Bayesian Linear Models

Andy Shen, Devin Francom

LANL: CCS-6

Task

Let $\boldsymbol{y} = (y_1, \ldots, y_n)$ be the regression response and $\boldsymbol{X}$ be a $n \times p$ matrix of covariates. Here is the traditional linear model likelihood:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

The Bayesian version just needs priors for the unknowns, which are $\boldsymbol{\beta}$ and σ^2, and then you just "turn the Bayesian crank", which means you multiply the likelihood and priors, get the full conditionals, and sample the posterior with MCMC. Use $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{I})$ and $\sigma^2 \sim IG(a, b)$ as priors, and derive the following:

$$\begin{aligned}
p(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{y}) &\propto ??\\
p(\boldsymbol{\beta} | \sigma^2, \boldsymbol{X}, \boldsymbol{y}) &\propto ??\\
p(\sigma^2 | \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{y}) &\propto ??
\end{aligned}$$

Hint: the full conditionals will be recognizable distributions (conjugate).

1 Posterior Distributions

Prior Distributions

We know that $\boldsymbol{\beta} \sim N(\mathbf{0}, \tau^2 \boldsymbol{I})$ and $\sigma^2 \sim IG(a, b)$.

Moreover,

$$
\begin{aligned}
p(\sigma^2) &= \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} exp\bigg[\frac{\beta}{\sigma^2}\bigg] \\
&\propto (\sigma^2)^{-\alpha-1} exp\bigg[\frac{\beta}{\sigma^2}\bigg]
\end{aligned}
$$

and

$$
p(\boldsymbol{\beta}) = (2\pi\tau^2)^{-\frac{p}{2}} exp\bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\bigg] \propto (\tau^2)^{-\frac{p}{2}} exp\bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\bigg]
$$

Likelihood Function

We know that $y_i \sim N(0, \sigma^2)$ and $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$.

Then,

$$
p(\boldsymbol{\beta} \mid \cdot) \propto N(\mathbf{0}, \tau^2 \ \mathbf{I}) \ L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y})
$$

and

$$
p(\sigma^2 \mid \cdot) \propto IG(a, b) \ L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y})
$$

Leveraging the fact that $\mathbb{E}(\boldsymbol{y}) = \boldsymbol{X\beta}$, the derivation of the likelihood is as follows (shoutout to Stats 100C Lec 3, Spring 2020):

$$
L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) = \prod_{i=1}^{N} p(y_i | \boldsymbol{\beta}, \sigma^2)
$$

$$
= \prod_{i=1}^{N} (2\pi\sigma^2) exp\bigg[-\frac{1}{2\sigma^2}(y_i - \mu_i)^2\bigg]
$$

$$
= (2\pi\sigma^2)^{-\frac{n}{2}} exp\bigg[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mu_i)^2\bigg]
$$

$$
\propto (\sigma^2)^{-\frac{n}{2}} exp\bigg[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mu_i)^2\bigg]
$$

$$
= (\sigma^2)^{-\frac{n}{2}} exp\bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})\bigg]
$$

Full Conditional Distributions

Multiplying the priors by the likelihood function, we get that:

$$p(\sigma^2 \mid \cdot) \propto (\sigma^2)^{-\alpha-1} exp\Bigg[-\frac{b}{\sigma^2}\Bigg](\sigma^2)^{-\frac{n}{2}} exp\Bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Bigg]$$

$$= (\sigma^2)^{-\alpha-1-\frac{n}{2}} exp\Bigg[-\frac{b}{\sigma^2}-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Bigg]$$

$$= (\sigma^2)^{-\alpha-1-\frac{n}{2}} exp\Bigg[-\frac{1}{\sigma^2}\Bigg(\frac{2b+(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})}{2}\Bigg)\Bigg]$$

