Spatial modeling of gun-related homicide rates

Phillip Yates

10/29/2020

Storing this here as a reference because it's a good example of Modeling for Spatial Correlation and it hasn't been covered in any class yet so I will want to look at this later.

First we will load the `tidyverse` library for data reshaping and visualization.

```
library(tidyverse)
```

The data from this analysis come from "Firearm legislation and firearm morality in the USA: a cross-sectional, state-level study" by Kalesan et al. (2016). The response variable, Y_i, is the log firearm-related death rate (i.e., the log of the number of deaths divided by the population) in 2010 in state i. This is regressed onto five potential cofounders.

1. log 2009 firearm death rate per 10,000 people
2. Firearm ownership rate quartile
3. Unemployment rate quartile
4. Non-firearm homicide rate quartile
5. Firearm export rate quartile

The covariate of interest is the number of gun control laws in effect in the state. This gives $p = 6$ covariates.

We fit the linear model

$$Y_i = \beta_0 + \sum_{j=1}^{p} X_i \beta_j + \epsilon_i.$$

We compute the usual non-spatial model with $\epsilon_i \sim \text{Normal}\left(0, \sigma^2\right)$ with the spatial model $\text{Cov}\left(\epsilon_1, \ldots, \epsilon_n\right) \sim \text{Normal}(\mathbf{0}, \mathbf{\Sigma})$. The covariance

$$\mathbf{\Sigma} = \tau^2 \mathbf{S} + \sigma^2 \mathbf{I}_n$$

is decomposed into a spatial covariance $\tau^2 \mathbf{S}$ and a non-spatial covariance $\sigma^2 \mathbf{I}_n$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix. The spatial covariance follows the conditionally-autoregressive model $\mathbf{S} = (\mathbf{M} - \rho \mathbf{A})^{-1}$, where $\mathbf{A}$ is the adjacency matrix with (i, j) element equal 1 if states i and j are neighbors and zero otherwise, and M is the diagonal matrix with i^{th} diagonal element equal to the number of states that neighbor state i.

Load the data

```
load("~/StatisticsMasters/bayesian/guns.rdata")
Y <- log(10000*Y/N)
Z[,1] <- log(Z[,1]) # Z is also the 2nd column in X
X <- cbind(1,Z,rowSums(X))

# Remove Alaska and Hawaii
Y <- Y[-c(2,11)]
X <- X[-c(2,11),]

n <- length(Y)
p <- ncol(X)
```

Fit the non-spatial model

```
ns_model <- "model{

  # Likelihood
  for(i in 1:n){
    Y[i] ~ dnorm(mu[i],taue)
    mu[i] <- inprod(X[i,],beta[])
  }
  # Priors
  for(j in 1:p){
    beta[j] ~ dnorm(0,0.01)
  }
  taue ~ dgamma(0.1,0.1)
  sig <- 1/sqrt(taue)
}"

library(rjags)
dat <- list(Y=Y,n=n,X=X,p=p)
init <- list(beta=rep(0,p))
model1 <- jags.model(textConnection(ns_model),inits=init,data=dat,quiet=TRUE)
update(model1,10000,progress.bar="none")
samp1 <- coda.samples(model1,variable.name=c("beta","sig"),n.iter=20000,
                      progress.bar="none")
summary(samp1)
```

```
##
## Iterations = 10001:30000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 20000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean       SD  Naive SE Time-series SE
## beta[1]  0.018309 0.082986 5.868e-04      0.0042871
## beta[2]  0.749915 0.082229 5.814e-04      0.0028864
## beta[3] -0.001870 0.019956 1.411e-04      0.0007665
## beta[4] -0.013559 0.015900 1.124e-04      0.0004122
## beta[5]  0.019993 0.017777 1.257e-04      0.0005595
## beta[6]  0.018221 0.018292 1.293e-04      0.0006310
## beta[7] -0.007745 0.004451 3.147e-05      0.0001011
## sig      0.101305 0.011458 8.102e-05      0.0001262
##
## 2. Quantiles for each variable:
##
##             2.5%       25%       50%       75%     97.5%
## beta[1] -0.13831 -0.038264  0.015396  0.072071 0.1884339
## beta[2]  0.59194  0.694109  0.748453  0.804361 0.9164515
## beta[3] -0.04113 -0.014979 -0.002198  0.011079 0.0384936
## beta[4] -0.04513 -0.024034 -0.013632 -0.002925 0.0176630
## beta[5] -0.01591  0.008396  0.020181  0.031929 0.0542824
```

```
## beta[6] -0.01760  0.006136  0.018304  0.030585 0.0538376
## beta[7] -0.01641 -0.010733 -0.007719 -0.004753 0.0008937
## sig      0.08177  0.093231  0.100326  0.108193 0.1268669
```

