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,

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} exp\Bigg[\frac{\beta}{\sigma^2}\Bigg]$$
$$\propto (\sigma^2)^{-\alpha-1} exp\Bigg[\frac{\beta}{\sigma^2}\Bigg]$$

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 $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$.

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

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

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

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

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

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

Joint Posterior

For the full posterior, we have that

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

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

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

The joint posterior distribution is effectively an inverse gamma distribution multiplied by two multivariate normal distributions. The full conditionals can also be derived from the joint posterior.

Full Conditional Distributions

Sigma^2

Multiplying the priors by the likelihood function, or looking at the joint posterior, we get that:

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

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

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

Therefore we have that

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

Beta

For the full conditional of $\boldsymbol{\beta}$, we have that

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

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

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

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

which we set proportional to

$$\propto exp\bigg[-\frac{1}{2\sigma_\beta^2}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})'\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\ (\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})\bigg]$$
$$\propto exp\bigg[-\frac{1}{2\sigma_\beta^2}\big[\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\big]\bigg]$$

From these results, we get that

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

2 MCMC Sampler

```
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")

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

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

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

```
p_sig <- function(a = 0, b = 0, 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 = b_term)
}
```

Beta

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

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

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

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:

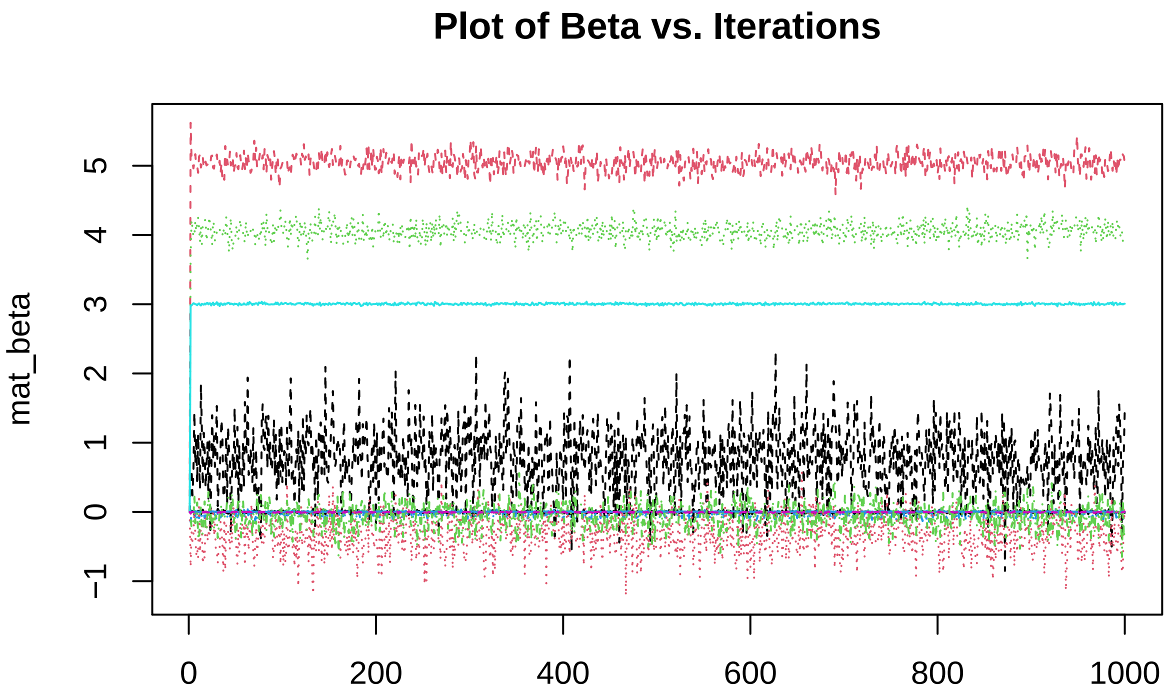
```
library(mvtnorm)
p_beta <- function(sig_sq, tau_sq = 100, p = ncol(X)-1) {
  sig <- solve( (1/sig_sq) * (t(X) %*% X) + (1/tau_sq) * diag(p+1) )
  mu <- (1/sig_sq) * sig %*% t(X) %*% y
  rmvnorm(1, mean = mu, sigma = sig)
}
```

Gibbs Sampler

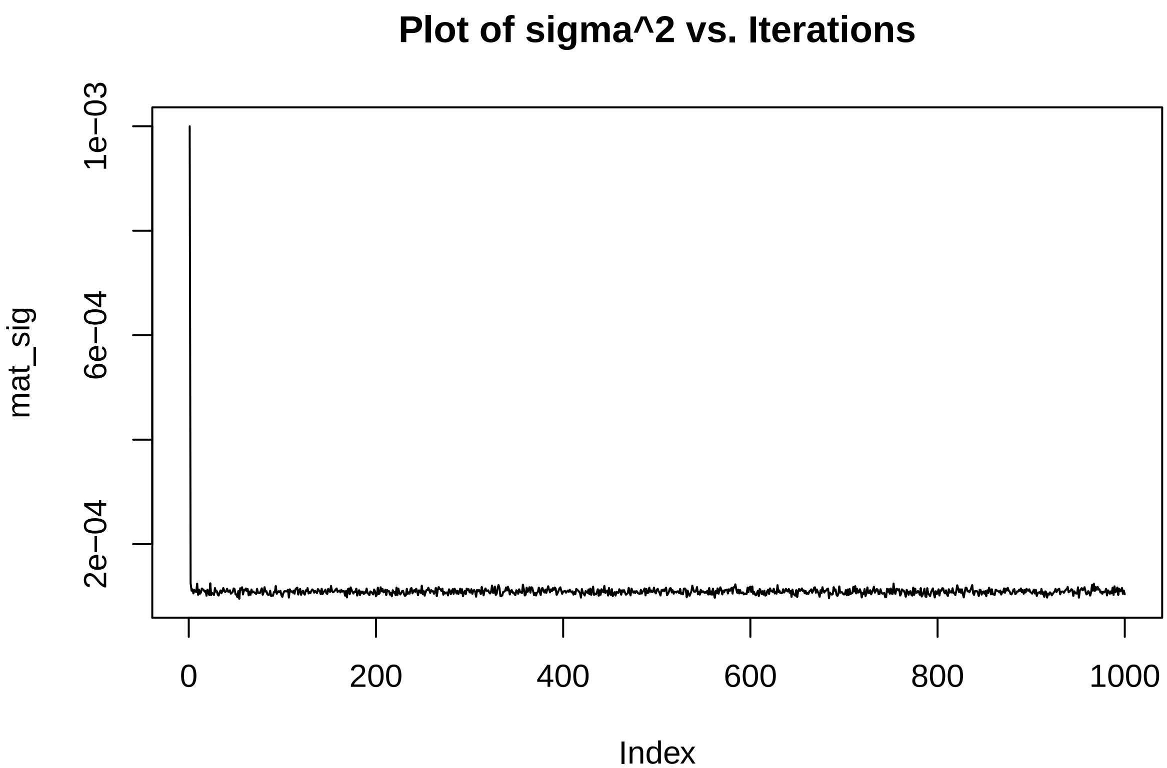
```r
set.seed(12)
gibbs <- function(its) {
  mat_beta <- matrix(NA, its, p+1)
  mat_sig <- rep(NA, its)
  mat_sig[1] <- 0.001
  mat_beta[1,] <- rep(0, p+1)
  for(i in 2:its) {
    mat_beta[i,] <- p_beta(sig_sq = mat_sig[i-1])
    mat_sig[i] <- p_sig(beta = mat_beta[i,])
  }
  list(mat_beta, mat_sig)
}

