

# MTH 511a - 2020: Lecture 28

Instructor: Dootika Vats

*The instructor of this course owns the copyright of all the course materials. This lecture material was distributed only to the students attending the course MTH511a: “Statistical Simulation and Data Analysis” of IIT Kanpur, and should not be distributed in print or through electronic media without the consent of the instructor. Students can make their own copies of the course materials for their use.*

In this lecture, we will focus only the Bayesian linear regression model, and discuss accept-reject algorithms that sample from the posterior.

## 1 Bayesian linear regression

Consider a Bayesian version of the linear regression model, where prior distributions are assigned to both the regression coefficient  $\beta$  and the variance,  $\sigma^2$ . Recall that the likelihood is

$$y = (y_1, \dots, y_n) \mid \beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n).$$

The parameters of interest are  $\beta$  and  $\sigma^2$  and popular prior distributions assume independent priors:

$$\beta \sim N_p(0, \sigma^2 I_p) \quad \text{and} \quad \sigma^2 \sim \text{Inverse Gamma}(a, b).$$

Note that the Inverse Gamma distribution has density

$$\pi(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{-a+1} e^{-b/\sigma^2}.$$

The posterior distribution is

$$\begin{aligned} \pi(\beta, \sigma^2 \mid y) &\propto \pi(\beta, \sigma^2) \prod_{i=1}^n f(y_i \mid \beta, \sigma^2) \\ &= \pi(\beta) \pi(\sigma^2) \prod_{i=1}^n f(y_i \mid \beta, \sigma^2) \\ &= \left(\frac{1}{\sigma^2}\right)^{-a+1} e^{-b/\sigma^2} \left(\frac{1}{\sigma^2}\right)^{p/2} \exp\left\{-\frac{\beta^T \beta}{2\sigma^2}\right\} \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2}\right\} \end{aligned}$$

$$= (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{(y - X\beta)^T(y - X\beta)}{2\sigma^2} - \frac{\beta^T \beta}{2\sigma^2} - \frac{b}{\sigma^2} \right\}$$

The above is the  $(p + 1)$ -dimensional posterior distribution and we want to obtain samples from this using accept-reject. We already know that accept-reject does not work well in higher dimensions. So in any

So, to run an accept-reject sampler consider a proposal distribution  $q(\beta, \sigma^2) = \pi(\beta)\pi(\sigma^2)$ . That is the proposal distribution is the same as the prior distribution. Then, if MLE exists,

$$\begin{aligned} \frac{\tilde{\pi}(\beta, \sigma^2|y)}{\pi(\beta)\pi(\sigma^2)} &= \frac{\pi(\beta)\pi(\sigma^2) \prod_{i=1}^n f(y_i|\beta, \sigma^2)}{\pi(\beta)\pi(\sigma^2)} \\ &= \prod_{i=1}^n f(y_i|\beta, \sigma^2) \\ &\leq \prod_{i=1}^n f(y_i|\hat{\beta}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}^2) := M. \end{aligned}$$

So an accept-reject sampler is theoretically possible to implement. However, we note that the dimensionality of the problem will certainly impede efficiency. So that it will be very inefficient to implement AR here and possibly close to impossible to get even one draw from the posterior distribution in reasonable time.

## 1.1 Linchpin variable samplers

As we have discussed plenty of times now, it is difficult to implement AR when the target is high-dimensional or when the upper bound is hard to get. In the first case, a *linchpin variable* trick can be very useful. Suppose the target density is

$$\pi(x, y).$$

Then, we can split the joint distribution as the product of conditional times marginal. That is

$$\pi(x, y) = \pi(x|y) \pi(y).$$

If  $X|Y$  is known in closed-form and we can sample from it, then we may try and get samples from the marginal distribution of  $y$ . This is beneficial since the dimension of  $y$  is smaller than  $(x, y)$ , and implementing AR on a smaller dimensional problem will be much easier. So the algorithm would be

- Generate  $Y \sim \pi(y)$
- Generate  $X \sim X|Y$
- Output  $(X, Y)$ .

The variable  $Y$  is called the linchpin variable with target density  $\pi(y)$ . We can use this quite easily in Bayesian linear regression.

