Spatial modeling of gun-related homicide rates

Linear models with correlated data

The data for this analysis come from "Firearm legislation and firearm mortality 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 confounders,

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

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

We compare the usual non-spatial model with $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$ with the spaital model $\text{Cov}(\varepsilon_1, \ldots, \varepsilon_n) \sim \text{Normal}(0, \Sigma)$. The covariance

$$\Sigma = \tau^2 S + \sigma^2 I_n$$

is decomposed into a spatial covariance $\tau^2 S$ and a non-spatial covariance $\sigma^2 I_n$. The spatial covariance follows the conditionally-autoregressive model $S = (M - \rho A)^{-1}$ wheree A is the adjacency matrix with (i, j) element is equal to 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

```
Y      <- log(10000*Y/N)
Z[,1] <- log(Z[,1])
X      <- cbind(1,Z,rowSums(X))

# Remove AK and HI
Y <- Y[-c(2,11)]
X <- X[-c(2,11),]

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

Fit the non-spatial model

```
library(rjags)
```

```
## Loading required package: coda
```

```
## Linked to JAGS 4.3.1
```

```
## Loaded modules: basemod,bugs
```

```r
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.names=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]   0.021214 0.088708 6.273e-04      0.0043749
## beta[2]   0.750225 0.085509 6.046e-04      0.0032627
## beta[3]  -0.002359 0.021341 1.509e-04      0.0008540
## beta[4]  -0.014008 0.016226 1.147e-04      0.0004424
## beta[5]   0.019923 0.018986 1.343e-04      0.0006048
## beta[6]   0.018247 0.019353 1.368e-04      0.0006819
## beta[7]  -0.007798 0.004395 3.108e-05      0.0001021
## sig       0.101678 0.011630 8.224e-05      0.0001456
##
## 2. Quantiles for each variable:
##
##             2.5%       25%       50%       75%     97.5%
## beta[1] -0.15514 -0.037932  0.019935  0.080379 0.1990856
## beta[2]  0.58209  0.693915  0.750716  0.805215 0.9198197
## beta[3] -0.04458 -0.016447 -0.002299  0.011702 0.0395890
## beta[4] -0.04630 -0.024803 -0.013780 -0.003052 0.0173855
## beta[5] -0.01727  0.007615  0.019964  0.032276 0.0579266
```

```
## beta[6] -0.01947  0.005078  0.018305  0.031424 0.0562396
## beta[7] -0.01634 -0.010717 -0.007792 -0.004868 0.0008632
## sig      0.08213  0.093450  0.100578  0.108716 0.1273492
```

Create an adjacency matrix for the states in the US

```r
library(maps)
library(sf)
```

```
## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

```r
library(spdep)
```

```
## Loading required package: spData
```

```
## To access larger datasets in this package, install the spDataLarge
## package with: 'install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')'
```

```r
library(rmapshaper)

# Create the USA state map
usa.state <- map(database = "state", fill = TRUE, plot = FALSE)

# Convert to an sf object
usa.sf <- st_as_sf(map("state", plot = FALSE, fill = TRUE))

# Clean the geometries to fix any issues
usa.sf <- st_make_valid(usa.sf)

# If there are still issues, simplify the geometries
usa.sf <- ms_simplify(usa.sf, keep_shapes = TRUE)

# Create neighborhood structure
usa.nb <- poly2nb(usa.sf)

# Convert to adjacency matrix
A <- nb2mat(usa.nb, style = "B")

# Remove DC (8th row/column)
A <- A[-8, ]
A <- A[, -8]

# Create the diagonal matrix
M <- diag(rowSums(A))
```

Fit the spatial model
```

```r
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)
}"


library(rjags)
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)
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]   0.049679 0.113167 8.002e-04      0.0077615
## beta[2]   0.760789 0.103858 7.344e-04      0.0047302
## beta[3]  -0.008806 0.027471 1.943e-04      0.0012984
## beta[4]  -0.010317 0.021361 1.510e-04      0.0007327
## beta[5]   0.014653 0.024287 1.717e-04      0.0010072
## beta[6]   0.015721 0.024859 1.758e-04      0.0011160
## beta[7]  -0.008352 0.005746 4.063e-05      0.0001743
## rho       0.376625 0.250032 1.768e-03      0.0035176
```

