

STAT 545 HW 5

Andres Efren Rocha Jayasinha

Updated: September 20, 2026

Table of Contents

Assignment 5: Gibb's Sampler	2
Exercise 5.7	2
Bayesian Lasso	3

Assignment 5: Gibb's Sampler

Exercise 5.7

Random variables X and N have joint distribution, defined up to a constant of proportionality,

$$f(x, n) \propto \frac{e^{-3x} x^n}{n!}, \quad \text{for } n = 0, 1, 2, \dots \text{ and } x > 0.$$

Note that X is continuous and N is discrete.

- (a) Implement a Gibbs sampler to sample from this distribution.
- (b) Use your simulation to estimate (i) $P(X^2 < N)$ and (ii) $E(XN)$.
- (c) Identify the prior and conditional likelihood that produce this target as a joint distribution, and write the corresponding hierarchical model.

Solution

Part A/B:

Input:

Listing 1: Gibb's Sampler

```
1 ### Exercise 5.17
2
3 trials <- 100000
4 simlist <- matrix(rep(0, trials*2), ncol=2) ## (X, N)
5 simlist[1,] <- c(1, 1)
6
7 for (i in 2:trials) {
8   simlist[i,1] <- rgamma(1, 1 + simlist[i-1,2], 3)
9   simlist[i,2] <- rpois(1, simlist[i,1]) }
10
11 mean( (simlist[,1]^2 < simlist[,2]) ) # P(X^2 < N)
12 mean(simlist[,1]*simlist[,2]) # E(XN)
```

Output: $P(X < N) = 0.26244$ and $E[XN] = 0.4898135$

Part C:

The joint distribution is given (up to proportionality) by

$$f(x, n) \propto \frac{e^{-3x} x^n}{n!}, \quad x > 0, n = 0, 1, 2, \dots$$

To identify a prior and conditional likelihood, write the joint in the form

$$p(x, n) = p(n | x) p(x).$$

Observe that

$$\frac{e^{-3x} x^n}{n!} = \left(\frac{e^{-x} x^n}{n!} \right) e^{-2x}.$$

The first factor,

$$\frac{e^{-x} x^n}{n!},$$

is exactly the probability mass function of a Poisson distribution with parameter x . Hence it is natural to interpret

$$p(n | x) = \frac{e^{-x} x^n}{n!},$$

so that

$$N | X = x \sim \text{Poisson}(x).$$

This is the *conditional likelihood*, describing how the data variable N is generated given the parameter X .

The remaining factor,

$$e^{-2x}, \quad x > 0,$$

depends only on x and is proportional to a Gamma density with shape $\alpha = 1$ and rate $\beta = 2$. Therefore,

$$p(x) \propto e^{-2x}, \quad x > 0,$$

which corresponds to

$$X \sim \text{Gamma}(1, 2),$$

that is, an Exponential distribution with rate 2. This distribution plays the role of the *prior*, representing the distribution assigned to X before observing N .

Thus, the joint distribution can be viewed as arising from the hierarchical model

$X \sim \text{Exponential}(2),$ $N \mid X = x \sim \text{Poisson}(x).$
--

In other words, one first draws X from an Exponential(2) distribution, and then, conditional on this value of X , draws N from a Poisson distribution with mean x . This prior-likelihood combination produces the given joint distribution.

Bayesian Lasso

Consider the linear regression setting with $Y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$ and let

$$Y \mid \beta, \sigma^2 \sim \mathcal{N}(X\beta, \sigma^2 I_n),$$

$$\beta_j \stackrel{iid}{\sim} \text{Laplace}(0, \lambda),$$

$$\sigma^2 \sim \text{Inverse-Gamma}(a, b),$$

where $\lambda, a, b > 0$ are constants. The priors give:

$$f(\beta_j) = \frac{\lambda}{2} \exp\{-\lambda|\beta_j|\}, \quad f(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\}.$$

In this problem you will estimate the model parameters by sampling from the posterior using MCMC.

- (a) Write the log-posterior $\log f(\beta, \sigma^2 \mid y)$ up to an additive constant.
- (b) Write the (log) full conditionals of the j th coefficient β_j and the variance parameter σ^2 up to an additive constant.
- (c) Consider the Metropolis update for β_j using a Gaussian random walk proposal $\beta_j^* = \beta_j + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, s^2)$. Derive the log acceptance ratio $\log r(\beta_j, \beta_j^*)$ for the update.
- (d) Write pseudocode for the Metropolis-within-Gibbs algorithm to sample from the posterior of (β, σ^2) .
- (e) Code up your algorithm from the previous part. Using the `diabetes` dataset in the `lars` package (see provided script), compute estimates of the regression coefficients and variance parameter. Use $a = b = 2$ with $\lambda = 1$. Tune the random walk step size until your coordinatewise acceptance rates are between 40% and 60%. Report your final acceptance rates, coordinatewise effective sample sizes, and estimates of posterior means for the regression coefficients.
- (f) Which variables are associated with an estimated increase in mean disease progression? A decrease? No association?
- (g) Increase λ to 2 and leave other settings the same. How do coefficient estimates change, if at all?