For the full posterior distribution, we have that

$$p(\boldsymbol{\beta}|\cdot) \propto exp\Bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Bigg]$$

$$\propto exp\Bigg[-\frac{1}{2\tau^2\sigma^2}\big[\sigma^2\boldsymbol{\beta}'\boldsymbol{\beta}+\tau^2(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\big]\Bigg]$$

$$\propto exp\Bigg[-\frac{1}{2\tau^2\sigma^2}\big[\sigma^2\boldsymbol{\beta}'\boldsymbol{\beta}+\tau^2(\boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y}+\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{X\beta})\big]\Bigg]$$

and

$$p(\boldsymbol{\beta},\sigma^2 \mid \boldsymbol{y},\boldsymbol{X}) \propto p(\sigma^2)\ p(\boldsymbol{\beta})\ L(\boldsymbol{\beta},\sigma^2|\boldsymbol{y})$$

$$\propto (\sigma^2)^{-\alpha-1} exp\Bigg[-\frac{b}{\sigma^2}\Bigg](\tau^2)^{-\frac{p}{2}} exp\Bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\Bigg](\sigma^2)^{-\frac{n}{2}} exp\Bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Bigg]$$

$$\propto (\sigma^2)^{-(\frac{n}{2}+\alpha+1)}(\tau^2)^{-\frac{p}{2}} exp\Bigg[-\frac{b}{\sigma^2}\Bigg] exp\Bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\Bigg] exp\Bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Bigg]$$

$$p(\boldsymbol{\beta},\sigma^2 \mid \boldsymbol{y},\boldsymbol{X}) \propto (\sigma^2)^{-(\frac{n}{2}+\alpha+1)}(\tau^2)^{-\frac{p}{2}} exp\Bigg[-\frac{b}{\sigma^2}\Bigg] exp\Bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\Bigg] exp\Bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Bigg]$$

The joint posterior distribution is effectively an inverse gamma distribution multiplied by two multivariate normal distributions.

From these results, we get that:

$$\sigma^2 \mid \cdot \sim IG\Bigg(\alpha+\frac{n}{2},\quad \frac{2b+(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})}{2}\Bigg)$$

and

$$\boldsymbol{\beta} \mid \cdot \sim \mathcal{N}\Bigg\{\Bigg(\boldsymbol{X}'\boldsymbol{X}+\frac{\sigma^2}{\tau^2}\boldsymbol{I}\Bigg)^{-1}\boldsymbol{X}'\boldsymbol{y},\Bigg(\frac{1}{\sigma^2}(\boldsymbol{X}'\boldsymbol{X})+\frac{1}{\tau^2}\boldsymbol{I}\Bigg)^{-1}\Bigg\}$$

2 MCMC Sampler

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
dat <- read.csv("data.csv")

y <- dat$y
X <- dat[,-c(1,2)] %>% as.matrix()
X <- cbind(1, X)
n <- nrow(X) # n
p <- ncol(X) - 1# p = 11, p + 1 = 12
```

Sigma

$$p(\sigma^2 \mid \cdot) \propto (\sigma^2)^{-\alpha-1-\frac{n}{2}} exp\left[-\frac{1}{\sigma^2}\left(\frac{2b + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2}\right)\right]$$

$$\sigma^2 \mid \cdot \sim IG\left(\alpha + \frac{n}{2},\quad \frac{2b + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2}\right)$$

We code the full conditional of σ^2 as follows:

```r
p_sig <- function(a = 1, b = 1, n = nrow(X), beta) {
  a_term <- a + (n/2)
  b_term <- 0.5 * (2*b + (t(y - (X %*% beta)) %*% (y - (X %*% beta)))) #y, X defined above
  1 / rgamma(1, shape = a_term, rate = 1/b_term)
}
```

Beta

The full conditional distribution of $\boldsymbol{\beta}$ is effectively the product of two multivariate normal distributions.

$$p(\boldsymbol{\beta}|\cdot) \propto exp\left[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta} - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right]$$

$$\propto (\tau^2)^{-\frac{p}{2}} exp\left[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right](\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right]$$

which is proportional to

$$\propto exp((\boldsymbol{\beta} - \boldsymbol{\mu_\beta})'\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\ (\boldsymbol{\beta} - \boldsymbol{\mu_\beta}))$$

After solving for the values of μ_β and $\mathbf{\Sigma}_\beta$, we get that

$$\boldsymbol{\beta} \mid \cdot \sim \mathcal{N}\left\{ \left(\boldsymbol{X'X} + \frac{\sigma^2}{\tau^2}\boldsymbol{I} \right)^{-1} \boldsymbol{X'y}, \left(\frac{1}{\sigma^2}(\boldsymbol{X'X}) + \frac{1}{\tau^2}\boldsymbol{I} \right)^{-1} \right\}$$

This is a p-variate normal distribution with mean vector and covariance matrix listed above.

The code for the full conditional distribution of $\boldsymbol{\beta}$ is as follows:

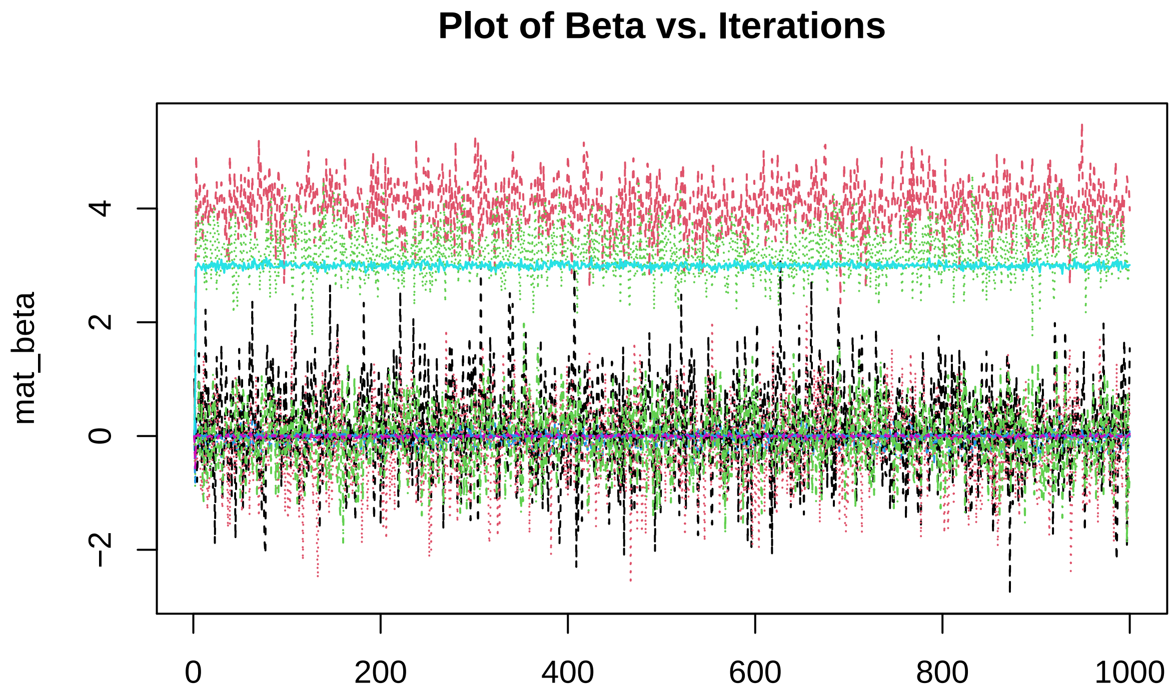
```
library(mvtnorm)
p_beta <- function(sig_sq, tau_sq = 1, p = ncol(X)) {
  sig <- solve( (1/sig_sq) * (t(X) %*% X) + (1/tau_sq) * diag(p) )
  mu <- (1/sig_sq) * sig %*% t(X) %*% y
  rmvnorm(1, mean = mu, sigma = sig)
}
```

Gibbs Sampler

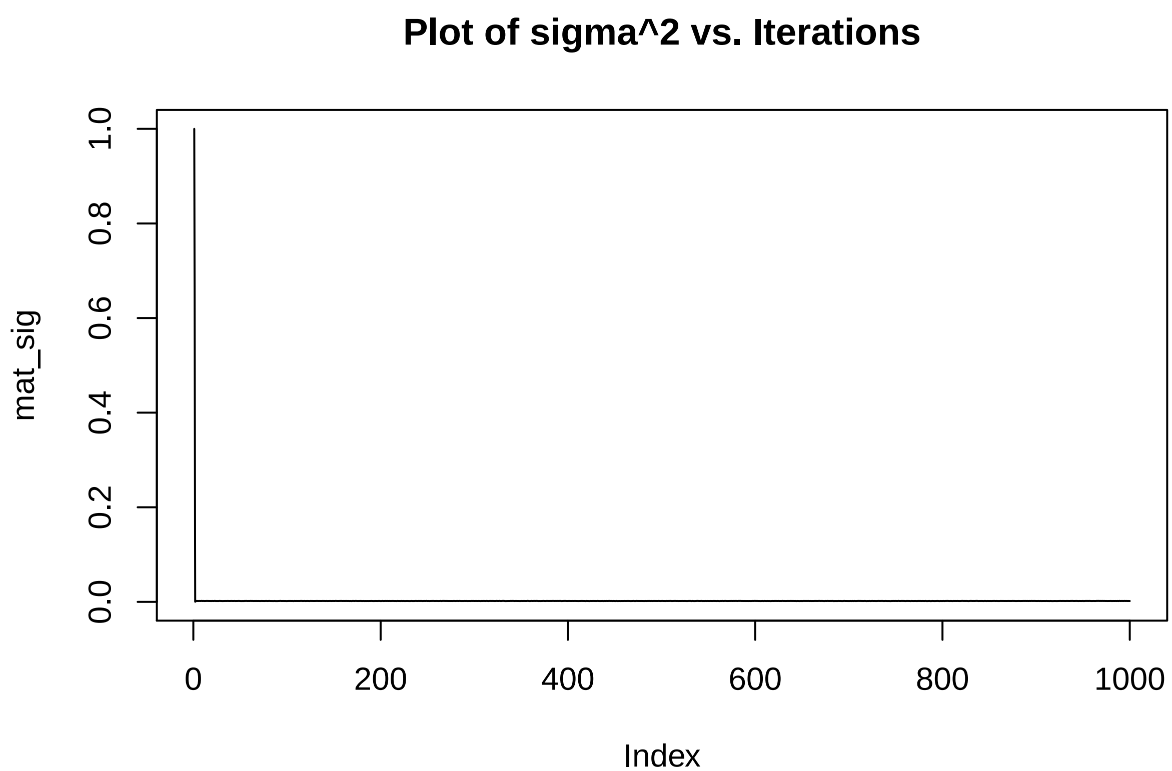
```
set.seed(12)
gibbs <- function(its) {
  mat_beta <- matrix(NA, its, p+1)
  mat_sig <- rep(NA, its)
  mat_sig[1] <- 1
  mat_beta[1,] <- c(1, rep(0, p))
  for(i in 2:its) {
    mat_beta[i,] <- p_beta(sig_sq = mat_sig[i-1])
    mat_sig[i] <- p_sig(beta = mat_beta[i,])
  }
  list(mat_beta, mat_sig)
}

