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

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

Objective: 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 $S = (M - \rho A)^{-1}$ where 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
library(ggplot2)
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)

 }"

 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]  1.145e-02 0.084051 5.943e-04      4.496e-03
## beta[2]  7.481e-01 0.080446 5.688e-04      2.812e-03
## beta[3]  1.533e-05 0.020255 1.432e-04      7.472e-04
## beta[4] -1.263e-02 0.015963 1.129e-04      4.548e-04
## beta[5]  2.026e-02 0.017771 1.257e-04      5.437e-04
## beta[6]  1.769e-02 0.018814 1.330e-04      6.079e-04
## beta[7] -7.649e-03 0.004312 3.049e-05      9.889e-05
## sig      1.015e-01 0.011523 8.148e-05      1.405e-04
##
## 2. Quantiles for each variable:
##
##             2.5%       25%        50%       75%     97.5%
## beta[1] -0.15844 -0.043488  0.0119550  0.067810 0.1734159
## beta[2]  0.58922  0.695017  0.7492111  0.803012 0.9027495
## beta[3] -0.03897 -0.013773 -0.0003669  0.013413 0.0408235
## beta[4] -0.04443 -0.023116 -0.0125683 -0.002058 0.0186851
## beta[5] -0.01399  0.008362  0.0201080  0.031896 0.0562895
## beta[6] -0.01914  0.005231  0.0176757  0.030174 0.0544307
## beta[7] -0.01622 -0.010462 -0.0076062 -0.004718 0.0006863
```

```
## sig      0.08203  0.093433  0.1004782  0.108515 0.1269191
```

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))
