

Bayesian Linear Models

Andy Shen, Devin Francom

LANL: CCS-6

Task

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

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

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 $\boldsymbol{\beta} \sim N(\mathbf{0}, \tau^2 \mathbf{I})$ and $\sigma^2 \sim IG(a, b)$ as priors, and derive the following:

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

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

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

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

1 Posterior Distributions

Prior Distributions

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

Moreover,

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

and

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

Likelihood Function

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

Then,

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

and

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

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

$$\begin{aligned} L(\beta, \sigma^2 \mid \mathbf{y}) &= \prod_{i=1}^N p(y_i \mid \beta, \sigma^2) \\ &= \prod_{i=1}^N (2\pi\sigma^2) \exp\left[-\frac{1}{2\sigma^2} (y_i - \mu_i)^2\right] \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu_i)^2\right] \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu_i)^2\right] \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta)\right] \end{aligned}$$

Full Conditional Distributions

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

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

and

$$\begin{aligned}
p(\beta \mid \cdot) &\propto \exp\left[-\frac{1}{2\tau^2}\beta'\beta - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right] \\
&\propto \exp\left[-\frac{1}{2\tau^2\sigma^2}[\sigma^2\beta'\beta + \tau^2(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)]\right] \\
&\propto \exp\left[-\frac{1}{2\tau^2\sigma^2}[\sigma^2\beta'\beta + \tau^2(\mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta)]\right]
\end{aligned}$$

and

$$\begin{aligned}
p(\beta, \sigma^2 \mid \mathbf{y}, \mathbf{X}) &\propto p(\sigma^2) p(\beta) L(\beta, \sigma^2 \mid \mathbf{y}) \\
&\propto (\sigma^2)^{-\alpha-1} \exp\left[-\frac{b}{\sigma^2}\right] (\tau^2)^{-\frac{p}{2}} \exp\left[-\frac{1}{2\tau^2}\beta'\beta\right] (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right] \\
&\propto (\sigma^2)^{-(\frac{n}{2}+\alpha+1)} (\tau^2)^{-\frac{p}{2}} \exp\left[-\frac{b}{\sigma^2}\right] \exp\left[-\frac{1}{2\tau^2}\beta'\beta\right] \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right] \\
p(\beta, \sigma^2 \mid \mathbf{y}, \mathbf{X}) &\propto (\sigma^2)^{-(\frac{n}{2}+\alpha+1)} (\tau^2)^{-\frac{p}{2}} \exp\left[-\frac{b}{\sigma^2}\right] \exp\left[-\frac{1}{2\tau^2}\beta'\beta\right] \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right]
\end{aligned}$$

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

From these results, we get that:

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

and

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

2 MCMC Sampler

```
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

dat <- read.csv("data.csv")
dat2 <- dat[,-1]

y <- dat[, "y"]
X <- dat[,-c(1,2)] %>% as.matrix()
X <- cbind(1, X)
nrow(X) # n

## [1] 1000
ncol(X) # p = 11, p + 1 = 12

## [1] 12
```

Sigma

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

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

```
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.

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

$$\propto \exp((\beta - \mu_\beta)' \Sigma_\beta^{-1} (\beta - \mu_\beta))$$

After solving for the values of μ_β and Σ_β , we get that

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

This is a p-variate normal distribution with mean $\left[(\mathbf{X}'\mathbf{X})^{-1} + \frac{\sigma^2}{\tau^2}\mathbf{I}\right]^{-1} \mathbf{X}'\mathbf{y}$ and covariance matrix $\left(\frac{1}{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1} + \frac{1}{\tau^2}\mathbf{I}\right)^{-1}$.

The code for the full conditional distribution of β is as follows:

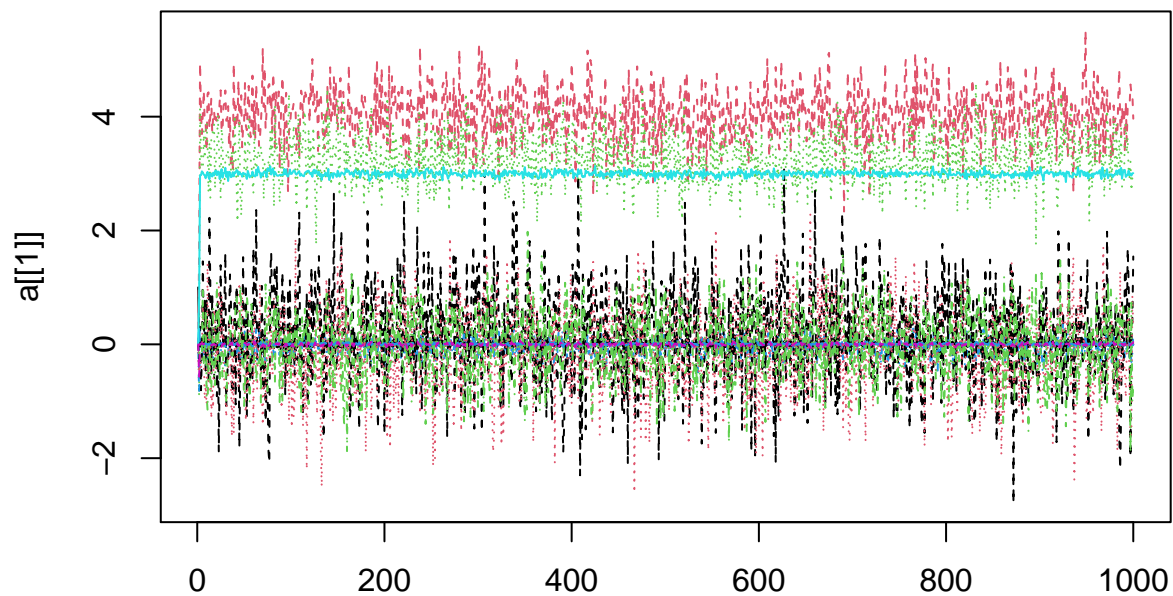
```
library(mvtnorm)
p_beta <- function(sig_sq, tau_sq = 1, p = ncol(X)) {
  sig <- solve( (1/sig_sq) * (t(X) %*% X) + (1/tau_sq) * diag(p) )
  mu <- (1/sig_sq) * sig %*% t(X) %*% y
  rmvnorm(1, mean = mu, sigma = sig)
}
```

Gibbs Sampler