Create an adjacency matrix for the states in the US

```
library(maps)
library(spdep)
library(maptools)
usa.state <- map(database="state",fill=TRUE,plot=FALSE)
# The strsplit function splits the elements of a character vector into substrings
state.ID <- sapply(strsplit(usa.state$names,":"),function(x) x[1])
usa.poly <- map2SpatialPolygons(usa.state,IDs=state.ID)
# The poly2nb function builds a neighbors list from a "polygon" list
usa.nb <- poly2nb(usa.poly)
# The nb2mat function builds a spatial weights matrix for neighbors list
A <- nb2mat(usa.nb,style="B")
A <- A[-8,] # Take out DC
A <- A[,-8]
M <- diag(rowSums(A))
```

Fit the spatial model

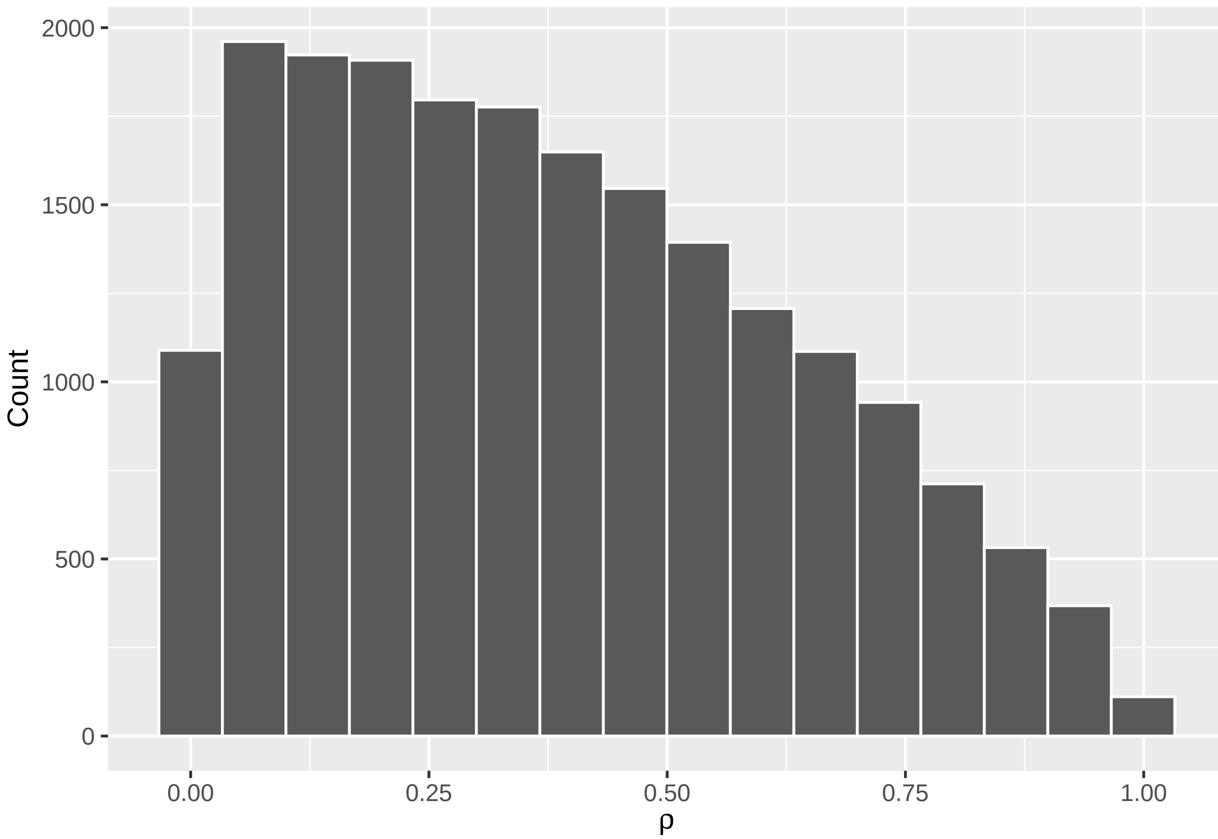
```
sp_model <- "model{
  # Likelihood
    for(i in 1:n){
      Y[i] ~ dnorm(mu[i]+S[i],taue)
    }
    S[1:n] ~ dmnorm(zero[1:n],taus*Omega[1:n,1:n])
    for(i in 1:n){
      mu[i] <- inprod(X[i,],beta[])
      zero[i] <- 0
    }
    Omega[1:n,1:n] <- M[1:n,1:n]-rho*A[1:n,1:n]

  # Priors
    for(j in 1:p){
      beta[j] ~ dnorm(0,0.01)
    }
    taue ~ dgamma(0.1,0.1)
    taus ~ dgamma(0.1,0.1)
    rho ~ dunif(0,1)
    sig[1] <- 1/sqrt(taue)
    sig[2] <- 1/sqrt(taus)
}"

dat <- list(Y=Y,n=n,X=X,A=A,M=M,p=p)
init <- list(rho=0.99,beta=lm(Y~X-1)$coef)
model2 <- jags.model(textConnection(sp_model),inits=init,data=dat,quiet=TRUE)
```

```r
update(model2,10000,progress.bar="none")
samp2 <- coda.samples(model2,variable.names=c("beta","rho","sig"),n.iter=20000,
                      progress.bar="none")
summary(samp2)
```

```
