Homework 4
of
STAT 632

Yanxun Xu

December 1, 2015

1 Problem 1

Solution:

1.1 (a)

As $Y \sim Normal(\theta_1 + \theta_2, 1)$, the data likelihood of y can be written as

$$\begin{aligned} p(y|\theta_1, \theta_2) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \theta_1 - \theta_2)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}[y^2 - 2y(\theta_1 + \theta_2) + (\theta_1 + \theta_2)^2]\right) \\ &\propto \exp\left(-\frac{1}{2}[y^2 - 2y(\theta_1 + \theta_2) + (\theta_1 + \theta_2)^2]\right). \end{aligned}$$

We assume that the prior distribution of θ_i follows a normal distribution with parameters a_i and b_i, given by

$$\begin{aligned} p(\theta_i) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2b_i^2}(\theta_i - a_i)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2b_i^2}(\theta_i^2 - 2a_i\theta_i + a_i^2)\right) \\ &\propto \exp\left(-\frac{1}{2b_i^2}(\theta_i^2 - 2a_i\theta_i + a_i^2)\right), i = 1, 2. \end{aligned}$$

According to Bayes' rule, the full conditional distribution $p(\theta_1|\theta_2, y)$ can be obtained by

$$\begin{aligned} p(\theta_1|\theta_2, y) &= \frac{p(\theta_1, \theta_2, y)}{p(\theta_2, y)} \\ &= \frac{p(y|\theta_1, \theta_2)p(\theta_1, \theta_2)}{p(y|\theta_2)p(\theta_2)} \\ &= \frac{p(y|\theta_1, \theta_2)p(\theta_1)p(\theta_2)}{p(y|\theta_2)p(\theta_2)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{p(y|\theta_1,\theta_2)p(\theta_1)}{p(y|\theta_2)} \\
&\propto p(y|\theta_1,\theta_2)p(\theta_1) \\
&\propto \exp\left(-\frac{1}{2}[y^2 - 2y(\theta_1+\theta_2) + (\theta_1+\theta_2)^2]\right)\exp\left(-\frac{1}{2b_1^2}(\theta_1^2 - 2a_1\theta_1 + a_1^2)\right) \\
&= \exp\left(-\frac{1}{2}[(1+\frac{1}{b_1^2})\theta_1^2 - 2(y-\theta_2+\frac{a_1}{b_1^2})\theta_1 + (y^2 - 2y\theta_2 + \theta_2^2 + \frac{a_1^2}{b_1^2})]\right) \\
&\sim Normal\left(\theta_1; \frac{y-\theta_2+\frac{a_1}{b_1^2}}{1+\frac{1}{b_1^2}}, \frac{1}{1+\frac{1}{b_1^2}}\right).
\end{aligned}
$$

Similarly, we can deduce the other full conditional distribution $p(\theta_2|\theta_1, y)$ as

$$
\begin{aligned}
p(\theta_2|\theta_1,y) &= \frac{p(\theta_1,\theta_2,y)}{p(\theta_1,y)} \\
&\propto p(y|\theta_1,\theta_2)p(\theta_2) \\
&\sim Normal\left(\theta_2; \frac{y-\theta_1+\frac{a_2}{b_2^2}}{1+\frac{1}{b_2^2}}, \frac{1}{1+\frac{1}{b_2^2}}\right).
\end{aligned}
$$

1.2 (b)

$$
\begin{aligned}
\theta_1|y &\sim N(\mu_1, \Sigma_{11}) = N\left(\frac{a_1 + b_1^2 y + a_1 b_2^2 - a_2 b_1^2}{1+b_1^2+b_2^2}, \left[\frac{b_1^2(1+b_2^2)}{1+b_1^2+b_2^2}\right]\right) \\
\theta_2|y &\sim N(\mu_2, \Sigma_{22}) = N\left(\frac{a_2 + b_2^2 y + a_2 b_1^2 - a_1 b_2^2}{1+b_1^2+b_2^2}, \left[\frac{b_2^2(1+b_1^2)}{1+b_1^2+b_2^2}\right]\right) \qquad (1)
\end{aligned}
$$

The data do update the prior distributions for these parameters.

1.3 (c)

Using the code attached, we run the Gibbs sampler described in Problem 1(a) for $t = 100$ and $t = 1000$ iterations, respectively. Starting the chains near the prior mean, the trace plots of θ_1, θ_2, and $\mu = \theta_1 + \theta_2$ are shown in Figure 1. As we can see from the figure, both θ_1 and θ_2 seem not to converge, but their summation μ always approximates to 0. Discarding the burn-in period $[0, \frac{t}{2}]$, the estimate posterior mean of $\mu = 0.1462$ for the former case and $\mu = 0.0614$ for the latter case.

1.4 (d)

Using the same code, we repeat the above process with different priors. The trace plots of θ_1, θ_2, and $\mu = \theta_1 + \theta_2$ are shown in Figure 2. As we can see from the figure, both θ_1 and θ_2 seem not to converge either, but their summation μ still fluctuates around 0. Discarding the burn-in period $[0, \frac{t}{2}]$, the estimate posterior mean of $\mu = 0.4740$ for the former case and $\mu = 0.5264$ for the latter case, which are greater than what we estimate in Problem 1(c).

