베이지안통계학 과제10 solution

Spring 2022

문제 13-1. (20점)

- $\hat{\beta}_0 = 29.107, \hat{\beta}_1 = 13.637, se(\hat{\beta}_0) = 15.969, se(\hat{\beta}_1) = 3.149, \hat{\sigma}^2 = 847.19$. (5점, 적절하게 구하지 않은 경우 부분점수 2점 부여)
- $\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}) \sim N(\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})$, 여기서 $\Sigma_{\boldsymbol{\beta}} = \left(\frac{1}{\sigma^2} XX^T\right)^{-1}$ 이고 $\mu_{\boldsymbol{\beta}} = \Sigma_{\boldsymbol{\beta}}\left(\frac{1}{\sigma^2} X^T \boldsymbol{y}\right)$. 그리고 $\pi(\sigma^2|\boldsymbol{\beta}, \boldsymbol{y}) \sim IG\left(\frac{n}{2}, \frac{1}{2} SSR(\boldsymbol{\beta})\right)$. (5점)
- 깁스 표본기법을 사용하여 추정치를 구하고 최소제곱 추정치와 적절히 비교. (5점)
- 사후밀도와 95% HPD구간을 그림으로 나타냄. (5점)

(1) 최소제곱 추정치를 다음과 같이 구한다.

```
library(dplyr)
```

```
##
## 다음의 패키지를 부착합니다: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data = matrix(scan("http://home.ewha.ac.kr/~msoh/Bayesianbook/car.txt"),
              nrow = 14, ncol = 2, byrow = T) %>% data.frame

names(data) = c("X", "Y")

lm_fit = lm(Y ~ X, data)
summary(lm_fit)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data)
##
## Residuals:
```

```
##      Min       1Q  Median       3Q      Max 
## -33.204 -20.383  -4.748  13.957  61.433 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   29.107     15.969   1.823 0.093341 .  
## X             13.637      3.149   4.330 0.000978 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 29.11 on 12 degrees of freedom
## Multiple R-squared:  0.6098, Adjusted R-squared:  0.5773 
## F-statistic: 18.75 on 1 and 12 DF,  p-value: 0.0009779
```

```
sum(lm_fit$residuals^2)/12 # sigma2 추정값
```

```
## [1] 847.1876
```

최소제곱법 결과로부터 β_0 의 추정값과 표준오차는 29.107, 15.969이고 β_1 의 추정값과 표준오차는 13.637, 3.149, 그리고 σ^2 의 lse 추정값은 $\frac{1}{14-2}\sum_{i=1}^{14}(y_i - \hat{y}_i)^2 = 847.19$ 이다.

(2) $\boldsymbol{\beta} = (\beta_0, \beta_1)$ 라 하자. 사전분포를 $\pi(\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sigma^2}$ 으로 가정할 때 $\boldsymbol{\beta}$ 와 σ^2 의 결합사후밀도함수는 다음과 같다 (식 전개는 강의노트 7, 8p 참고).

$$
\begin{aligned}
\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\beta}, \sigma^2) \\
&\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})^T(\boldsymbol{y} - X\boldsymbol{\beta})\right\}\frac{1}{\sigma^2} \\
&= (\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})^T(\boldsymbol{y} - X\boldsymbol{\beta})\right\}
\end{aligned}
$$

그리고 $\boldsymbol{\beta}$ 와 σ^2 의 완전조건부 사후분포를 구해보자. 먼저 $\boldsymbol{\beta}$ 의 완전조건부 사후분포는

$$
\begin{aligned}
\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}) &\propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})^T(\boldsymbol{y} - X\boldsymbol{\beta})\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\beta}^T X^T\boldsymbol{y} + \boldsymbol{\beta}^T X^T X\boldsymbol{\beta})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T\left(\frac{1}{\sigma^2}XX^T\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\left(\frac{1}{\sigma^2}X^T\boldsymbol{y}\right)\right]\right\} \\
&\sim N(\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})
\end{aligned}
$$

를 따르며, 여기서 $\Sigma_{\boldsymbol{\beta}} = \left(\frac{1}{\sigma^2}XX^T\right)^{-1}$ 이고 $\mu_{\boldsymbol{\beta}} = \Sigma_{\boldsymbol{\beta}}\left(\frac{1}{\sigma^2}X^T\boldsymbol{y}\right)$ 이다.

