LECTURE 20: AREAL DATA MODEL FITTING

CLASS INTRO

INTRO QUESTIONS

Last Time:

- Discuss the mechanism for include spatial random effects in for areal data. How is this simulated / how can models account for this?

Today:

- Model Fitting for Areal Data

SPDEP

ADJACENCY MATRIX

Using the code below, create an adjacency matrix for Montana. Then identify the neighbors for Gallatin county.

```
MT.counties <- map('county', 'montana', fill = T, plot = F)
map('county', 'montana')
county.ID <- sapply(strsplit(MT.counties$names, ','),
                    function(x) x[2])
mt.poly <- map2SpatialPolygons(MT.counties, IDs = county.ID)
mt.nb <- poly2nb(mt.poly)
mt.adj.mat <- nb2mat(mt.nb, style = 'B')
```

These functions also work with general shape files.

MORAN'S I / GEARY'S C

Recall:

- Moran's I
  $$I = \frac{n \sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i \neq j} \ w_{ij}) \sum_i (Y_i - \bar{Y})^2}.$$
  This is a spatial analogue measuring the lagged autocorrelation.

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

MORAN'S I / GEARY'S C

- The R package `spdep` contains built in functions for Moran's I and Geary's C.
- `moran.test()` and `geary.test()` both take a numeric vector (response) and a `listw` object created by `nb2listw` as arguments.

MORAN’S I / GEARY’S C

Using the Tester - Rosendale election results, compute and interpret Moran’s I and Geary’s C with the proportion voting for Tester.

```
Tester <- read_csv('Tester_Results.csv')
Tester <- Tester %>%
  mutate(Tester_Prop = TESTER / (TESTER + ROSENDALE + BRECKENRIDGE))

#drop Yellowstone National Park
mt.poly.noYNP <- mt.poly[1:56,]
mt.nb.noYNP <- poly2nb(mt.poly.noYNP)
mt.listw <- nb2listw(mt.nb.noYNP, style = 'W')
```

SAR / CAR

The `spdep` package also contains the functionality to fit SAR / CAR models.

Follow the include tutorial code and answer these four questions.

1. Summarize the data set, note Z is a standardized (standard normal) response for PROPCAS
2. What is nyadjmat
3. Summarize the results from SAR
4. Choose a model between lm0, SAR, and CAR

BAYESIAN MODELS FOR AREAL DATA

CARBAYES

Similar to earlier functionality, there are R packages for analyzing areal data using Bayesian methods. We will look at `CARBayes` **Tutorial**

PROPERTY VALUES IN GLASGOW TUTORIAL

- Using the `CARBayes` package, answer the following questions.

1. Describe the data set
2. What are the results of `moran.mc`? What is the purpose of using `resid.model` as the response?
3. Interpret and describe the results of the `S.CARleroux()` model call.

JAGS

- Again JAGS is a possibility for any situation, it just requires sampling model and prior along with explicit documentation.

- Recall
  $$Y_i|\psi_i \sim Poisson(E_i \exp(\psi_i))$$
  $$\psi_i = \boldsymbol{x_i^T}\boldsymbol{\beta} + \theta_i + \phi_i$$
  where $\boldsymbol{x_i}$ are spatial covariates, θ_i corresponds to region wide heterogeneity, and ψ_i captures local clustering.

JAGS MODEL CODE

```
car_model <- "model {
  for (i in 1 : regions) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + beta0 + beta1*x1 + phi[i] + theta[i]
    theta[i] ~ dnorm(0.0, tau.h)
    xi[i] <- theta[i] + phi[i]
    SMRhat[i] <- 100 * mu[i] / E[i]
    SMRraw[i] <- 100 * O[i] / E[i]
  }
  phi[1:regions] ~ car.normal(adj[], weights[], num[], tau.c)

  beta0 ~ dnorm(0.0, 1.0E-5)
  beta1 ~ dnorm(0.0, 1.0E-5)

  tau.h ~ dgamma(1.0E-3, 1.0E-3)
  tau.c ~ dgamma(1.0E-3, 1.0E-3)

  sd.h <- sd(theta[]) #marginal SD of heterogeneity effects
  sd.c <- sd(phi[]) # marginal SD of clustering effects
  alpha <- sd.c / (sd.h + sd.c)
  }
}"
```