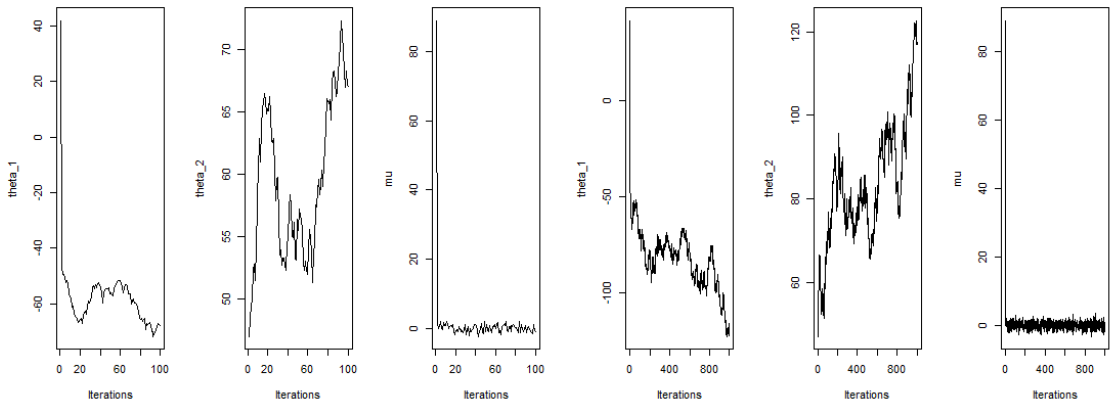

Figure 1: The trace plot of θ_1, θ_2, and μ when the priors are $\theta_i \sim Normal(50, 1000^2), i = 1, 2$ for $t = 100$ iterations (Left) and $t = 1000$ iterations (Right)

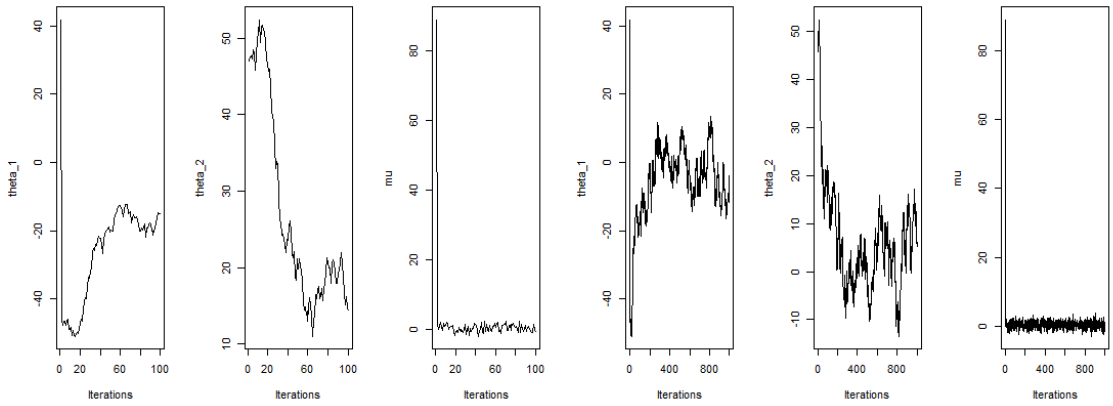

Figure 2: The trace plot of θ_1, θ_2, and μ when the priors are $\theta_i \sim Normal(50, 10^2), i = 1, 2$ for $t = 100$ iterations (Left) and $t = 1000$ iterations (Right)

2 Problem 9.2

Solution:

2.1 (a)

Using the code attached, we fit a Bayesian regression model using the g-prior with $g = 532$, $\nu_0 = 2$, and $\sigma_0^2 = 1$. The posterior distribution for each regressor is shown in each panel of Figure 3. The blue vertical lines represent the posterior means and confidence intervals between two green vertical line is the 95% confidence intervals. Also, we give the numerical results of the posterior means and confidence intervals for all of the parameters as show in Table 1.

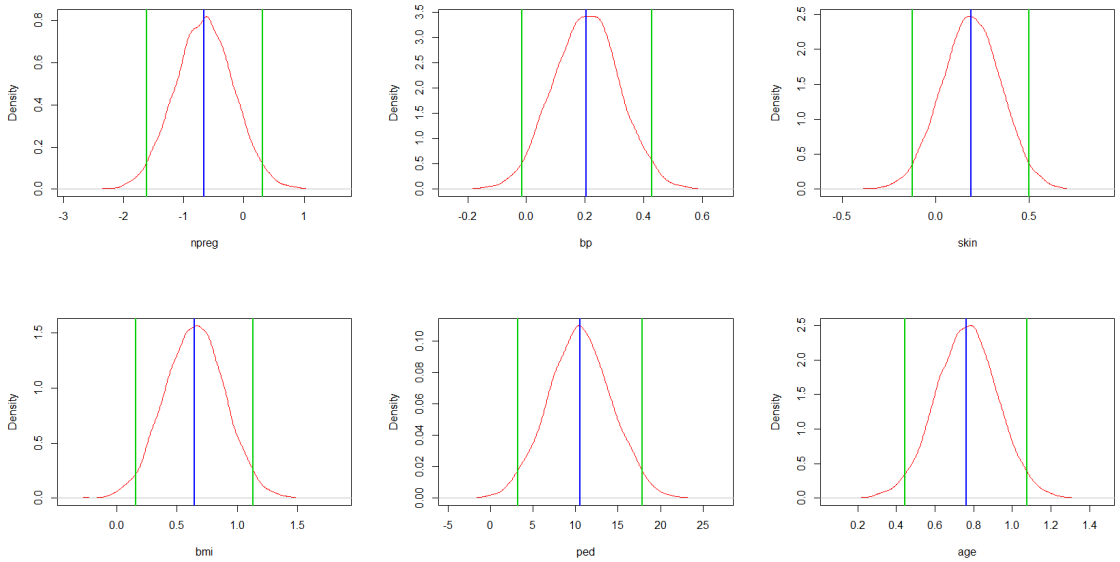

Figure 3: The posterior distribution for each parameter with the corresponding posterior mean (Blue lines) and 95% confidence interval (Between green lines)

