Homework 9 - Stat 534
Due Monday April 17, 2017

1. Empirical Bayes - NC SIDS data: This is a well-known data set on the incidence of SIDS (Sudden Infant Death Syndrome) deaths in North Carolina. The data are available in the `nc.sids` data set in the `spdep` package. We will use emprical Bayes' methods to produce smoothed disease maps of SIDS incidence in the 1974-78 period. I have been letting you do most of the coding yourself this semester but getting the plots out for a problem like this is too much so you can use the `R` code below. Briefly discuss the results. In addition to the maps you can look at the raw SMR values and the smoothed estimates and summarize them in some meaningful way. I am particularly interested in counties in which the data dominate and counties in which the empirical Bayes' priors dominate. I will send an `R` script file with the code included also. I do not know what all those commands mean myself but I know they work.

```
require(maptools)
require(spdep)
require(rgdal)
nc.sids<-readOGR(
      "C:/Users/r71q845/Documents/R/win-library/3.3/spdep/etc/shapes/sids.shp")
# correct coordinate system so that order in the vectors below will be OK
proj4string(nc.sids)<-CRS("+proj=longlat +ellps=clrk66")
# two different row names for different purposes below
row.names(nc.sids)<-sapply(slot(nc.sids, "polygons"), function(x) slot(x, "ID"))
rn <- nc.sids$FIPSNO
# Get the raw SMR values
r<-sum(nc.sids$SID74)/sum(nc.sids$BIR74)
Expected<-nc.sids$BIR74 * r
SMR<-nc.sids$SID74/Expected
# Get the empirical Bayes' estimates
require(DCluster)
# Poisson Gamma
EB.pg<-empbaysmooth(nc.sids$SID74,Expected)$smthrr
# Poisson-LogNormal
EB.LogN<-exp(lognormalEB(nc.sids$SID74,Expected)$smthrr)
# Marshall's Global
EB.Global<-EBest(nc.sids$SID74,Expected)$estmm
# Marshall's Local
# need the neighbors designation
ncCR85_nb <- read.gal(system.file("etc/weights/ncCR85.gal", package="spdep")[1],
  region.id=rn)
EB.Local<-EBlocal(nc.sids$SID74,Expected,ncCR85_nb)$est
# put it all together and produce the plots.
nc<-cbind(nc.sids,SMR,EB.pg,EB.LogN,EB.Global,EB.Local)
names(nc)<-c(names(nc.sids)[1:20],"SMR","EB.pg","EB.LogN","EB.Global","EB.Local")
```

```
spplot(nc,c("SMR","EB.pg","EB.LogN","EB.Global","EB.Local"))
```

2. Continuing with the NC SIDS data. We assume the Poisson-Gamma model.

   (a) Find the posterior distribution assuming a more informative Gamma prior with $\alpha = \beta = 4$.

   (b) Find the posterior distribution assuming a non-informative Gamma prior with $\alpha = \beta = 0.1$

   (c) Summarize the role of the data and the prior in the posterior distributions by comparing the observed SMR and the posterior means from the 2 Bayesian approaches. A plot would help and you should be able to modify the code above to produce those plots. Just add the vectors of the poterior distribution to `nc.sids` using `cbind` and then use `spplot` to get plots of SMR, and the pure Bayes' posterior means.

3. See the notes for the details on the Global Estimator (page 85). Marshall assumed, with E_i being the expected count that

$$Z_i|\gamma_i \sim Poi(E_i\gamma_i)$$

but made no assumption about the distribution of the spatially varying relative risks or SMR values γ_i. We denote the estimates of SMR by $r_i = Z_i/E_i$. We make no distributional assumptions about the γ_i values but do assume that the prior means and variances exist. Denote them m_{γ_i} and v_{γ_i}, respectively. Marshall found method-of-moments estimators of the marginal or unconditional mean and variance of the γ_i.

   (a) Find the conditional mean and variance of r_i. That is find $E(r_i|\gamma_i)$ and $Var(r_i|\gamma_i)$.

   (b) Show that the marginal mean of r_i is equal to the prior mean of γ_i.

   (c) Find the marginal variance of r_i. Denote the (Hint: it will be a function of the prior mean and variance of γ_i).

The method-of-moments estimators are given in the notes.