```
## sig[1]     0.104814 0.014003 9.902e-05       0.0002100
## sig[2]     0.149694 0.026505 1.874e-04       0.0005028
##
## 2. Quantiles for each variable:
##
##                  2.5%       25%        50%       75%     97.5%
## beta[1]  -0.16929 -0.0252607  0.051215  0.123892 0.270851
## beta[2]   0.55447  0.6932273  0.759008  0.828118 0.965786
## beta[3]  -0.06182 -0.0273626 -0.009359  0.009617 0.046529
## beta[4]  -0.05254 -0.0245496 -0.010112  0.003542 0.032281
## beta[5]  -0.03438 -0.0010765  0.015144  0.030661 0.061628
## beta[6]  -0.03326 -0.0007065  0.015867  0.032029 0.065108
## beta[7]  -0.01964 -0.0121538 -0.008393 -0.004537 0.003133
## rho       0.01672  0.1637923  0.342251  0.564700 0.886932
## sig[1]    0.08117  0.0949497  0.103480  0.113277 0.135785
## sig[2]    0.10598  0.1308526  0.146892  0.165388 0.209569
```

```
rho <- samp2[[1]][,8]
hist(rho,breaks=100)
```


Histogram of rho
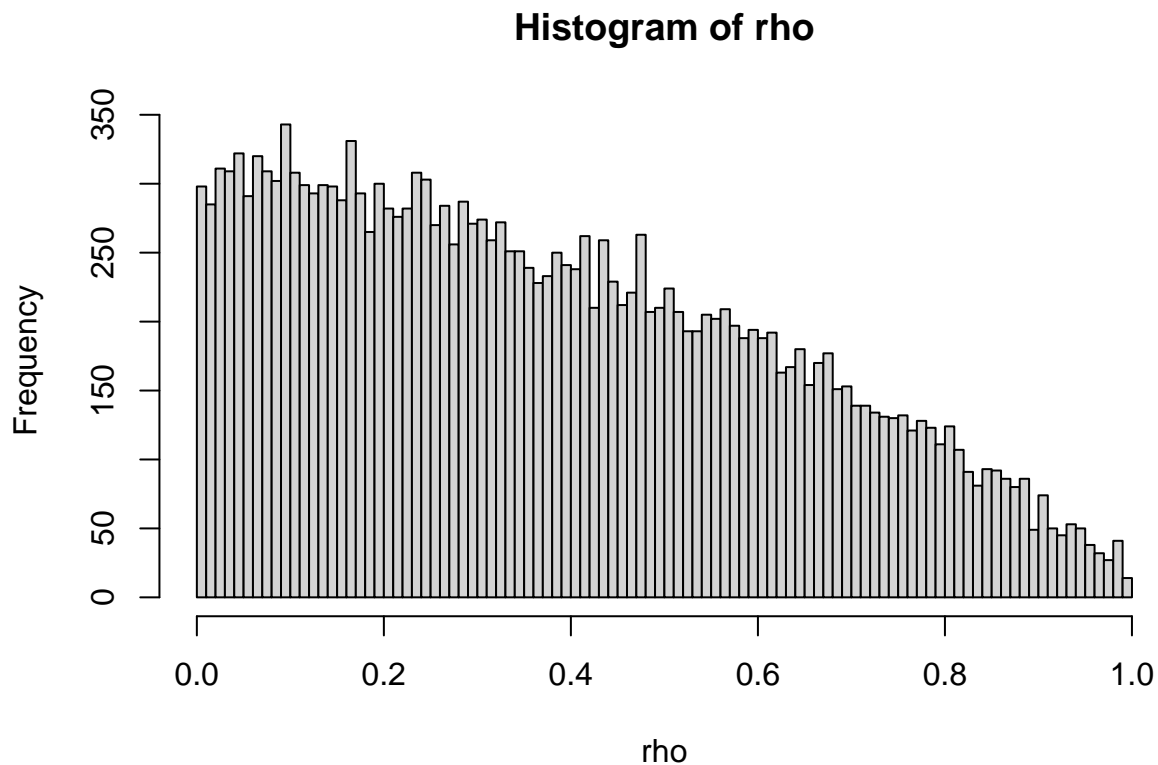
Summary: The spatial dependence parameter is estimated to be near one, indicating strong spatial dependence.

Compare the results across models

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