Table 1: The posterior mean and 95% confidence interval for each parameter

Regressor	Mean	95% confidence interval
(Intercept)	52.2316	$[35.1472, 69.2597]$
npreg	-0.6548	$[-1.6131, 0.3205]$
bp	0.2045	$[-0.0183, 0.4290]$
skin	0.1949	$[-0.1200, 0.5079]$
bmi	0.6407	$[0.1664, 1.1315]$
ped	10.5213	$[3.1344, 17.7420]$
age	0.7644	$[0.4518, 1.0787]$

2.2 (b)

Using the code attached, we perform the model selection and averaging procedure using the g-prior with $g = 532$, $\nu_0 = 2$, and $\sigma_0^2 = 1$. The posterior distribution for each regressor is shown in each panel of Figure 5.

The blue vertical lines represent the posterior means and confidence intervals between two green vertical line is the 95% confidence intervals. Also, we give the numerical results of $P(\beta_j \neq 0|\boldsymbol{y})$ and confidence intervals for all of the parameters as show in Table 2.

Comparing to the results in Problem 9.2(a), for the regressor npreg, bp, and skin, of which 0 is included in the 95% confidence interval by Bayesian inference, the non-zero probability $P(\beta_j \neq 0|\boldsymbol{y})$ is far away from 1 and their posterior distribution narrow down to 0; for the other regressor whose $P(\beta_j \neq 0|\boldsymbol{y}) \approx 1$, their posterior means and confidence intervals are almost the same by both methods.

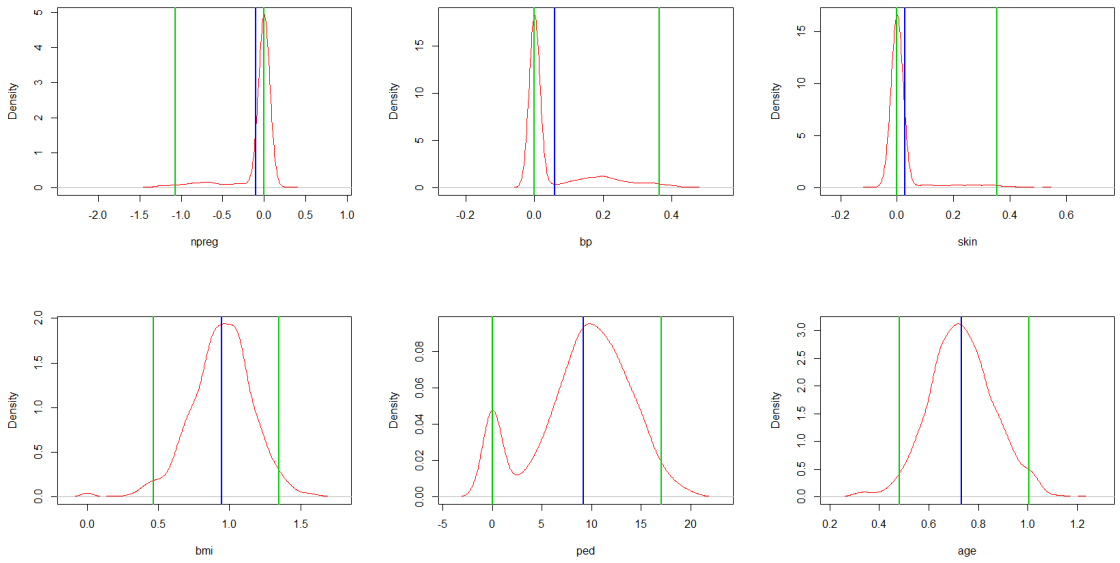

Figure 4: The posterior distribution for each parameter with the corresponding posterior mean (Blue lines) and 95% confidence intervals (Between green lines)

Table 2: $P(\beta_j \neq 0|\boldsymbol{y})$ and the posterior 95% confidence interval for each parameter

Regressor	$P(\beta_j \neq 0\|\boldsymbol{y})$	95% confidence interval
(Intercept)	1.000	$[42.2292, 76.6873]$
npreg	0.167	$[-1.0712, 0.0000]$
bp	0.293	$[0.0000, 0.3648]$
skin	0.129	$[0.0000, 0.3531]$
bmi	0.995	$[0.4650, 1.3501]$
ped	0.892	$[0.0000, 17.0726]$
age	1.000	$[0.4826, 1.0066]$

3 Problem 9.3

Solution:

3.1 (a)

Using the code attached, we fit a Bayesian regression model using the g-prior with $g = 47$, $\nu_0 = 2$, and $\sigma_0^2 = 1$. The posterior distribution for each regressor is shown in each panel of Figure 5. The blue vertical lines represent the posterior means and the intervals between two green vertical line is the 95% confidence intervals. Also, we give the numerical results of the posterior means and confidence intervals for all of the parameters as shown in Table 3. The least squares estimates are also given in Table 3, from which we can see the 95% confidence intervals have usually a wider range than those inferred by Bayesian approach.