its <- 1000
a <- gibbs(its = its)
mat_beta <- a[[1]] #beta values
mat_sig <- a[[2]] #sig^2 values
```

Results

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




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



Thanks to Stats 100C, we know that:

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

```
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[,,drop = TRUE]
```

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

```
mean(mat_sig)
```

```
## [1] 0.000109914
```

Our sampled σ^2 values closely match the results from the frequentist `lm()` command.

Alternatively, s^2 can be found with

```
(summary(mod)$sigma)^2
```

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

The data frame below shows our sampled values of $\boldsymbol{\beta}$ vs. the output from `lm()`.

```
beta_df <- data.frame(
  "bayesian_vals" = colMeans(mat_beta),
  "freq_vals" = mod$coefficients
)
beta_df
```

```
##         bayesian_vals     freq_vals
## X       -2.343123e-03 -2.436116e-03
## Xa       5.040528e+00  5.048074e+00
## Xb       4.054103e+00  4.054884e+00
## Xc      -4.031743e-02 -4.004919e-02
## Xx_n     2.303489e-04  2.562824e-04
## Xx_m    -1.646753e-05 -3.616514e-06
## Xvel1    7.677796e-01  7.584315e-01
## Xvel2   -3.266386e-01 -3.210139e-01
## Xvel3   -5.647037e-02 -5.395596e-02
## XG1      4.556258e-03  4.487446e-03
## Xdelta2  3.001002e+00  3.003534e+00
## Xdelta3 -4.739770e-03 -4.994286e-03
```

```
confint(mod) # frequentist confidence interval
```

```
##                 2.5 %       97.5 %
## X       -0.0356008624 0.0307286313
## Xa       4.8119905173 5.2841583057
## Xb       3.8187732329 4.2909941988
## Xc      -0.1148669459 0.0347685688
## Xx_n    -0.0012390717 0.0017516365
```

```
## Xx_m    -0.0007525664 0.0007453334
## Xvel1   -0.1781601375 1.6950232173
## Xvel2   -0.8550656296 0.2130377656
## Xvel3   -0.4404899177 0.3325779994
## XG1     -0.0029865103 0.0119614013
## Xdelta2  2.9811035879 3.0259644311
## Xdelta3 -0.0162096406 0.0062210695
```

```r
t(apply(mat_beta, 2, quantile, probs=c(.025, 0.975))) #bayesian credible interval
```

```
##               2.5%        97.5%
##  [1,] -0.0358126292 0.0320460860
##  [2,]  4.7993671699 5.2843666833
##  [3,]  3.8173711900 4.2991317081
##  [4,] -0.1134091212 0.0330874874
##  [5,] -0.0012526292 0.0017688427
##  [6,] -0.0007676316 0.0007873408
##  [7,] -0.1530076602 1.6405456089
##  [8,] -0.8790163224 0.1873298922
##  [9,] -0.4300229621 0.3007667691
## [10,] -0.0029461987 0.0121564939
## [11,]  2.9824074703 3.0257335873
## [12,] -0.0161852931 0.0061561152
```