이어서 σ^2 의 완전조건부 사후분포는

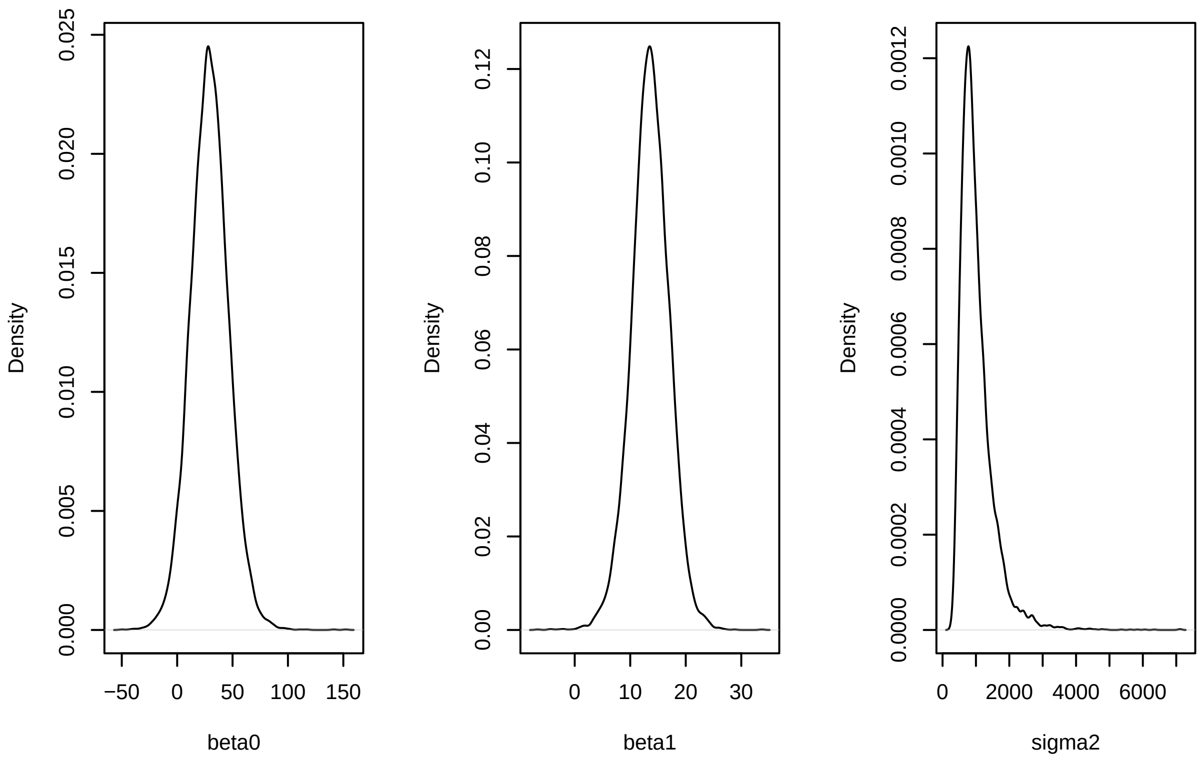
$$
\pi(\sigma^2|\boldsymbol{\beta}, \boldsymbol{y}) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})^T(\boldsymbol{y} - X\boldsymbol{\beta})\right\}
$$

$$
\sim IG\left(\frac{n}{2}, \frac{1}{2}SSR(\boldsymbol{\beta})\right)
$$

을 따른다.

(3) 깁스표본 알고리즘으로부터 10,000개의 샘플을 추출한다. (강의노트 9p 참고)

```r
library(mvtnorm)
library(invgamma)
K = 10000
n = nrow(data)
X = cbind(rep(1, n), data$X)
y = data$Y

# 초기값 설정
sigma2 = 1

beta = matrix(0, 2, K)
for(k in 1:K){
  # beta의 추출
  Sigma_B = solve( (t(X) %*% X)/sigma2[k] )
  mu_B = Sigma_B %*% ( (t(X) %*% y)/sigma2[k] )

  beta[,k] = t(rmvnorm(1, mu_B, Sigma_B))

  # sigma^2의 추출
  SSR = t(y - X %*% beta[,k]) %*% (y - X %*% beta[,k])
  a = n/2
  b = SSR/2
  sigma2[k+1] = rinvgamma(1, a, b)
}

# burn-in 기간 제외
beta = beta[, -c(1:500)]
sigma2 = sigma2[-c(1:500)]

par(mfrow = c(1, 3))
plot(density(beta[1,]), xlab = "beta0", main = "")
plot(density(beta[2,]), xlab = "beta1", main = "")
plot(density(sigma2), xlab = "sigma2", main = "")
```