##
## Iterations = 11001:31000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 20000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean       SD  Naive SE Time-series SE
## beta[1]  0.045755 0.112939 7.986e-04      0.0076482
## beta[2]  0.760519 0.106056 7.499e-04      0.0044500
## beta[3] -0.007284 0.026648 1.884e-04      0.0012876
## beta[4] -0.010997 0.021054 1.489e-04      0.0007331
## beta[5]  0.015524 0.023492 1.661e-04      0.0009524
## beta[6]  0.015186 0.025590 1.810e-04      0.0011993
## beta[7] -0.008098 0.005755 4.069e-05      0.0001607
## rho      0.379281 0.250670 1.773e-03      0.0035509
## sig[1]   0.104950 0.014132 9.993e-05      0.0001960
## sig[2]   0.149645 0.026389 1.866e-04      0.0005220
##
## 2. Quantiles for each variable:
##
##             2.5%        25%       50%       75%    97.5%
## beta[1] -0.16014 -0.0304465  0.039668  0.116492 0.280788
## beta[2]  0.55968  0.6896451  0.756840  0.828257 0.979974
## beta[3] -0.06210 -0.0242202 -0.006783  0.010293 0.044540
## beta[4] -0.05394 -0.0245636 -0.010777  0.003069 0.029439
## beta[5] -0.03155  0.0004606  0.015772  0.030816 0.062237
## beta[6] -0.03703 -0.0011419  0.015435  0.032379 0.063756
## beta[7] -0.01920 -0.0119231 -0.008219 -0.004373 0.003521
## rho      0.01558  0.1675387  0.349020  0.563487 0.895657
## sig[1]   0.08132  0.0950696  0.103605  0.113300 0.136616
## sig[2]   0.10638  0.1308305  0.146743  0.165495 0.208839
```

```r
rho <- samp2[[1]][,8]
ggplot(NULL,aes(x=rho))+geom_histogram(bins=16,color="white")+
  labs(x=expression(rho),y="Count")
```

```
## Don't know how to automatically pick scale for object of type mcmc. Defaulting to continuous.
```