We say a variable is strongly predictive of crime rates when either 0 is not contained in its 95% confidence interval by Bayesian approach or p-value is remarkably small by ordinary least squares (OLS) method (the one with asterisk in Table 3). As we can see from Table 3, those strong predictors are M (percentage of males aged $14 - 24$), Ed (mean years of schooling), U2 (unemployment rate of urban males 3539), Ineq (income inequality), and Prob (probability of imprisonment). Among those regressor, Ed and Ineq are the two most significant positive factors, which means the more years of schooling and the more income inequality, the higher crime rates. Prob are the only one remarkable negative factor, which means the higher probability of imprisonment, the lower crime rates.

Table 3: The posterior mean and 95% confidence interval inferred by Bayesian approach and the estimate coefficient and p-value obtained by ordinary least squares method for each parameter

	Bayes		*OLS*		
Regressor	$\hat{\beta}_{Bayes}$	95% confidence interval	$\hat{\beta}_{ols}$	95% confidence interval	p-value
(Intercept)	-0.0003	$[-0.1403, 0.1395]$	-0.0005	$[-0.1614, 0.1605]$	0.9954
M	0.2816	$[0.0358, 0.5179]$	0.2865	$[0.0096, 0.5635]$	0.0430^*
So	-0.0012	$[-0.3355, 0.3341]$	-0.0001	$[-0.3755, 0.3753]$	0.9995
Ed	0.5350	$[0.2035, 0.8640]$	0.5445	$[0.1782, 0.9108]$	0.0049^{**}
Po1	1.4535	$[-0.0148, 2.9322]$	1.4716	$[-0.1939, 3.1372]$	0.0813
Po2	-0.7784	$[-2.3199, 0.7514]$	-0.7818	$[-2.5186, 0.9551]$	0.3657
LF	-0.0661	$[-0.3396, 0.2123]$	-0.0660	$[-0.3789, 0.2470]$	0.6703
M.F	-0.1278	$[-0.1556, 0.4104]$	0.1313	$[-0.1852, 0.4478]$	0.4040
Pop	-0.0700	$[-0.3017, 0.1622]$	-0.0703	$[-0.3291, 0.1886]$	0.5837
NW	0.1070	$[-0.1990, 0.4181]$	0.1091	$[-0.2416, 0.4597]$	0.5305
U1	-0.2662	$[-0.6226, 0.0858]$	-0.2705	$[-0.6716, 0.1305]$	0.1787
U2	0.3624	$[0.0365, 0.6860]$	0.3687	$[0.0019, 0.7356]$	0.0489^*
GDP	0.2330	$[-0.2356, 0.6993]$	0.2381	$[-0.2901, 0.7662]$	0.3650
Ineq	0.7123	$[0.2823, 1.1328]$	0.7263	$[0.2487, 1.2038]$	0.0041^{**}
Prob	-0.2806	$[-0.5262, -0.0367]$	-0.2852	$[-0.5581, -0.0124]$	0.0411^*
Time	-0.0611	$[-0.2961, 0.1751]$	-0.0616	$[-0.3288, 0.2056]$	0.6417

3.2 (b)

3.2.1 (i)

Using the code attached, we randomly divide the crime data set into a training data set and a test data set and plot $\hat{\boldsymbol{y}}_{ols}$ versus $\boldsymbol{y}_{te}$, as shown in Figure 6 (Left). The corresponding prediction error is computed and its value is 0.5443.