*Example 1* (Bayesian linear regression). Recall the posterior distribution in Bayesian linear regression as:

$$\pi(\beta, \sigma^2 | y) = (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2} - \frac{\beta^T \beta}{2\sigma^2} - \frac{b}{\sigma^2} \right\}$$

First, note that we prefer  $\sigma^2$  to be the linchpin variable since it is univariate, and  $\beta$  is  $p$ -variate. So we need to find the distribution  $\beta | \sigma^2$  and the marginal distribution of  $\sigma^2$ . Let  $A = (X^T X + I)$ .

$$\begin{aligned} & \int \pi(\beta, \sigma^2 | y) d\beta \\ & \propto \int (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T y - 2\beta^T X^T y + \beta^T X^T X \beta}{2\sigma^2} - \frac{\beta^T \beta}{2\sigma^2} - \frac{b}{\sigma^2} \right\} d\beta \\ & = (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T y}{2\sigma^2} - \frac{b}{\sigma^2} \right\} \int \exp \left\{ -\frac{\beta^T X^T X \beta - 2\beta^T X^T y}{2\sigma^2} - \frac{\beta^T \beta}{2\sigma^2} \right\} d\beta \\ & = (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T y}{2\sigma^2} - \frac{b}{\sigma^2} \right\} \int \exp \left\{ -\frac{\beta^T (X^T X + I) \beta - 2\beta^T X^T y}{2\sigma^2} \right\} d\beta \\ & = (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T y}{2\sigma^2} - \frac{b}{\sigma^2} \right\} \int \exp \left\{ -\frac{\beta^T A \beta - 2\beta^T A A^{-1} X^T y}{2\sigma^2} \right. \\ & \quad \left. - \frac{(A^{-1} X^T y)^T A (A^{-1} X^T y)}{2\sigma^2} + \frac{(A^{-1} X^T y)^T A (A^{-1} X^T y)}{2\sigma^2} \right\} d\beta \\ & = (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T y}{2\sigma^2} - \frac{b}{\sigma^2} + \frac{y^T X A^{-1} A A^{-1} X^T y}{2\sigma^2} \right\} \\ & = \times \int \exp \left\{ -\frac{\beta^T A \beta - 2\beta^T A A^{-1} X^T y + y^T X A^{-1} A A^{-1} X^T y}{2\sigma^2} \right\} \\ & = (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T y}{2\sigma^2} - \frac{b}{\sigma^2} + \frac{y^T X A^{-1} X^T y}{2\sigma^2} \right\} \int \exp \left\{ -\frac{(\beta - A^{-1} X^T y)^T A (\beta - A^{-1} X^T y)}{2\sigma^2} \right\} \end{aligned}$$

So  $\beta | \sigma^2$  is a multivariate normal distribution

$$\beta | \sigma^2, y \sim N_p (A^{-1} X^T y, \sigma^2 A^{-1}) ,$$

and the integral integrates to a known constant.

$$\int \pi(\beta, \sigma^2 | y) d\beta \propto (\sigma^2)^{-n/2-p/2-a+1} \exp \left\{ -\frac{y^T (I - X A^{-1} X^T) y}{2\sigma^2} - \frac{b}{\sigma^2} \right\} \cdot (\sigma^2)^{p/2} \det(A)^{p/2}$$

$$\propto (\sigma^2)^{-n/2-a+1} \exp \left\{ -\frac{y^T(I - XA^{-1}X^T)y}{2\sigma^2} - \frac{b}{\sigma^2} \right\}.$$

So the marginal posterior distribution for  $\sigma^2|y$  is

$$\sigma^2|y \sim \text{Inverse Gamma} \left( \frac{n}{2} + a, \frac{y^T(I - XA^{-1}X^T)y}{2} + b \right).$$

We have just done the following decomposition

$$\pi(\beta, \sigma^2|y) = \pi(\beta|\sigma^2, y)\pi(\sigma^2|y),$$

where both those densities are available in closed-form and samples can be generated easily from them in the following way:

1. Generate  $\sigma^2 \sim \text{Inverse Gamma}$  as indicated above
2. Generate  $\beta|\sigma^2 \sim \text{Normal distribution}$  as indicated above.
3.  $(\beta, \sigma^2)$  is one draw from the posterior. Repeat for many draws, and estimate posterior mean and quantiles.