```

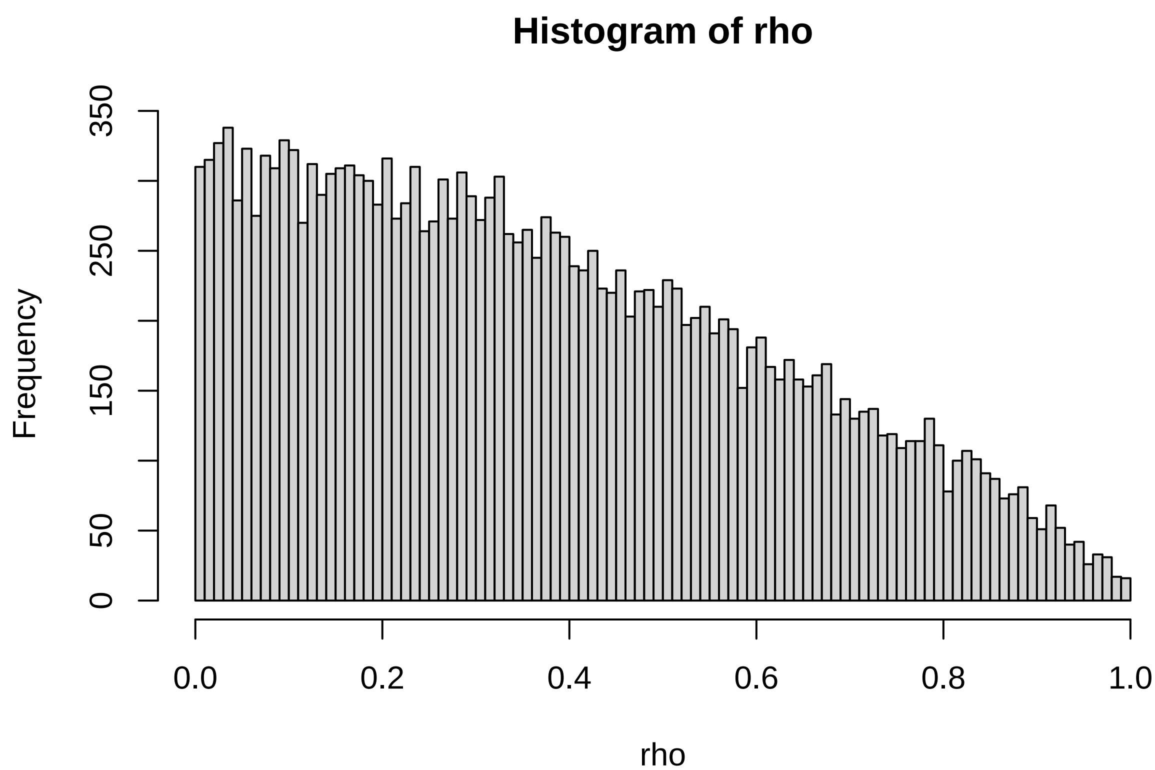
Fit the spatial model

```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)
  }"

  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]  0.042300 0.116704 8.252e-04      0.0086964
## beta[2]  0.755409 0.106192 7.509e-04      0.0050309
## beta[3] -0.007245 0.027262 1.928e-04      0.0012877
## beta[4] -0.009619 0.020608 1.457e-04      0.0006513
## beta[5]  0.014667 0.023664 1.673e-04      0.0009645
## beta[6]  0.016603 0.025705 1.818e-04      0.0012383
## beta[7] -0.008376 0.005715 4.041e-05      0.0001624
## rho      0.369467 0.246625 1.744e-03      0.0034671
## sig[1]   0.104583 0.013949 9.864e-05      0.0001785
## sig[2]   0.150161 0.026556 1.878e-04      0.0005469
```

```
##
## 2. Quantiles for each variable:
##
##              2.5%        25%       50%       75%    97.5%
## beta[1] -0.17449 -0.0391792  0.038075  0.118476 0.283996
## beta[2]  0.55121  0.6843267  0.753195  0.823345 0.972330
## beta[3] -0.06337 -0.0242785 -0.006577  0.010672 0.045646
## beta[4] -0.05044 -0.0232337 -0.009903  0.004117 0.030872
## beta[5] -0.03171 -0.0014322  0.014944  0.030488 0.060527
## beta[6] -0.03451 -0.0005287  0.016865  0.033625 0.067481
## beta[7] -0.01965 -0.0122196 -0.008369 -0.004593 0.002891
## rho      0.01585  0.1618265  0.334456  0.548208 0.882068
## sig[1]   0.08096  0.0948630  0.103313  0.112819 0.135750
## sig[2]   0.10690  0.1312643  0.147104  0.165820 0.210375
```