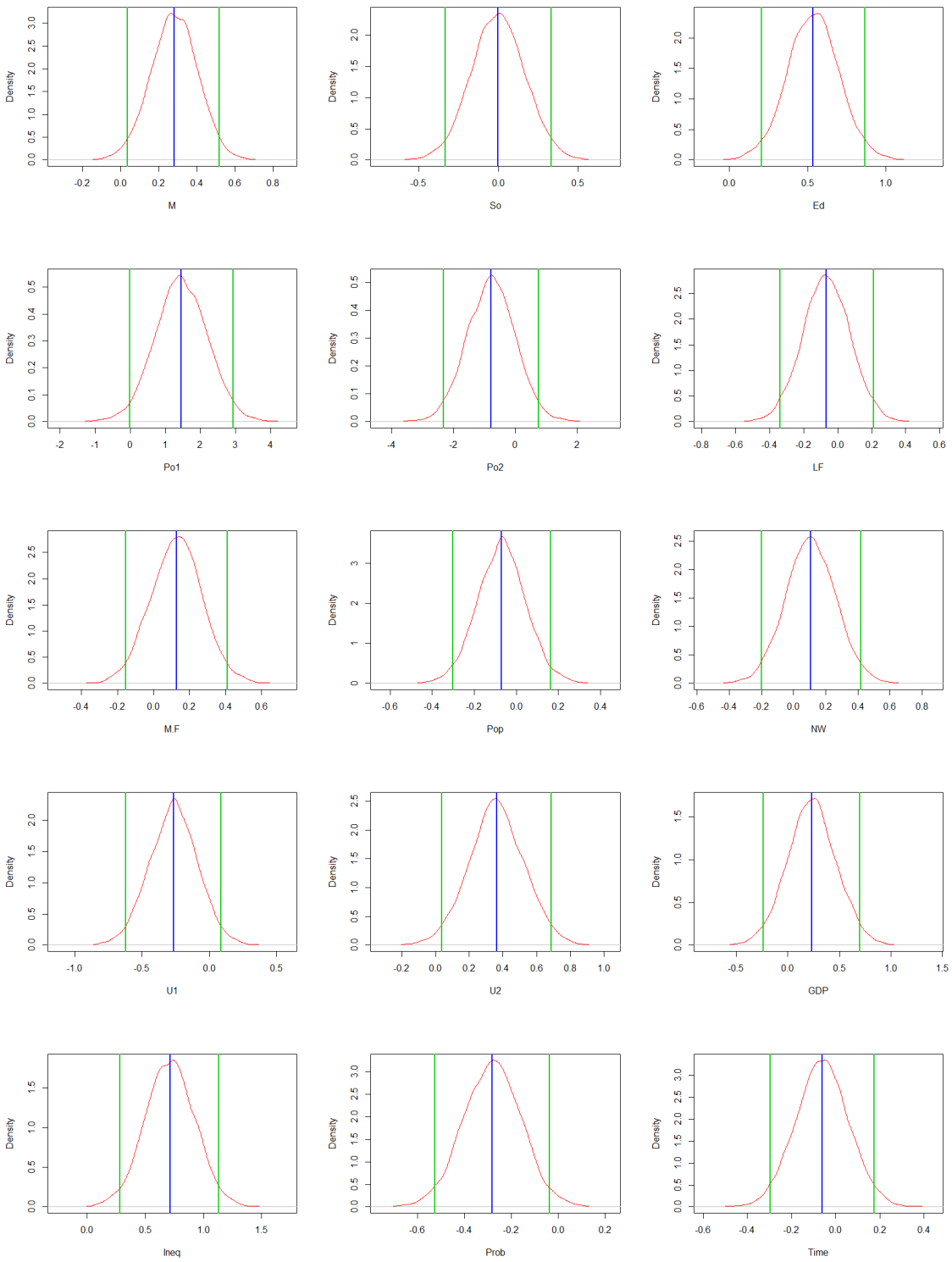

Figure 5: The posterior distribution for each parameter with the corresponding posterior mean (Blue lines) and 95% confidence interval (Between green lines)

3.2.2 (ii)

Using the same training set to estimate the regression coefficients and the same test set to evaluate the performance, we plot $\hat{\boldsymbol{y}}_{Bayes}$ versus $\boldsymbol{y}_{te}$, as shown in Figure 6 (Left). The corresponding prediction error is computed and its value is 0.5219, which is smaller than the OLS prediction error.

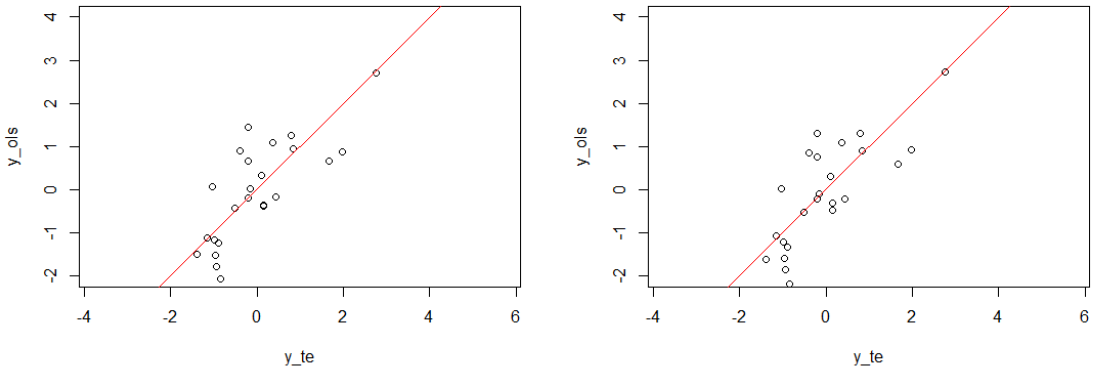

Figure 6: $\boldsymbol{y}_{te}$ versus prediction based on the ordinary least squares (Left) and the Bayesian approach (Right)

3.3 (c)

Using the code attached, we repeat the procedures in Problem 9.3(b) for 1000 times, the average prediction error for the OLS is 1.2074 and for the Bayesian methods is 1.1375. Again, it validate the truth that the Bayesian methods outperform the OLS.

4 Problem 4

Solution:

The marginal posterior density of σ^2 is proportional to $p(\sigma^2)p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{z},\sigma^2)$.

As σ^2 follows an inverse-gamma distribution with parameter $\frac{\nu_0}{2}$ and $\frac{\nu_0\sigma_0^2}{2}$, we have

$$\begin{aligned} p(\frac{1}{\sigma^2}) &= p(\gamma) \\ &= \left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}} \Gamma\Big(\frac{\nu_0}{2}\Big)^{-1} \gamma^{\frac{\nu_0}{2}-1} \exp\left(-\gamma\frac{\nu_0\sigma_0^2}{2}\right). \end{aligned}$$

For any given $\boldsymbol{z}$ with p_z non-zero entries, let $\boldsymbol{X}_z$ be the $n \times p_z$ matrix corresponding to the variables j for which $z_j = 1$ and $\boldsymbol{\beta}_z$ be the $p_z \times 1$ vector consisting of the entries of $\boldsymbol{\beta}$ for which $z_j = 1$. Using the rules of marginal probability, the latter term in this product can be expressed as

