregressive (SAR)

\[ y(s) = \phi \sum_{s'} \frac{W_{ss'}}{W_s} y(s') + \epsilon \]

\[ y \sim \mathcal{N}(0, \sigma^2 ((I - \phi W)^{-1})(I - \phi W)^{-1}^t) \]

• Conditional Autoregressive (CAR)

\[ y(s)|y_{-s} \sim \mathcal{N} \left( \sum_{s'} \frac{W_{ss'}}{W_s} y(s'), \sigma^2 \right) \]

\[ y \sim \mathcal{N}(0, \sigma^2 (I - \phi W)^{-1}) \]
Generalization

- Adopting different weight matrices, $W$
  - Between SAR and CAR model we move to a generic weight matrix definition (beyond average of nearest neighbors)
  - In time we varied $p$ in the $AR(p)$ model, in space we adjust the weight matrix.
  - In general having a symmetric $W$ is helpful, but not required

- More complex Variance (beyond $\sigma^2 I$)
  - $\sigma^2$ can be a vector (differences between areal locations)
  - E.g. since areal data tends to be aggregated - adjust variance based on sample size