S2208 MATH8050 Data Analysis - Section 001: Homework 8 Due on 11/02/22

Adithya Ravi, C09059838

2022-11-01

```r
# load packages
library(MASS)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

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

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

```r
library(ggplot2)
library(stats)

sessionInfo()
```

Solutions

Question1

1a

$p(w1, w2, w3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_1, ..., Y_N) \propto p(Y_1, ..., Y_N | w1, w2, w3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2) p(w1, w2, w3) p(\varepsilon^2) p(\mu_0) p(\sigma_0^2)$

$$\propto \prod_{j=1}^{N} (\sum_{i=1}^{3} w_i \exp(\frac{1}{\sqrt{2\pi\varepsilon^2}}(Y_j - \mu_i)))[\prod_{k=1}^{3} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp(-\frac{1}{2\sigma_0^2}(\mu_i - \mu_0)^2)](\varepsilon^2)^{-3} \exp(-\frac{2}{\varepsilon^2})(\frac{1}{\sqrt{2\pi 3}}) \exp(-\frac{1}{2.3}\mu_0^2)(\sigma_0^2)^{-3} exp(-\frac{2}{\sigma_0^2}))$$

$$\propto \prod_{j=1}^{N} (\sum_{i=1}^{3} w_i \exp(\frac{1}{\sqrt{2\pi\varepsilon^2}}(Y_j - \mu_i)))[\prod_{k=1}^{3} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp(-\frac{1}{2\sigma_0^2}(\mu_i - \mu_0)^2)](\tau^3)(\exp(-2\tau_0)) \frac{1}{\sqrt{2\pi 3}} \exp(-\frac{1}{2.3}\mu_0^2) \phi_0^3 \exp(-2\phi_0)$$

1b

$$p(w_1, w_2, w_3 | \mu_1, \mu_2, \mu_3, \epsilon^2, Y_1, ..., Y_N)$$

$$\propto \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{1}{2\epsilon^2}(Y_i - \mu_j)^2})$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{1}{2\epsilon^2}} (\sum_{j=1}^{3} w_j e^{Y_i - \mu_j)^2})$$

$$= (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{1}{2\epsilon^2}})^N \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j e^{Y_i - \mu_j)^2})$$

$$\propto \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j e^{Y_i - \mu_j)^2})$$

Consider the following,

$$p(\mu_1 | \mu_2, \mu_3, w_1, w_2, w_3, Y_1, ...Y_N, \epsilon^2, \mu 0, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{\sigma_0^2}(\mu_1 - \mu_0)^2} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{((\frac{1}{2\epsilon^2}(Y_i - \mu_j)^2)}))$$

$$\propto e^{(\mu_1 - \mu_0)^2} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j (e^{(Y_i - \mu_j))^2}))$$

$$p(\mu_2 | \mu_1, \mu_3, w_1, w_2, w_3, Y_1, ...Y_N, \epsilon^2, \mu 0, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{\sigma_0^2}(\mu_2 - \mu_0)^2} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{((\frac{1}{2\epsilon^2}(Y_i - \mu_j)^2)}))$$

$$\propto e^{(\mu_2 - \mu_0)^2} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j (e^{(Y_i - \mu_j))^2}))$$