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{z},\sigma^2) &= \int p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{z},\boldsymbol{\beta},\sigma^2)p(\boldsymbol{\beta}|\boldsymbol{X},\sigma^2)d\boldsymbol{\beta} \\ &= \int (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}_z\boldsymbol{\beta}_z)^T(\boldsymbol{y}-\boldsymbol{X}_z\boldsymbol{\beta}_z)\right)|2\pi g\sigma^2(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}|^{-1} \exp\left(-\frac{1}{2g\sigma^2}\boldsymbol{\beta}_z^T\boldsymbol{X}_z^T\boldsymbol{X}_z\boldsymbol{\beta}_z\right)d\boldsymbol{\beta}_z \\ &= \int (2\pi\sigma^2)^{-\frac{n}{2}}|2\pi g\sigma^2(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}|^{-1} \exp\left(-\frac{1}{2\sigma^2}[(\boldsymbol{y}-\boldsymbol{X}_z\boldsymbol{\beta}_z)^T(\boldsymbol{y}-\boldsymbol{X}_z\boldsymbol{\beta}_z) + \frac{\boldsymbol{\beta}_z^T\boldsymbol{X}_z^T\boldsymbol{X}_z\boldsymbol{\beta}_z}{g}]\right)d\boldsymbol{\beta}_z \\ &= \int (2\pi\sigma^2)^{-\frac{n}{2}}|2\pi g\sigma^2(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}|^{-1} \exp\left(-\frac{1}{2\sigma^2}[\boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{y}^T\boldsymbol{X}_z\boldsymbol{\beta}_z + (1+\frac{1}{g})\boldsymbol{\beta}_z^T\boldsymbol{X}_z^T\boldsymbol{X}_z\boldsymbol{\beta}_z]\right)d\boldsymbol{\beta}_z \\ &= \int (2\pi\sigma^2)^{-\frac{n}{2}}|2\pi g\sigma^2(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}|^{-1} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T\boldsymbol{y} - \frac{1}{2}(\boldsymbol{\beta}_z-\boldsymbol{m})^T\boldsymbol{V}^{-1}(\boldsymbol{\beta}_z-\boldsymbol{m}) + \frac{1}{2}\boldsymbol{m}^T\boldsymbol{V}^{-1}\boldsymbol{m}\right)d\boldsymbol{\beta}_z \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}(1+g)^{-\frac{p_z}{2}} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T\boldsymbol{y} + \frac{1}{2}\boldsymbol{m}^T\boldsymbol{V}^{-1}\boldsymbol{m}\right)\int |2\pi\boldsymbol{V}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_z-\boldsymbol{m})^T\boldsymbol{V}^{-1}(\boldsymbol{\beta}_z-\boldsymbol{m})\right)d\boldsymbol{\beta} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}(1+g)^{-\frac{p_z}{2}} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}^T\boldsymbol{y} + \frac{1}{2}\boldsymbol{m}^T\boldsymbol{V}^{-1}\boldsymbol{m}\right) \\ &= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}(1+g)^{-\frac{p_z}{2}} \exp\left(-\frac{1}{2\sigma^2}SSR_g^z\right) \\ p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{z},\gamma) &= (2\pi)^{-\frac{n}{2}}(\gamma)^{\frac{n}{2}}(1+g)^{-\frac{p_z}{2}} \exp\Big(-\frac{\gamma}{2}SSR_g^z\Big), \end{aligned}$$

where $\boldsymbol{V} = \frac{g}{g+1}\sigma^2(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}$, $m = \frac{g}{g+1}(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}\boldsymbol{X}_z^T\boldsymbol{y}$, and $SSR_g^z = \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{m}^T\boldsymbol{V}^{-1}\boldsymbol{m} = \boldsymbol{y}^T[\boldsymbol{I} - \frac{g}{g+1}\boldsymbol{X}_z(\boldsymbol{X}_z^T\boldsymbol{X}_z)^{-1}\boldsymbol{X}_z]\boldsymbol{y}$.

Thus, we obtain that

$$p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{z},\gamma)p(\gamma) = (2\pi)^{-\frac{n}{2}}(\gamma)^{\frac{n}{2}}(1+g)^{-\frac{p_z}{2}} \exp\Big(-\frac{\gamma}{2}SSR_g^z\Big)\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}} \Gamma\Big(\frac{\nu_0}{2}\Big)^{-1} \gamma^{\frac{\nu_0}{2}-1} \exp\left(-\gamma\frac{\nu_0\sigma_0^2}{2}\right),$$

which is as the same as Equation 9.9 (Page 165).