Solution

Part A:

The posterior satisfies

$$f(\beta, \sigma^2 \mid y) \propto f(y \mid \beta, \sigma^2) f(\beta) f(\sigma^2).$$

From $Y \mid \beta, \sigma^2 \sim \mathcal{N}(X\beta, \sigma^2 I_n)$,

$$f(y \mid \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \|y - X\beta\|^2\right\},$$

so

$$\log f(y \mid \beta, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|y - X\beta\|^2 + C.$$

Independence of the Laplace priors gives

$$f(\beta) = \prod_{j=1}^p \frac{\lambda}{2} \exp\{-\lambda|\beta_j|\} \Rightarrow \log f(\beta) = -\lambda \sum_{j=1}^p |\beta_j| + C.$$

For $\sigma^2 \sim \text{Inverse-Gamma}(a, b)$ with density

$$f(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\},$$

$$\log f(\sigma^2) = -(a+1) \log(\sigma^2) - \frac{b}{\sigma^2} + C.$$

Adding terms yields the log-posterior (up to an additive constant):

$$\log f(\beta, \sigma^2 \mid y) = -\left(\frac{n}{2} + a + 1\right) \log(\sigma^2) - \frac{1}{2\sigma^2} \|y - X\beta\|^2 - \frac{b}{\sigma^2} - \lambda \sum_{j=1}^p |\beta_j| + C.$$

Part B:

The log-posterior from part (a), up to an additive constant, is

$$\log f(\beta, \sigma^2 \mid y) = -\frac{1}{2\sigma^2} \|y - X\beta\|^2 - \lambda \sum_{k=1}^p |\beta_k| - \left(\frac{n}{2} + a + 1\right) \log(\sigma^2) - \frac{b}{\sigma^2} + C.$$

Full conditional of β_j .

Let x_j denote the j th column of X , and define the partial residual

$$r_j = y - \sum_{k \neq j} x_k \beta_k = y - X_{-j} \beta_{-j}.$$

Then

$$y - X\beta = r_j - x_j \beta_j.$$

Expanding the quadratic term,

$$\|r_j - x_j \beta_j\|^2 = r_j^\top r_j - 2\beta_j x_j^\top r_j + \beta_j^2 x_j^\top x_j.$$

Dropping terms not involving β_j , the log full conditional is

$$\log f(\beta_j \mid \beta_{-j}, \sigma^2, y) = -\frac{1}{2\sigma^2} (\beta_j^2 x_j^\top x_j - 2\beta_j x_j^\top r_j) - \lambda |\beta_j| + C.$$

Equivalently, letting

$$A_j = x_j^\top x_j, \quad B_j = x_j^\top r_j,$$

$$\log f(\beta_j \mid \dots) = -\frac{A_j}{2\sigma^2} \beta_j^2 + \frac{B_j}{\sigma^2} \beta_j - \lambda |\beta_j| + C.$$

Full conditional of σ^2 .

Keeping only the σ^2 terms,

$$\log f(\sigma^2 \mid \beta, y) = -\left(\frac{n}{2} + a + 1\right) \log(\sigma^2) - \frac{1}{\sigma^2} \left(b + \frac{1}{2} \|y - X\beta\|^2\right) + C.$$

This is the kernel of an Inverse-Gamma distribution, hence

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

Part C:

We update β_j using a Gaussian random-walk proposal

$$\beta_j^* = \beta_j + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, s^2).$$

Since this proposal is symmetric,

$$q(\beta_j^* | \beta_j) = q(\beta_j | \beta_j^*),$$

the proposal densities cancel in the Metropolis–Hastings ratio.