```
b1  <- samp1[[1]][,7]
 b2  <- samp2[[1]][,7]
 r   <- c(-0.035,0.015)
 d1  <- density(b1,from=r[1],to=r[2])
 d2  <- density(b2,from=r[1],to=r[2])

 plot(NA,xlim=r,ylim=c(0,max(d1$y)),
         xlab="Beta",ylab="Posterior density")
 lines(d1$x,d1$y)
 lines(d2$x,d2$y,lty=2)
 legend("topright",c("Non-spatial","Spatial"),lty=1:2,bty="n",cex=1.5)
```
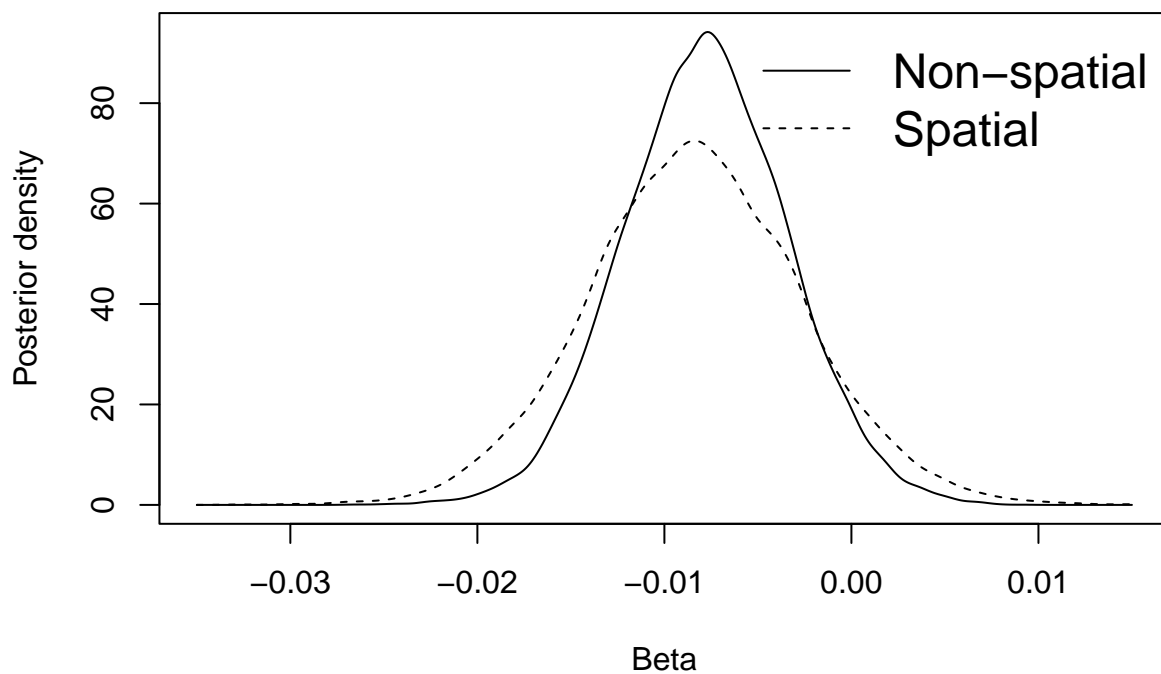


```
 mean(b1<0)
```

```
## [1] 0.9622
```

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

```
## [1] 0.92765
```

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.