$$p(\mu_3 | \mu_1, \mu_2, w_1, w_2, w_3, Y_1, ...Y_N, \epsilon^2, \mu 0, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{\sigma_0^2}(\mu_3 - \mu_0)^2} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{((\frac{1}{2\epsilon^2}(Y_i - \mu_j)^2)}))$$

$$\propto e^{(\mu_3 - \mu_0)^2} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j (e^{(Y_i - \mu_j))^2}))$$

Consider the following,

$$p(\epsilon^2 | \mu_1, \mu_2, \mu_3, w_1, w_2, w_3, Y_1, ..., Y_N) \propto (\epsilon^2)^{-3} e^{-\frac{2}{\epsilon^2}} \prod_{i=1}^{N} (\sum_{j=1}^{3} w_j \frac{1}{2\pi\epsilon^2} e^{-\frac{1}{2\epsilon^2}(Y_i - \mu_j)^2})$$

$$\propto (\epsilon^2)^{-3} e^{-\frac{2}{\epsilon^2}} (\frac{1}{2\pi\epsilon^2} e^{-\frac{1}{2\epsilon^2}})^N$$

And,

$$p(\mu_0|\mu_1, \mu_2, \mu_3, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi 3}} e^{-} \frac{1}{2.3} \mu_0^2 \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(\mu_i - \mu_0)^2}$$

$$\propto e^{\mu_0^2} \prod_{i=1}^{3} e^{(\mu_i - \mu_0)^2}$$

$$= exp(\mu_0^2 + \sum_{i=1}^{3} (\mu_i - \mu_0)^2)$$

$$p(\sigma_0^2|\mu_0, \mu_1, \mu_2, \mu_3) \propto \frac{1}{\sqrt{2.\pi.3}} e^{-\frac{1}{2.3}\mu_0^2} \left(\prod_{i=1}^{3} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(\mu_i - \mu_0)^2} \right) \frac{2^2}{\Gamma(2)} (\sigma_0^2)^{-3} e^{-\frac{2}{\sigma_0^2}}$$

$$\propto (\sigma_0^2)^{-3} e^{-\frac{2}{\sigma_0^2}} \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(\mu_i - \mu_0)^2}$$

1c

$$p(w_1, w_2, w_3|\mu_1, \mu_2, \mu_3, \epsilon^2, Y_1, ..., Y_N, Z_1, ..., Z_N)$$

$$\propto \prod_{i=1}^{N} (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{-} \frac{1}{2\epsilon^2} (Y_i - \mu z_i)^2$$

$$\propto \prod_{i=1}^{N} (e^(Y_i - \mu z_i)^2)$$

And,

$$p(\mu_1|\mu_2, \mu_3, w_1, w_2, w_3, Y_1, ...Y_N, \epsilon^2, \mu_0, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{\sigma_0^2}(\mu_1 - \mu_0)^2} \prod_{i=1}^{N} (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{((\frac{1}{2\epsilon^2}(Y_i - \mu_{z_i})^2)})$$

$$\propto e^{(\mu_1 - \mu_0)^2} \sum_{i=1}^{N} (e^{(Y_i - \mu_{z_i})^2})$$

$$p(\mu_2|\mu_1, \mu_3, w_1, w_2, w_3, Y_1, ...Y_N, \epsilon^2, \mu_0, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{\sigma_0^2}(\mu_2 - \mu_0)^2} \prod_{i=1}^{N} (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{((\frac{1}{2\epsilon^2}(Y_i - \mu_{z_i})^2)})$$

$$\propto e^{(\mu_2 - \mu_0)^2} \sum_{i=1}^{N} (e^{(Y_i - \mu_{z_i})^2})$$

$$p(\mu_3|\mu_1, \mu_2, w_1, w_2, w_3, Y_1, ...Y_N, \epsilon^2, \mu_0, \sigma_0^2) \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{\sigma_0^2}(\mu_3 - \mu_0)^2} \prod_{i=1}^{N} (\frac{1}{\sqrt{2\pi\epsilon^2}} e^{((\frac{1}{2\epsilon^2}(Y_i - \mu_{z_i})^2)})$$

$$\propto e^{(\mu_3 - \mu_0)^2} \sum_{i=1}^{N} (e^{(Y_i - \mu_{z_i})^2})$$

1d