Hence the acceptance ratio is

$$r(\beta_j, \beta_j^*) = \frac{f(\beta_j^* | \beta_{-j}, \sigma^2, y)}{f(\beta_j | \beta_{-j}, \sigma^2, y)}.$$

Taking logarithms,

$$\log r(\beta_j, \beta_j^*) = \log f(\beta_j^* | \beta_{-j}, \sigma^2, y) - \log f(\beta_j | \beta_{-j}, \sigma^2, y).$$

From part (b), the log full conditional (up to a constant) is

$$\log f(\beta_j | \beta_{-j}, \sigma^2, y) = -\frac{1}{2\sigma^2} \|r_j - x_j \beta_j\|^2 - \lambda |\beta_j| + C,$$

where

$$r_j = y - X_{-j} \beta_{-j}.$$

Therefore,

$$\log r(\beta_j, \beta_j^*) = -\frac{1}{2\sigma^2} (\|r_j - x_j \beta_j^*\|^2 - \|r_j - x_j \beta_j\|^2) - \lambda (|\beta_j^*| - |\beta_j|).$$

The Metropolis update accepts β_j^* with probability

$$\alpha = \min\{1, \exp(\log r(\beta_j, \beta_j^*))\}.$$

Part d:

Let T be the number of MCMC iterations. Initialize $\beta^{(0)} \in \mathbb{R}^p$ and $(\sigma^2)^{(0)} > 0$. For each coordinate j , choose a random-walk step size $s_j > 0$.

Metropolis-within-Gibbs sampler:

1. For $t = 1, 2, \dots, T$:

(a) For $j = 1, 2, \dots, p$ (coordinate update for β_j):

i. Define the partial residual

$$r_j^{(t)} = y - X_{-j} \beta_{-j}^{(t)},$$

where $\beta_{-j}^{(t)}$ denotes the current values of all coefficients except β_j (using updated values for coordinates $1, \dots, j-1$ and previous-iteration values for $j+1, \dots, p$).

ii. Propose

$$\beta_j^* = \beta_j^{(t-1)} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, s_j^2).$$

iii. Compute the log acceptance ratio

$$\log r = -\frac{1}{2(\sigma^2)^{(t-1)}} (\|r_j^{(t)} - x_j \beta_j^*\|^2 - \|r_j^{(t)} - x_j \beta_j^{(t-1)}\|^2) - \lambda (|\beta_j^*| - |\beta_j^{(t-1)}|).$$

iv. Set $\alpha = \min\{1, \exp(\log r)\}$. Draw $u \sim \text{Uniform}(0, 1)$. If $u < \alpha$, accept and set $\beta_j^{(t)} = \beta_j^*$; otherwise reject and set $\beta_j^{(t)} = \beta_j^{(t-1)}$.

(b) Update σ^2 from its full conditional:

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

2. Store $\beta^{(t)}$ and $(\sigma^2)^{(t)}$.

Part E:

Listing 2: Metropolis within Gibb's Sampler

```
1 install.packages(c("lars", "mcmcse")) # if needed
2
3 # ----- load data -----
4 data(diabetes, package = "lars")
5
```

```

6 # diabetes disease progression after one year
7 y <- scale(diabetes$y, center = TRUE, scale = FALSE) |> as.numeric()
8
9 # possible explanatory attributes
10 x <- scale(diabetes$x, center = TRUE, scale = TRUE)
11 n <- nrow(x)
12 p <- ncol(x)
13 colnames(x) <- colnames(diabetes$x)
14
15 #####
16 # Bayesian lasso: MH-within-Gibbs skeleton
17 # y | beta, sig2 ~ N(x beta, sig2 I)
18 # beta_j iid ~ Laplace(0, lambda) with f(b) = (lambda/2) exp{-lambda |b|}
19 # sig2 follows Inv-Gamma(a, b) with density proportional to (sig2)^{-(a+1)} exp{-b/sig2}
20 #
21 # Gibbs: sig2 | beta, y is conjugate
22 # MH: beta_j | beta_{-j}, sig2, y
23 #####
24
25 set.seed(1)
26
27 # ----- hyperparameters -----
28 a <- 2
29 b <- 2
30 lambda <- 1
31
32 # ----- MCMC settings -----
33 n_iter <- 20000
34 burn <- 2000
35 thin <- 1
36 s_beta <- 2.6 # RW step size for beta_j proposals (tune)
37
38 # ----- initialize -----
39 beta_curr <- rep(0, p)
40 sig2_curr <- 1
41
42 # ----- full conditional for sig2 -----
43 rcond_sig2 <- function(beta, a, b) {
44   mu_hat <- as.numeric(x %*% beta)
45   rss <- sum((y - mu_hat)^2)
46   shape <- a + n / 2
47   rate <- b + 0.5 * rss
48   1 / rgamma(1, shape = shape, rate = rate)
49 }
50
51 # ----- FILL IN: proposal + log acceptance ratio for ONE beta_j update -----
52 rprop_beta_j <- function(beta_j_curr, s_beta) {
53   # Gaussian random-walk proposal
54   beta_j_curr + rnorm(1, mean = 0, sd = s_beta)
55 }
56
57 log_r_beta_j <- function(beta_j_new, beta_j_curr, j,
58                           beta_curr, sig2_curr, lambda) {
59   # symmetric RW proposal => log r = log pi(new) - log pi(curr) for beta_j | rest
60   xj <- x[, j]
61   mu_curr <- as.numeric(x %*% beta_curr)
62   res_curr <- y - mu_curr
63   rss_curr <- sum(res_curr^2)
64
65   delta <- beta_j_new - beta_j_curr
66   res_new <- res_curr - xj * delta
67   rss_new <- sum(res_new^2)
68
69   # likelihood difference + Laplace prior difference (constants cancel)
70   log_r <- -(rss_new - rss_curr) / (2 * sig2_curr) -

