# Homework 3

Stat 597a: Spatial Models

*Claire Kelling*

*Due XXXXX, 2017*

## Problem 1:

Fit a linear model relating rent per square meter to the covariates using least squares, and extract the coefficient estimates. You can ignore the Location variable for now since we will later treat this as a random effect. Also note that the room indicator variables include one that is redundant, so treat a single room as the baseline (i.e., leave Room1 out of the model, so the intercept corresponds to a single room and coefficients for the others represent adjustments to the intercept for a different number of rooms).

## Problem 2:

There are two SpatialPolygons objects associated with this dataset, districts.sp and parks.sp. The first corresponds to city districts in which apartments may be located. The second corresponds to districts with no possible apartments, such as parks or fields. Create an nb object with neighbors for the districts, defining neighbors as districts that share a common boundary. Make a plot showing the districts, then add the parks shaded a different color. Think about the way parks are treated; what else could we do?

## Problem 3:

There are 380 districts in districts.sp, and the corresponding district numbers are indicated by the Location variable in rents. How many of the 380 districts appear in the rent dataset? I've included a matrix H that provides a mapping between the districts as they're ordered in districts.sp and as they appear in the rents dataframe. Use H to create a new vector containing the number of observations in each district, and make a color or grayscale plot to illustrate this. Note that inference for unobserved districts will still be possible under a hierarchical mixed effects model, since we can "borrow strength" from nearby districts that do have observations.

## Problem 4

We will now create a Gibbs sampler to sample from the posterior distribution under the following Bayesian model. Let X be the matrix of covariates, including the intercept term. Let n be the number of data points in Y and m be the number of spatial locations in $\eta$. Data model:

$$Y|\beta, \eta, \sigma^2 \sim N(X\beta + H\eta, \sigma^2 I)$$

Process model:

$$p(\eta|\tau^2) \propto (\tau^2)^{-(m-1)/2} \exp\{\frac{-1}{2\tau^2}\eta^T(D_w - W)\eta\}$$

where W is the matrix of 0's and 1's indicating the neighborhood structure from problem 2, and $D_w$ is a diagonal matrix with diagonal entries $\sum_j W_{1j}, ..., \sum_j W_{nj}$. That is, $\eta$ follows an (improper) intrinsic

autoregressive model. The $-(m-1)/2$ exponent on $\tau^2$ is due to the fact that the matrix $D_w - ????W$ has rank m-1 rather than m.

Prior model: Specify independent priors for $\beta$, $\sigma^2$, and $\tau^2$, with

$$p(\beta) \propto 1, \text{ and } \sigma^2, \tau^2 \sim InverseGamma(0:001, 0:001)$$

The full conditional distributions for $\beta$, $\eta$ $\sigma^2$, and $\tau^2$ are given at the end of this assignment. Construct a Gibbs sampler that cycles through each of the full conditionals and stores the results for B = 10,000 iterations. The full conditionals are given below.

A few notes to keep in mind when constructing the sampler:

- The matrix W can be computed from your nb object in problem 2; see help(nb2mat). I also included objects X and y with the data file.

- The function rinvgamma is in the library MCMCpack.

- IMPORTANT: The intrinsic autoregressive model is an example of a pair-wise difference prior. It defines proper distributions for the differences $n_i - n_j$, but it also implicitly contains a distribution for $\frac{1}{m}\sum_{i=1}^{m} \eta_i$ that has infinite variance. In practice, since there is also an intercept term in $X\beta$, we impose the constraint $\sum_{i=1}^{m} \eta_i = 0$ when we sample from the full conditional for $\eta$. **Do this numerically by subtracting the mean $\frac{1}{m}\sum_{i=1}^{m} \eta_i^{(j)}$ from $\eta^{(j)}$ in each iteration j.**

Turn in the following:

- Your map with the neighbors from problem

- Your map of the apartment counts for each district

- Trace plots and ACF plots for $\sigma^2$ and $\tau^2$

- A table with posterior means of the $\beta$'s and 95

- A color or grayscale map of the posterior means for the vector $\eta$

- A color or grayscale map of the posterior standard deviations for the vector $\eta$