```
data=read.csv("Mixture.csv")
data=data.frame(success=19,size=57)
sum(data[,1])
```

```
## [1] 19
```

```
sum(data[1,])
```

```
## [1] 76
```

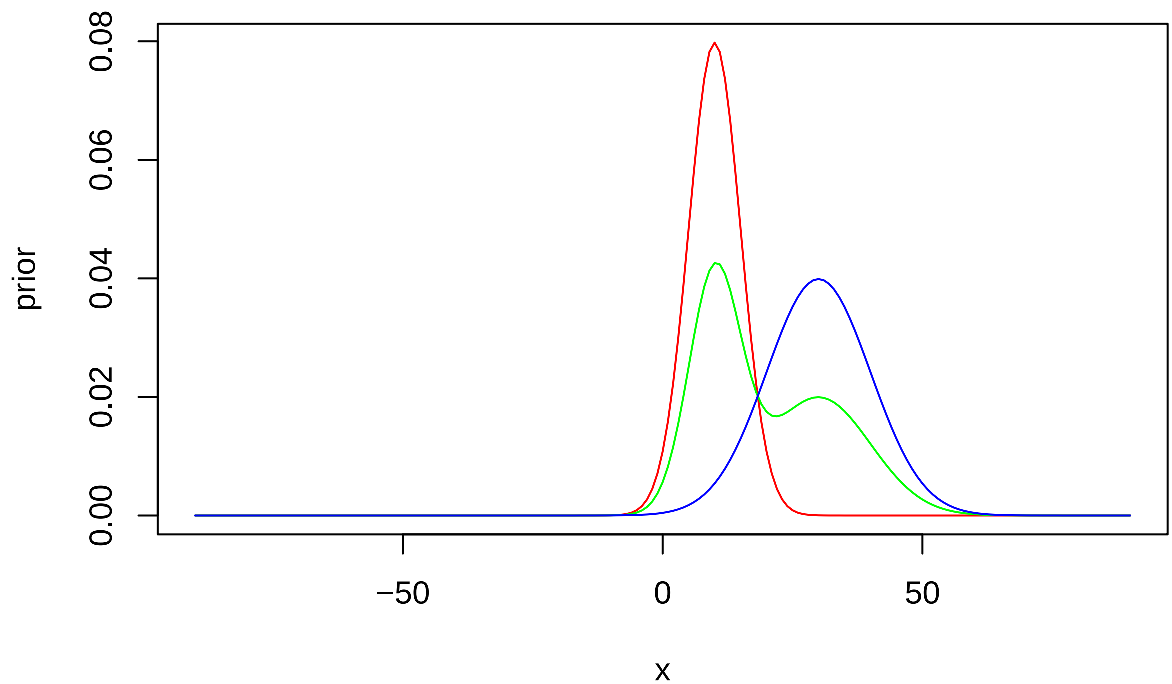
```
data[1,1]/sum(data[,1])
```

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

```
print(sum)
```

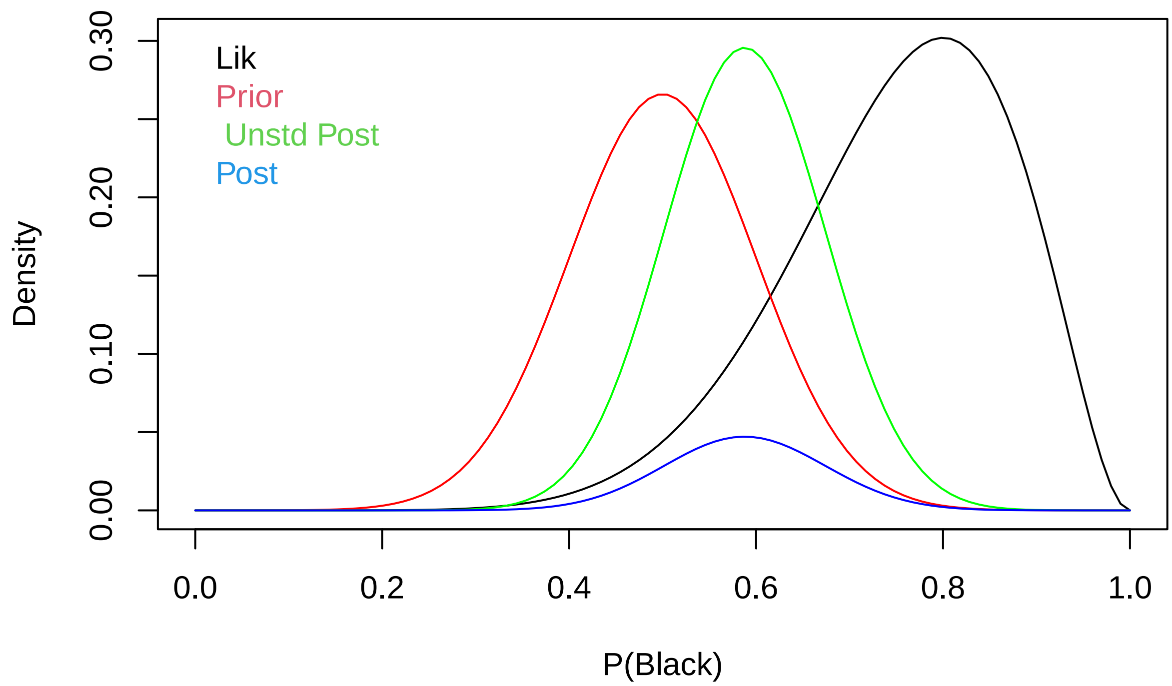
```
## function (..., na.rm = FALSE)  .Primitive("sum")
```