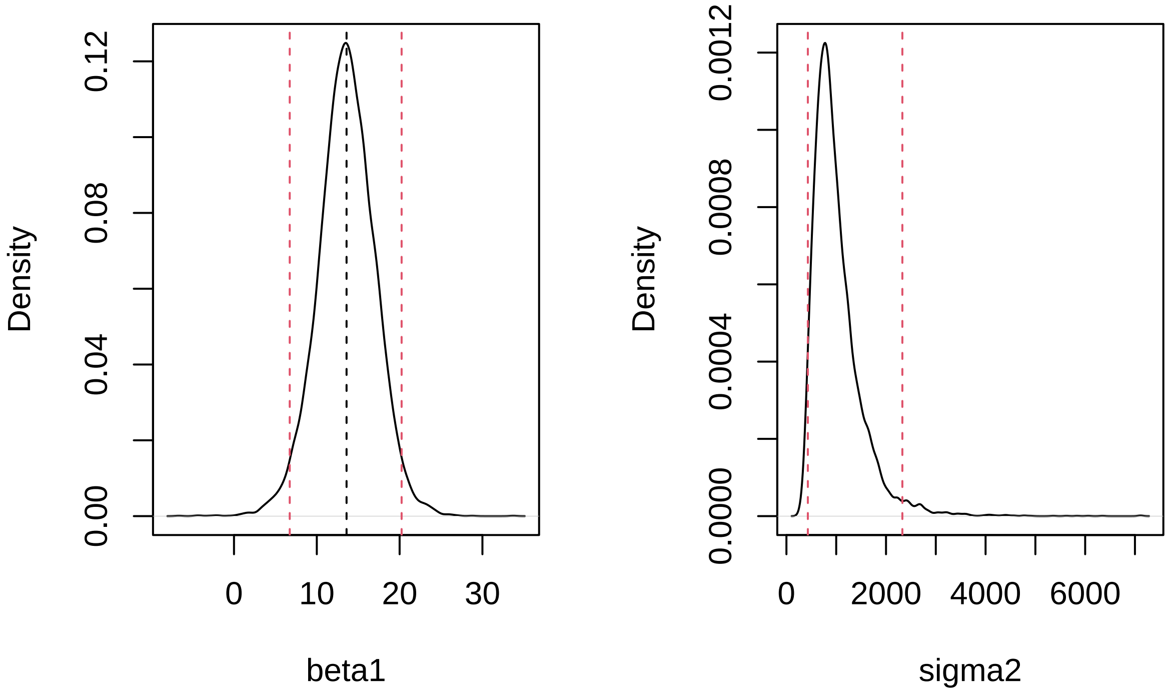

```
cat(" beta0 추정치 =", mean(beta[1,]), "표준오차 =", sd(beta[1,])/sqrt(n), '\n',
    "beta1 추정치 =", mean(beta[2,]),  "표준오차 =", sd(beta[2,])/sqrt(n),'\n',
    "simga2 추정치 =", b/(a - 1), "표준오차 =", sqrt(( b^2/(a-1)^2/(a-2) ) / n))
```

```
##  beta0 추정치 = 29.33865 표준오차 = 4.619683
##  beta1 추정치 = 13.59455 표준오차 = 0.9159843
##  simga2 추정치 = 1175.052 표준오차 = 140.4456
```

최소제곱 방법과 추정치는 비슷하지만 깁스 표본기법에서의 표준오차가 더 작게 나타났다 (σ^2의 평균과 분산은 역감마분포의 평균과 분산으로 계산하였음.).

(4) β_1와 σ^2의 주변 사후밀도함수와 95% HPD구간은 아래 그림과 같다. β_1의 사후구간에 0이 포함되지 않아 유의수준 5% 하에서 기울기가 유의함을 알 수 있다.

```
par(mfrow = c(1, 2))
plot(density(beta[2,]), xlab = "beta1", main = "")
abline( v = c(mean(beta[2,]), quantile(beta[2, ], c(0.025, 0.975))), lty = 2, col = c(1, 2, 2) )
plot(density(sigma2), xlab = "sigma2", main = "")
abline( v = quantile(sigma2, c(0.025, 0.975)), lty = 2, col = 2)
```


0.00
0.04
0.08
0.12
Density
0
10
20
30
beta1
0.0000
0.0004
0.0008
0.0012
Density
0
2000
4000
6000
sigma2