2020 BIOSTAT 555
Statistical Methods for Spatial Epidemiology
Assignment 3

To be submitted to the canvas site by the start of class on Monday 24th February, 2020.

Hand in your `R` code as an appendix.

1. In this question we will smooth the SMRs for the Ohio lung cancer data using the disease mapping Poisson-Lognormal-Spatial model:

$$\begin{aligned} Y_i|\beta_0, S_i, \epsilon_i &\underset{\text{ind}}{\sim} \text{Poisson}(E_i \text{e}^{\beta_0} \text{e}^{S_i+\epsilon_i}) \\ \epsilon_i|\sigma^2_\epsilon &\underset{\text{iid}}{\sim} \text{N}(0, \sigma^2_\epsilon) \\ S_1, \ldots, S_n|\sigma^2_s &\sim \text{ICAR}(\sigma^2_s) \end{aligned}$$

   for i, $i = 1, \ldots, n$.

   (a) Using the `inla` function in `R` fit this model using the `bym2` model, with the default prior for β_0 and the following prior specification for the spatial and non-spatial random effects:

   ```
   f(Region, model="bym2", graph="ohio_map.graph", scale.model=T, constr=T,
   hyper=list(phi=list(prior="pc", param=c(0.5, 0.5), initial=1),
   prec=list(prior="pc.prec", param=c(0.3,0.01), initial=5)))
   ```

   These choices correspond to the prior belief that there is a 1% chance that the total residual standard deviation is greater than 0.3, and a 50% chance that the proportion of the variance that is spatial is bigger than 0.5. Report both the posterior medians and 95% intervals for β_0, the total variance of the random effects, and the proportion of the total variance attributed to the spatial random effect.

(b) Extract the relative risk estimates and provide a map of these. Compare these estimates with the SMRs and with those obtained from the Poisson-Lognormal model (i.e., the model with IID random effects only) that you fit in Assignment 2.

2. Using the `SUMMER` package, carry out small area estimation for the outcome smoking for Health Reporting Areas (HRAs).

   (a) Calculate:

      i. Simple proportion of smokers in each area (i.e., as if we had a binomial sampling model).

      ii. Combine a binomial sampling model with the BYM2 smoothing model, and obtain posterior median estimates of the smoking proportion in each HRA.

      iii. Weighted (direct) estimates for each area.

      iv. Smoothed direct estimates (again take the posterior medians).

   (b) Plot the point estimates against each other and comment.

   (c) Plot the standard errors (for (i) and (iii)) and the posterior standard deviations (for (ii) and (iv)) against each other and comment.

   (d) Summarize this analysis, in terms of the most appropriate model, and describing the variation in smoking prevalence across HRAs in King County.