We now implement Bayesian linear regression for the cars dataset

```
#####
## Linchpin variable sampler
## for Bayesian linear regression for cars
#####
set.seed(1)

# loading the dataset
data(cars)
n <- dim(cars)[1]
X <- cbind(1, cars$speed)
y <- cars$dist
p <- dim(X)[2]
a <- 1 # prior parameters
b <- 1 # prior parameters
```

Drawing the samples is easy, since no AR step is required

```
# We implement Monte Carlo sampling using linchping
N <- 1e4
A <- t(X)%*%X + diag(p)
A.inv <- solve(A)

sig2 <- numeric(length = N)
beta <- matrix(0, nrow = N, ncol = p)
```

```

rate.sig <- ( t(y) %*% (diag(1,n) - X %*% A.inv %*% t(X)) %*% y )/2 + b

# sampling Inverge Gamma for sigma2
sig2 <- 1 / rgamma(N, shape = n/2 + a, rate = rate.sig)

# Sampling beta from multivariate normal
# mean + sqrt(covariance) %*% rnorm
foo <- svd(A.inv) #Singular values decomposition of A^{-1}
Ainv.sqrt <- foo$u %*% diag(foo$d^(1/2)) %*% t(foo$v)

for(i in 1:N)
{
  beta[i,] <- A.inv %*% t(X) %*%y + Ainv.sqrt %*% rnorm(p, sd =
    sqrt(sig2[i])) # Getting beta estimates
}

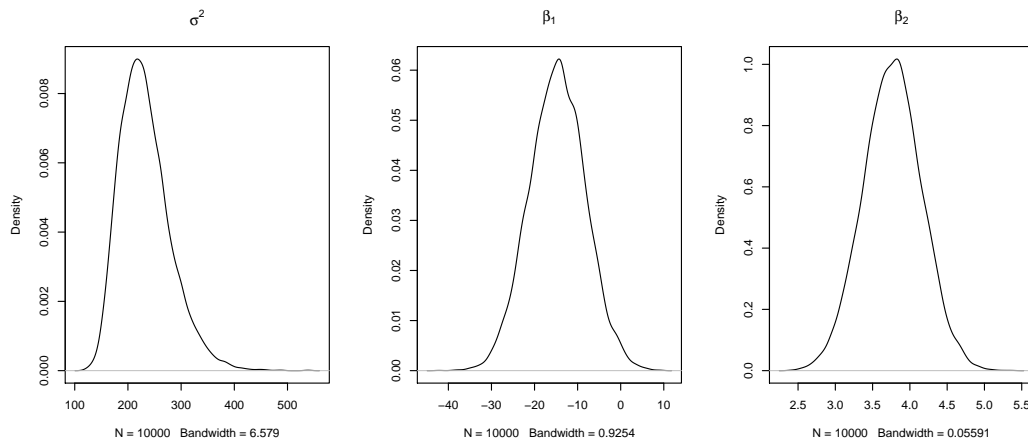
```

We can view the posterior marginal density plots of the samples

```

par(mfrow = c(1,3))
plot(density(sig2), main = expression(sigma^2))
plot(density(beta[,1] ), main = expression(beta[1]))
plot(density(beta[,2] ), main = expression(beta[2]))

```



We can also find the posterior means and quantiles:

```

poster <- cbind(sig2, beta)
colMeans(poster)
#      sig2
#232.404843 -14.713624 3.766138

apply(poster, 2, quantile, c(.025, .975))

```

```
#           sig2
#2.5% 157.9957 -27.370729 2.993072
#97.5% 342.2522 -1.772828 4.539354
```

---

Note that the posterior credible interval for both  $\beta_1$  and  $\beta_2$  do not have 0 in the interval, implying both regression coefficients are important and should be treated as non-zero.

## 2 Questions to think about

- Implement accept-reject for the cars dataset and see for yourself how well the algorithm works here.
- Suppose the marginal posterior distribution for  $\sigma^2|y$  was not from a nice known family. What could we have done then?
- What is a MAP estimator of  $\beta$  in this problem?