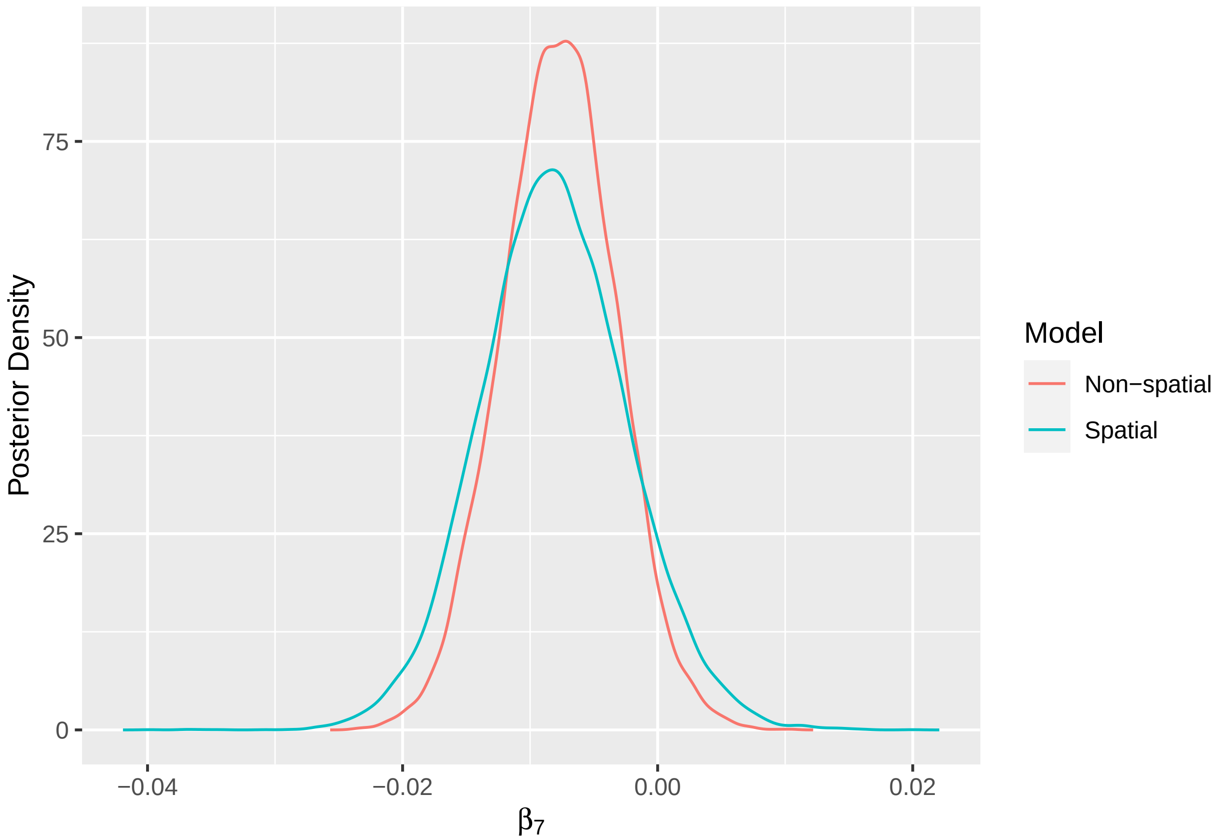


Summary: The spatial dependence parameter is estimated to be near 0.35, indicating weak spatial dependence

Compare the results across models

The objective is to determine if the coefficient corresponding to the number of gun laws, β_7, is non-zero. Below we compare its posterior distribution for the spatial and non-spatial models.

```
b1 <- samp1[[1]][,7]
b2 <- samp2[[1]][,7]
d1 <- density(b1)
d2 <- density(b2)
Model <- c(rep("Non-spatial",length(d1$x)),rep("Spatial",length(d2$x)))
x <- c(d1$x,d2$x)
y <- c(d1$y,d2$y)
density.data <- as.data.frame(x=x,y=y,Model=Model)
ggplot(density.data,aes(x=x,y=y,color=Model))+geom_line()+
  labs(x=expression(beta[7]),y="Posterior Density")
```




```r
mean(b1<0)
```

```
## [1] 0.96055
```

```r
mean(b2<0)
```

```
## [1] 0.9194
```

Summary: Both models provide evidence of a negative relationship between the number of gun laws and the firearm-related death rate. However, there is more uncertainty in the spatial model, which is likely more realistic.