```
set.seed(12)
gibbs <- function(its, p = ncol(X)) {
  mat_beta <- matrix(NA, its, p)
  mat_sig <- rep(NA, its)
  mat_sig[1] <- 1
  mat_beta[1,] <- c(1, rep(0, p-1))
  for(i in 2:its) {
    mat_beta[i,] <- p_beta(sig_sq = mat_sig[i-1])
    #browser()
    mat_sig[i] <- p_sig(beta = mat_beta[i,])
  }
  list(mat_beta, mat_sig)
}
its <- 1000
a <- gibbs(its = its)

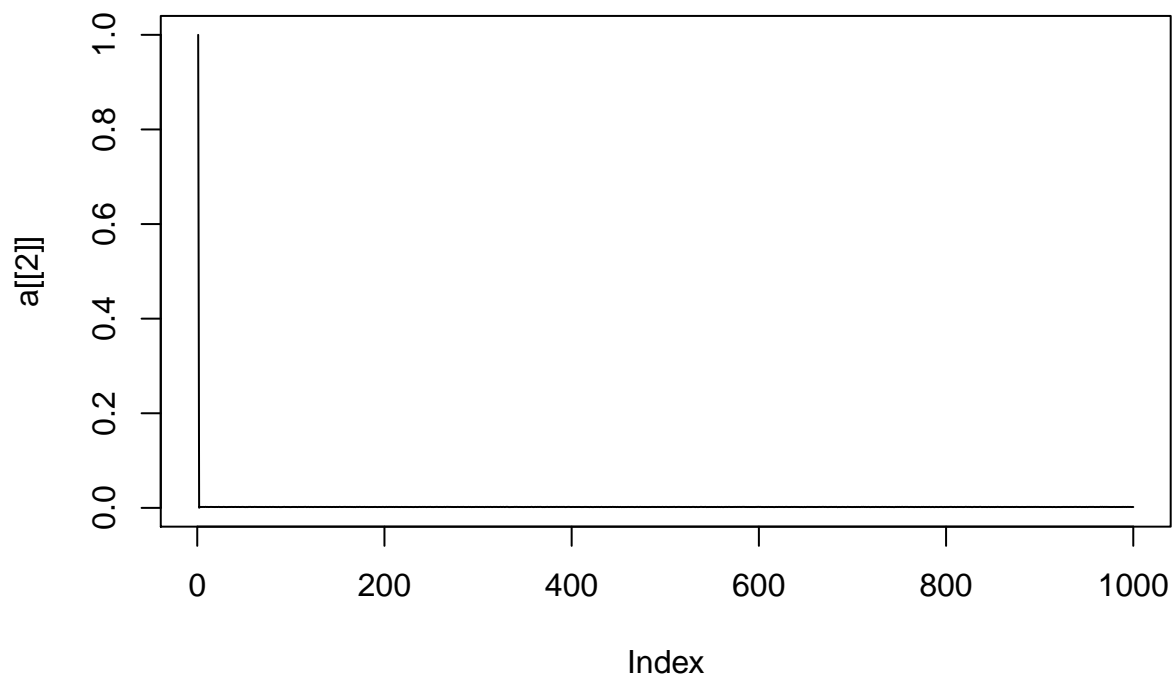
matplot(a[[1]], type = "l", main = "Plot of Beta vs. Iterations")
```

Plot of Beta vs. Iterations



```
plot(a[[2]], type = "l", main = "Plot of sigma^2 vs. Iterations")
```

Plot of sigma^2 vs. Iterations



We know that:

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

```

mod <- lm(y ~ ., data = dat2)
mod2 <- lm(y ~ 0 + X)
all(mod$coefficients == mod2$coefficients)

## [1] TRUE

res <- mod$residuals
S_beta <- t(res) %*% res
s_sq <- S_beta / (nrow(X) - (ncol(X) - 1) - 1) #or just 1000 - 12
s_sq #same thing from summary table

##           [,1]
## [1,] 0.0001086339
summary(mod)$sigma^2 #from summary table

## [1] 0.0001086339
mean(a[[2]])

## [1] 0.002870207
colMeans(a[[1]])

## [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

mod$coefficients

##      (Intercept)          a          b          c          x_n
## -2.436116e-03  5.048074e+00  4.054884e+00 -4.004919e-02  2.562824e-04
##           x_m          vel1          vel2          vel3          G1
## -3.616514e-06  7.584315e-01 -3.210139e-01 -5.395596e-02  4.487446e-03
##           delta2          delta3
##  3.003534e+00 -4.994286e-03

```