A Code

```
# Problem 1
# (c) and (d)
a_1 <- 50;
a_2 <- 50;
b_1 <- 1000;
b_2 <- 1000;
y <- 0;
n <- length(y);
t <- 1000;

set.seed(2);
theta_1 <- rep(NA, t, 1);
theta_2 <- rep(NA, t, 1);
theta_1[1] <- runif(1, 40, 50);
theta_2[1] <- runif(1, 40, 50);

for(i in 2:t) {
  theta_1[i] <- rnorm(1, (n*mean(y)-n*theta_2[i-1]+a_1/b_1^2)/(n+1/b_1^2), sqrt(1/((n+1/b_1^2))));
  theta_2[i] <- rnorm(1, (n*mean(y)-n*theta_1[i]+a_2/b_2^2)/(n+1/b_2^2), sqrt(1/((n+1/b_2^2))));
}

par(mfrow = c(1, 3));
plot(theta_1, type = 'l', xlab = "Iterations", ylab = "theta_1");
plot(theta_2, type = 'l', xlab = "Iterations", ylab = "theta_2");
plot(theta_1+theta_2, type = 'l', xlab = "Iterations", ylab = "mu");
par(mfrow = c(1, 1));

mean(theta_1[(t/2+1):t]+theta_2[(t/2+1):t]);

# Problem 9.2
# (a)
azdiabetes = read.table("azdiabetes.dat", header = TRUE);
y <- as.matrix(azdiabetes[,2]);
X <- as.matrix(cbind(rep(1, dim(y)[1], 1), azdiabetes[,-c(2, 8)]));
g <- length(y);
nu_0 <- 2;
sigma2_0 <- 1;
n <- dim(X)[1];
p <- dim(X)[2];
t <- 10000;

H_g <- (g/(g+1))*X%*%solve(t(X)%*%X)%*%t(X);
SSR_g <- t(y)%*%(diag(1, nrow = n)-H_g)%*%y;
sigma2 <- 1/rgamma(t, (nu_0+n)/2, (nu_0*sigma2_0+SSR_g)/2);
Vb <- g*solve(t(X)%*%X)/(g+1);
Eb <- Vb%*%t(X)%*%y;
```

```
E <- matrix(rnorm(t*p, 0, sqrt(sigma2)), t, p);
beta <- t(t(E%*%chol(Vb))+c(Eb));

for (i in 1:p) {
  plot(density(beta[,i+1]), col = 2, lwd = 1, xlab = names(azdiabetes)[i+1], main = "");
  abline(v = quantile(beta[,i+1], c(0.025, 0.975))[1], col = 3, lwd = 2);
  abline(v = quantile(beta[,i+1], c(0.025, 0.975))[2], col = 3, lwd = 2);
  abline(v = mean(beta[,i+1]), col = 4, lwd = 2);
}

# (b)
lpy.X <- function(y, X) {
  n <- dim(X)[1];
  p <- dim(X)[2];
  g <- length(y);
  nu_0 <- 1;
  sigma2_0 <- try(summary(lm(y ~ -1+X))$sigma^2, silent = TRUE);
  if (p == 0) {
    H_g <- 0;
    sigma2_0 <- mean(y^2);
  }
  if (p > 0) {
    H_g <- (g/(g+1))*X%*%solve(t(X)%*%X)%*%t(X);
  }
  SSR_g <- t(y)%*%(diag(1, nrow = n)-H_g)%*%y;

  -(1/2)*(n*log(pi)+p*log(1+g)+(nu_0+n)*log(nu_0*sigma2_0+SSR_g)-nu_0*log(nu_0*sigma2_0))+
    lgamma((nu_0+n)/2)-lgamma(nu_0/2);
}

t <- 1000;
Z <- matrix(NA, t, p);
B <- matrix(0, t, p);
z <- rep(1, p);
lpy.c <- lpy.X(y, X[, z == 1, drop = FALSE])

for(i in 1:t) {
  for(j in sample(1:p)) {
    zp <- z;
    zp[j] <- 1-zp[j];
    lpy.p <- lpy.X(y, X[, zp == 1, drop = FALSE]);
    r <- (lpy.p-lpy.c)*(-1)^(zp[j] == 0);
    z[j] <- rbinom(1, 1, 1/(1+exp(-r)));
    if (z[j] == zp[j]) {
      lpy.c <- lpy.p;
    }
    Z[i, ] <- z;
```

```
    H_g <- (g/(g+1))*X[, Z[i,] == 1, drop = FALSE]%*%solve(t(X[, Z[i,] == 1, drop = FALSE])%*%
      X[, Z[i,] == 1, drop = FALSE])%*%t(X[, Z[i,] == 1, drop = FALSE]);
    SSR_g <- t(y)%*%(diag(1, nrow = n)-H_g)%*%y;
    sigma2 <- 1/rgamma(t, (nu_0+n)/2, (nu_0*sigma2_0+SSR_g)/2);

    Vb <- g*solve(t(X[, Z[i,] == 1, drop = FALSE])%*%X[, Z[i,] == 1, drop = FALSE])/(g+1);
    Eb <- Vb%*%t(X[, Z[i,] == 1, drop = FALSE])%*%y;
    E <- matrix(rnorm(sum(Z[i, ]), 0, sqrt(sigma2)), 1, sum(Z[i, ]));
    B[i, Z[i,] == 1] <- t(t(E%*%chol(Vb))+c(Eb));
  }
}

for (i in 1:p) {
  sum(B[,i] != 0)/t;
}

for (i in 1:p) {
  plot(density(B[,i+1]), col = 2, lwd = 1, xlab = names(azdiabetes)[i+1], main = "");
  abline(v = quantile(B[,i+1], c(0.025, 0.975))[1], col = 3, lwd = 2);
  abline(v = quantile(B[,i+1], c(0.025, 0.975))[2], col = 3, lwd = 2);
