COA 616 Geostatistics in Environmental Sciences

Lecture 12 – Modelling areal data

Wei Wu

November 29, 2016

Modeling areal data (SAR)

Simultaneous autoregressive model (SAR): Uses a regression on the values from the other areas to account for the spatial dependence. The error terms ε are modeled so that they depend on each other in the following way:

$$Z = \mu + e$$

$$e_i = \sum_{i=1}^{m} b_{ij} e_i + \varepsilon_i$$

ε_{i} represents residual errors, $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_m) \sim N(\vec{0}, \vec{\Lambda})$ with $\vec{\Lambda}$ diagonal

$\mathrm{b_{ij}}$ values are used to represent spatial dependence between areas,

$\mathrm{b_{ii}}$ must be set to 0 so that each area is not regressed on itself.

If we express the error terms as $\mathrm{e} = \mathrm{B}(\mathrm{Y} - \mathrm{X}^{\mathrm{T}}\beta) + \varepsilon$, the model can also be expressed as

$$\mathrm{Z} = \mathrm{X}^{\mathrm{T}}\beta + B(Y - X^{T}\beta) + \varepsilon$$

Hence, the model can be fomulated in a matrix form as follows :

$$(\mathrm{I} - \mathrm{B})(\mathrm{Y} - \mathrm{X}^{\mathrm{T}}\beta) = \varepsilon$$

where $\mathrm{B} = (\mathrm{b_{ij}})$, I is the m dimensional identity matrix

$I - B$ must be non - singular

SAR cont.

$$[Z] \sim N(X^T\beta, (I-B)^{-1}\Sigma_\varepsilon(I-B^T)^{-1})$$

Often $\Sigma_\varepsilon = \sigma^2 I$ then

$$[Z] \sim N(X^T\beta, \sigma^2(I-B)^{-1}(I-B^T)^{-1})$$

Let $\mathrm{B} = \lambda\mathrm{W}$ where λ is a spatial autocorrelation parameter and W is a matrix that represents spatial dependence (often symmetric), then

$$\mathrm{var(Z)} = \sigma^2(I-\lambda\mathrm{W})^{-1}(I-\lambda\mathrm{W}^T)^{-1}$$

Modeling areal data (CAR)

Conditional autoregressive model (CAR): Relies on the conditional distribution of the spatial error term.

$$e_i \mid e_{j\sim i} \sim N\left(\sum_{j\sim i}\frac{c_{ij}e_j}{\sum_{j\sim i}c_{ij}}, \frac{\sigma_{e_i}{}^2}{\sum_{j\sim i}c_{ij}}\right)$$

$$Z(A_i) \mid Z(A_j) \sim N(\mu_i + \sum_{j=1}^{n} c_{ij}(Z(A_j)-\mu_j), \tau_i^2)$$

$$Z \sim N(\vec{\mu}, (\vec{I}_n - \vec{C})^{-1}\vec{T})$$

where $\mathrm{c_{ij}}$ *are* dependence parameters similar to $\mathrm{b_{ij}}$

$\vec{C} = (c_{ij})$, and $\vec{\mathrm{T}} = \mathrm{diag}(\tau_1^2, \tau_2^2, ..., \tau_3^2)$

$$\sigma_j^2 c_{ij} = \sigma_i^2 c_{ji}$$

The structure of **B** and **C** are specified by the shape of lattice. One common way is with a singular parameter that scales neighborhood matrix **W** that indicates whether the regions are neighbor or not.

$$w_{ij} = \begin{cases} 1 & \text{if region } \mathrm{A_i} \text{ shares the same border or vertex with region } \mathrm{A_j} \\ 0 & \text{if i = j} \\ 0 & \text{otherwise} \end{cases}$$

CAR cont.

CAR models are typically specified in a hierarchical Bayesian framework, with inference based on Markov Chain Monte Carlo simulation.

1. Data likelihood

$Z_k \sim f(z_k \mid \mu_k, \nu^2)$ for k = 1, 2, ..., m

$g(\mu_k) = x_k^T \beta + \phi_k + O_k$

ϕ : random effects

O : known offsets

2. Priors

2.1) Independent prior

$\phi_k \sim N(0, \sigma^2)$

$\sigma^2 \sim U(0, M_\sigma)$

2.2) Global priors

$$\phi_k \mid \phi_{-k}, W, \tau^2 \sim N\left(\frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}}, \frac{\tau^2}{\sum_{i=1}^n w_{ki}}\right)$$

$\tau^2 \sim U(0, M_\tau)$

CARBayes

Besag-York-Mollie (BYM) (1991) CAR model (CARbym())

$$\varphi_k = \phi_k + \theta_k$$

$$\phi_k \mid \phi_{-k}, W, \tau^2 \sim N\left(\frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}}, \frac{\tau^2}{\sum_{i=1}^n w_{ki}}\right)$$

$$\theta_k \sim N(0, \theta^2)$$

$$\tau^2, \theta^2 \sim Inverse-Gamma(a, b)$$

Leroux et al. (1999) (CARleroux())

$$\varphi_k = \phi_k$$

$$\phi_k \mid \phi_{-k}, W, \tau^2, \rho \sim N\left(\frac{\rho\sum_{i=1}^n w_{ki}\phi_i}{\rho\sum_{i=1}^n w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{i=1}^n w_{ki} + 1 - \rho}\right)$$

$$\tau^2 \sim Inverse-Gamma(a, b)$$

$$\rho \sim Uniform(0,1)$$