```

```

71     lambda * (abs(beta_j_new) - abs(beta_j_curr))
72
73     log_r
74 }
75
76 # ----- storage -----
77 keep_idx <- which(seq_len(n_iter) > burn & ((seq_len(n_iter) - burn) %% thin == 0))
78 n_keep <- length(keep_idx)
79 beta_draws <- matrix(NA_real_, n_keep, p)
80 sig2_draws <- numeric(n_keep)
81
82 # ----- MH-within-Gibbs loop -----
83 k <- 0
84 acc_beta <- integer(p) # coordinatewise accept counts
85 for (t in 1:n_iter) {
86
87     # --- update beta one coordinate at a time via MH ---
88     for (j in 1:p) {
89         beta_j_curr <- beta_curr[j]
90         beta_j_new <- rprop_beta_j(beta_j_curr, s_beta)
91
92         log_r <- log_r_beta_j(beta_j_new, beta_j_curr, j,
93                               beta_curr, sig2_curr, lambda)
94
95         if (log(runif(1)) < min(0, log_r)) {
96             beta_curr[j] <- beta_j_new
97             acc_beta[j] <- acc_beta[j] + 1L
98         }
99     }
100
101     # --- update sig2 via Gibbs ---
102     sig2_curr <- rcond_sig2(beta_curr, a, b)
103
104     # --- save ---
105     if (t %in% keep_idx) {
106         k <- k + 1
107         beta_draws[k, ] <- beta_curr
108         sig2_draws[k] <- sig2_curr
109     }
110 }
111
112 # ----- FILL IN: coordinatewise acceptance rates -----
113 acc_rate_beta <- acc_beta / n_iter
114 names(acc_rate_beta) <- colnames(x)
115 acc_rate_beta
116
117 # ----- FILL IN: coordinatewise ess -----
118 library(mcmcse)
119 ess_beta <- apply(beta_draws, 2, mcmcse::ess)
120 names(ess_beta) <- colnames(x)
121 ess_sig2 <- mcmcse::ess(sig2_draws)
122 ess_beta
123 ess_sig2
124
125 # ----- FILL IN: estimates -----
126 beta_post_mean <- colMeans(beta_draws)
127 sig2_post_mean <- mean(sig2_draws)
128
129 beta_post_mean
130 sig2_post_mean

```

Output:

	Age	Sex	BMI	MAP	TC	LDL	HDL	TCH	LTG	GLU
Acceptance Rate	0.4180	0.5392	0.7096	0.7037	0.4382	0.4363	0.6893	0.5288	0.7065	0.5394
ESS	2968.18	1986.64	1275.29	1580.95	2045.35	2579.21	1608.57	1983.19	1138.78	2135.00
Posterior Mean	0.2499	-1.6547	23.2438	8.6590	-0.5053	-0.4415	-4.9045	1.5876	19.6723	1.7049

Table 1: MCMC diagnostics and posterior means for all regression coefficients

Part F:

Based on the table in Part E, BMI, LTG and MAP are associated with increase in mean disease progression (positive coeff.). HDL is associated with disease in mean disease progression (negative coeff.). The rest of the variables have little to no association (close to 0 coeff.)

Part G:

	Age	Sex	BMI	MAP	TC	LDL	HDL	TCH	LTG	GLU
Acceptance Rate	0.2825	0.2704	0.7232	0.6774	0.2662	0.2693	0.5291	0.4305	0.7203	0.4052
ESS	2828.27	2728.03	1426.01	1803.99	3135.07	2869.95	2119.10	1768.83	1104.64	2190.37
Posterior Mean	0.3006	-0.1834	19.2743	4.0529	0.0827	0.0693	-2.0003	1.2974	15.4730	1.1556

Table 2: MCMC diagnostics and posterior means for all regression coefficients ($\lambda = 2$)

We see that the values for the variables coefficient's change, however relative to one another they exhibit the same result as we saw in Part F.