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



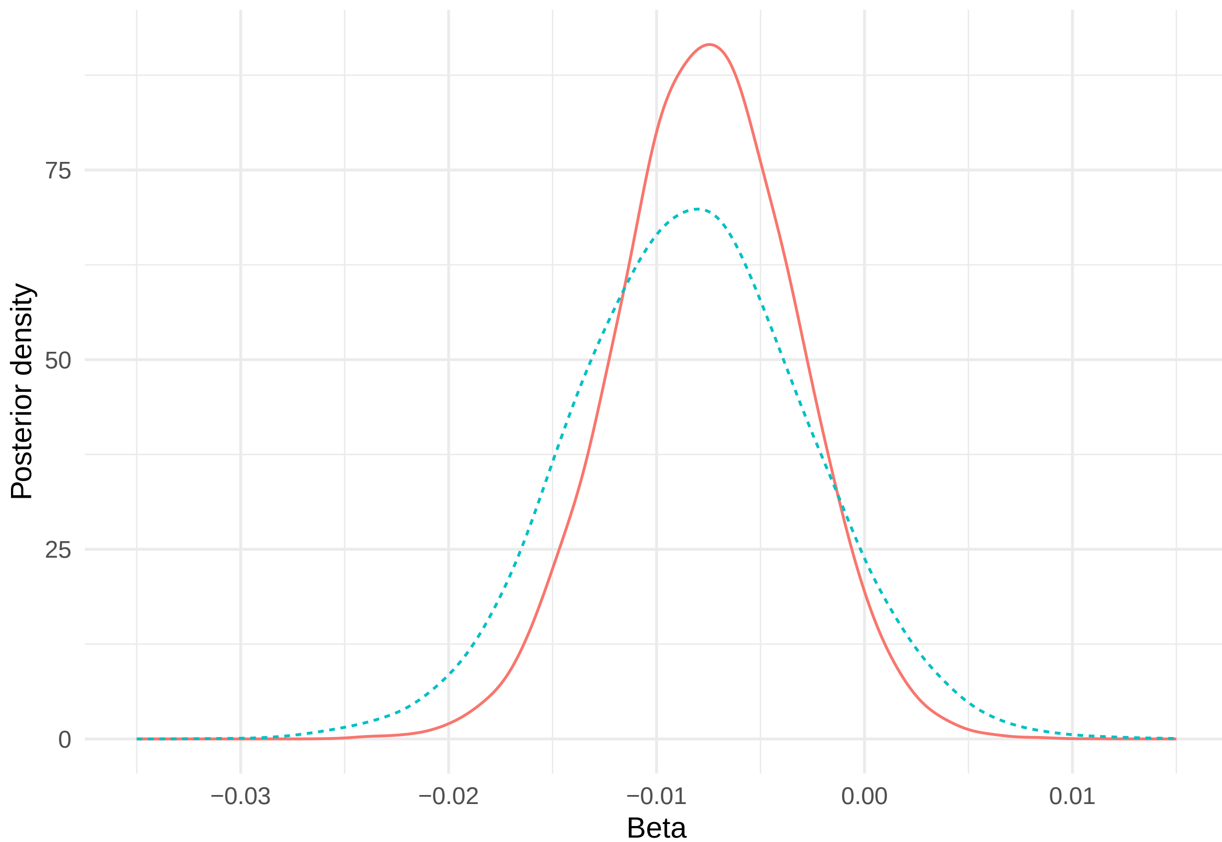
Conclusion: The spatial dependence parameter is close to one, which implies a strong spatial dependence.

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. So, we compare its posterior distribution for the spatial and non-spatial models.

```r
b1  = samp1[[1]][,7]
b2  = samp2[[1]][,7]
r   = c(-0.035,0.015)
# Combine data into a data frame for ggplot2
data <- data.frame(
  Beta = c(b1, b2),
  Group = factor(rep(c("Non-spatial", "Spatial"), c(length(b1), length(b2))))
)
ggplot(data, aes(x = Beta, color = Group, linetype = Group)) +
  geom_density(adjust = 1.5) +
  xlim(r) +
  labs(x = "Beta", y = "Posterior density") +
  theme_minimal() +
  theme(legend.position = "topright")
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_density()‘).
```




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

```
## [1] 0.9643
```

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

```
## [1] 0.92905
```

Based on the analysis comparing non-spatial and spatial models for predicting gun-related homicide rates, we can draw several conclusions:

1. **Negative Relationship**: Both the non-spatial and spatial models indicate a negative relationship between the number of gun laws and firearm-related death rates. This suggests that as the number of gun control laws increases, the rate of firearm-related deaths decreases.

2. **Posterior Distributions**: The posterior distributions for the coefficient corresponding to the number of gun laws (β_7) are shown for both models. The density plots reveal that the distribution for the spatial model is slightly more spread out, indicating higher uncertainty compared to the non-spatial model.

3. **Spatial Dependence**: The spatial dependence parameter (ρ) was estimated to be near one, indicating strong spatial dependence among the states. This means that the firearm-related death rates are not independent across states but are influenced by neighboring states' rates.

4. **Model Comparison**: While both models suggest a negative effect of gun laws on firearm-related deaths, the spatial model, which accounts for correlations between neighboring states, provides a more realistic representation by incorporating spatial dependence. The increased uncertainty in the spatial model is expected due to this additional complexity.

5. **Policy Implications**: The findings support the effectiveness of gun control laws in reducing firearm-related deaths. However, policymakers should consider the spatial dependence highlighted by the spatial model, as it suggests that the impact of gun laws in one state can influence neighboring states.