Lasso

John Tipton

May 4, 2015

First, let's load the packages and some helper functions

```
##
## libraries and functions
##

library(statmod)
source('~/Linear-Model/dinvgamma.R')
source('~/Linear-Model/rMVN.R')

source("~/Linear-Model/bayesianLassoRegression/fixedEffectModel/mcmc.lm.lasso.R")

make.model.plot <- function(out){
  n.burn <- floor(n.mcmc / 5) + 1
  layout(matrix(1:9, 3))
  matplot(t(out$beta.save[, n.burn:n.mcmc]), type = 'l')
  hist(out$beta.save[2,][n.burn:n.mcmc], main = 'Posterior of Beta2')
  abline(v = beta[2], col = 'red')
  plot(out$sigma.squared.epsilon.save[n.burn:n.mcmc], type = 'l')
  abline(h = sigma.squared.epsilon, col = 'red')
  plot(out$lambda.squared.save, type = 'l')
}
```

Then we simulate some data

```
##
## Simulate some data
##

N <- 1000
n <- 100
beta <- -3:3
sigma.squared.epsilon <- 0.25
tau <- length(beta)

make.lm.data <- function(N, n, beta, sigma.sqaured.epsilon){
  tau <- length(beta)
  X <- matrix(nrow = N, ncol = tau)
  for(i in 1:tau){
    X[, i] <- rnorm(N, 0, 1)
  }
    Y <- X %*% beta + rnorm(N, 0, sigma.squared.epsilon)
```

```
  data.frame(Y, X)
}

data <- make.lm.data(N, n, beta, sigma.squared.epsilon)
```

Subsample the data

```
samp <- sample(1:N, n)
data.samp <- data[samp, ]
```

Examine a linear regression model

```
lm(Y ~ . ,data = data)
```

```
##
## Call:
## lm(formula = Y ~ ., data = data)
##
## Coefficients:
## (Intercept)           X1           X2           X3           X4
##   -0.009582    -2.998608    -1.996605    -1.001350     0.001352
##          X5           X6           X7
##    1.002206     2.004336     2.999200
```

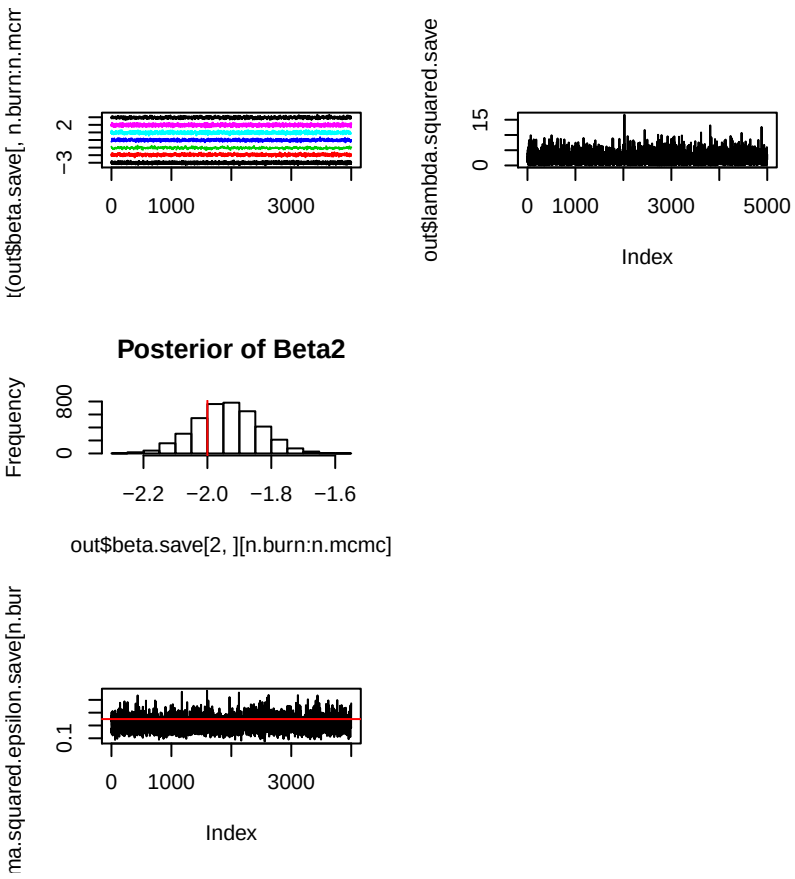
Specify priors for a Bayesian model

```
##
## Setup priors
##

# hyperparameters for mu.beta and sigma.squared.beta
alpha.epsilon <- 1
beta.epsilon <- 1
alpha.lambda <- 1
beta.lambda <- 1

n.mcmc <- 5000

##
## Fit mcmc
##

Y <- data.samp[, 1]
X <- as.matrix(data.samp[, 2:(tau + 1)], ncol = tau)

out <- mcmc.lm.lasso(Y, X, n.mcmc, alpha.epsilon, beta.epsilon, alpha.lambda, beta.lambda)
```

```
##   100  200  300  400  500  600  700  800  900 1000 1100 1200 1300 1400 1500 1600 1700 1800
```

Examine model output

```
make.model.plot(out)
```



Examine estimates $\hat{\beta}$

```
library(pander)
results=data.frame(rbind(c(beta), c(rowMeans(out$beta))), row.names=c("Truth", "Estimate"))
names(results)=c("Beta1", "Beta2", "Beta3", "Beta4", "Beta5", "Beta6", "Beta7")
pandoc.table(results, style="rmarkdown")
```

```
##
##
## |           |  Beta1  |  Beta2  |  Beta3  |  Beta4   |  Beta5  |
## |:--------------:|:-------:|:-------:|:-------:|:--------:|:-------:|
## |   **Truth**    |   -3    |   -2    |   -1    |    0     |    1    |
## |  **Estimate**  | -3.002  | -1.939  |  -1.04  | -0.01437 | 0.9778  |
##
## Table: Table continues below
##
##
##
```

```
## |           |  Beta6  |  Beta7  |
## |:--------------:|:-------:|:-------:|
## |   **Truth**    |    2    |    3    |
## |  **Estimate**  |  1.971  |  2.967  |
```