its <- 1000
a <- gibbs(its = its)
mat_beta <- a[[1]]
mat_sig <- a[[2]]
```

```
matplot(mat_beta, type = "l", main = "Plot of Beta vs. Iterations")
```




```r
plot(mat_sig, type = "l", main = "Plot of sigma^2 vs. Iterations")
```



We know that:

$$s^2 = \frac{S(\hat{\beta})}{n-p-1}$$

```r
mod <- lm(y ~ 0 + X)
res <- mod$residuals
S_beta <- t(res) %*% res
s_sq <- S_beta / (n - p - 1) #or just 1000 - 12
s_sq
```

```
##              [,1]
## [1,] 0.0001086339
```

```r
mean(mat_sig)
```

```
## [1] 0.002870207
```

```r
colMeans(mat_beta)
```

```
##  [1]  1.295607e-02  4.033032e+00  3.268582e+00 -4.772263e-02 -8.423255e-05
##  [6]  6.407258e-06  1.808754e-01 -1.691941e-01 -1.767072e-02  5.358390e-03
## [11]  2.991415e+00 -4.076436e-03
```

```r
mod$coefficients
```

```
##             X            Xa            Xb            Xc          Xx_n
## -2.436116e-03  5.048074e+00  4.054884e+00 -4.004919e-02  2.562824e-04
##          Xx_m         Xvel1         Xvel2         Xvel3           XG1
## -3.616514e-06  7.584315e-01 -3.210139e-01 -5.395596e-02  4.487446e-03
##       Xdelta2       Xdelta3
##  3.003534e+00 -4.994286e-03
```