  abline(v = mean(B[,i+1]), col = 4, lwd = 2);
}



# Problem 9.3
# (a)
crime = read.table("crime.dat", header = TRUE);
y <- as.matrix(crime[,1]);
X <- as.matrix(cbind(rep(1, dim(y)[1], 1),crime[,-1]));
g <- length(y);
nu_0 <- 2;
sigma2_0 <- 1;
n <- dim(X)[1];
p <- dim(X)[2];
t <- 10000;

H_g <- (g/(g+1))*X%*%solve(t(X)%*%X)%*%t(X);
SSR_g <- t(y)%*%(diag(1, nrow = n)-H_g)%*%y;
sigma2 <- 1/rgamma(t, (nu_0+n)/2, (nu_0*sigma2_0+SSR_g)/2);
Vb <- g*solve(t(X)%*%X)/(g+1);
Eb <- Vb%*%t(X)%*%y;
E <- matrix(rnorm(t*p, 0, sqrt(sigma2)), t, p);
beta <- t(t(E%*%chol(Vb))+c(Eb));

for (i in 1:p) {
  i <- 15;
  plot(density(beta[,i+1]), col = 2, lwd = 1, xlab = names(crime)[i+1], main = "");
```

```
  abline(v = quantile(beta[,i+1], c(0.025, 0.975))[1], col = 3, lwd = 2);
  abline(v = quantile(beta[,i+1], c(0.025, 0.975))[2], col = 3, lwd = 2);
  abline(v = mean(beta[,i+1]), col = 4, lwd = 2);
}

g = lm(y ~ M+So+Ed+Po1+Po2+LF+M.F+Pop+NW+U1+U2+GDP+Ineq+Prob+Time, data = crime)
summary(g)
MSE = sum(g$residuals^2) / (n - 16);
X_ols = as.matrix(cbind(1, X));
round(g$coefficients + qt(1 - 0.05 / 2, df = n - 16) * sqrt(diag(MSE * solve(t(X_ols) %*% X_ols))), 4);
round(g$coefficients - qt(1 - 0.05 / 2, df = n - 16) * sqrt(diag(MSE * solve(t(X_ols) %*% X_ols))), 4);

# (b)
set.seed(1);
tr_index <- sample(1:n, n/2);
te_index <- (1:n)[-tr_index];
X_tr <- X[tr_index,];
X_te <- X[te_index,];
y_tr <- y[tr_index,];
y_te <- y[te_index,];

beta_ols <- solve(t(X_tr)%*%X_tr)%*%t(X_tr)%*%y_tr;
y_ols <- X_te%*%beta_ols;

plot(y_te, y_ols, xlab = "y_te", ylab = "y_ols", xlim = c(-2, 4), ylim = c(-2, 4), asp = 1);
abline(c(0, 0), c(1, 1), col = "red");

sum((y_ols-y_te)^2)/length(y_te)

# (c)
g <- length(y_tr);
nu_0 <- 2;
sigma2_0 <- 1;
n <- dim(X_tr)[1];
p <- dim(X_tr)[2];
t <- 10000;

H_g <- (g/(g+1))*X_tr%*%solve(t(X_tr)%*%X_tr)%*%t(X_tr);
SSR_g <- t(y_tr)%*%(diag(1, nrow = n)-H_g)%*%y_tr;
sigma2 <- 1/rgamma(t, (nu_0+n)/2, (nu_0*sigma2_0+SSR_g)/2);
Vb <- g*solve(t(X_tr)%*%X_tr)/(g+1);
Eb <- Vb%*%t(X_tr)%*%y_tr;
E <- matrix(rnorm(t*p, 0, sqrt(sigma2)), t, p);
beta_bayes <- colMeans(t(t(E%*%chol(Vb))+c(Eb)));
y_bayes <- X_te%*%beta_bayes;

plot(y_te, y_bayes, xlab = "y_te", ylab = "y_bayes", xlim = c(-2, 4), ylim = c(-2, 4), asp = 1);
abline(c(0, 0), c(1, 1), col = "red");
```

```
sum((y_bayes-y_te)^2)/length(y_te);

# (d)
tt <- 1000;
error_ols <- rep(NA, 1, tt);
error_bayes <- rep(NA, 1, tt);
for (i in 1:tt) {
  set.seed(i);
  n <- 47;
  tr_index <- sample(1:n, n/2);
  te_index <- (1:n)[-tr_index];
  X_tr <- X[tr_index,];
  X_te <- X[te_index,];
  y_tr <- y[tr_index,];
  y_te <- y[te_index,];
  beta_ols <- solve(t(X_tr)%*%X_tr)%*%t(X_tr)%*%y_tr;
  y_ols <- X_te%*%beta_ols;
  error_ols[i] <- sum((y_ols-y_te)^2)/length(y_te)

  g <- length(y_tr);
  nu_0 <- 2;
  sigma2_0 <- 1;
  n <- dim(X_tr)[1];
  p <- dim(X_tr)[2];
  t <- 10000;
  H_g <- (g/(g+1))*X_tr%*%solve(t(X_tr)%*%X_tr)%*%t(X_tr);
  SSR_g <- t(y_tr)%*%(diag(1, nrow = n)-H_g)%*%y_tr;
  sigma2 <- 1/rgamma(t, (nu_0+n)/2, (nu_0*sigma2_0+SSR_g)/2);
  Vb <- g*solve(t(X_tr)%*%X_tr)/(g+1);
  Eb <- Vb%*%t(X_tr)%*%y_tr;
  E <- matrix(rnorm(t*p, 0, sqrt(sigma2)), t, p);
  beta_bayes <- colMeans(t(t(E%*%chol(Vb))+c(Eb)));
  y_bayes <- X_te%*%beta_bayes;
  error_bayes[i] <- sum((y_bayes-y_te)^2)/length(y_te);
}
mean(error_ols)
mean(error_bayes)
```