```
x <- seq(from = -90, to = 90, by = 1)
data <- dnorm(x, mean = 30, sd = 10)
prior <- dnorm(x, mean = 10, sd = 5)
posterior <- 0.5 * dnorm(x, mean = 10, sd = 5) +
  0.5 * dnorm(x, mean = 30, sd = 10)
plot(x, prior, type = "l", col = "red")
lines(x, posterior, type = "l", col = "green")
lines(x, data, type = "l", col = "blue")
```

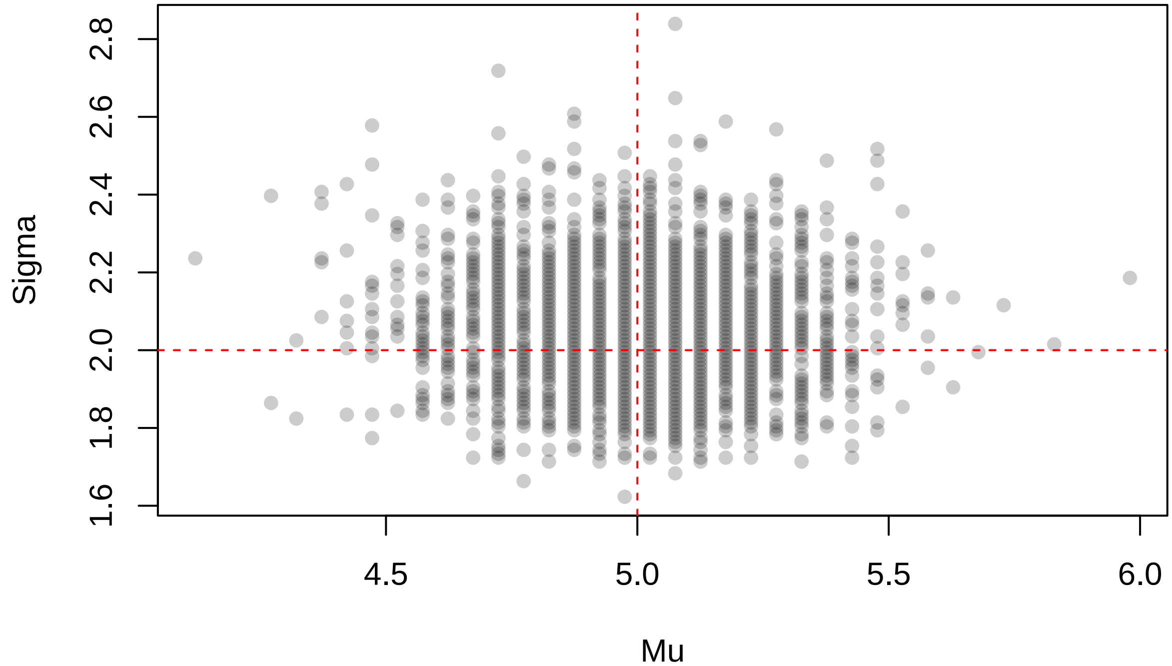


1e

```
rangeP = seq(0,1,length.out = 100)
plot(rangeP, dbinom(x=8, prob=rangeP,size=10),
     type = "l", xlab="P(Black)", ylab="Density")
lines(rangeP,dnorm(x=rangeP, mean=0.5,sd=0.1)/15,col="red")
lik <- dbinom(x=8, prob = rangeP, size =10)
prior <- dnorm(x = rangeP, mean = 0.5, sd = 0.1)
lines(rangeP, lik*prior, col="green")
unstdPost<- lik * prior
stdPost <- unstdPost / sum(unstdPost)
lines(rangeP, stdPost, col = "blue")
legend("topleft",
       legend = c("Lik", "Prior", " Unstd Post", "Post"),
       text.col = 1:4, bty="n")
```




```r
trueMu <- 5
trueSig <- 2
set.seed(100)
randomSample <- rnorm(100, trueMu, trueSig)
grid<-expand.grid(mu = seq(0,10,length.out=200),
                  sigma = seq(1, 3, length.out = 200))
lik<- sapply(1:nrow(grid), function(x){
  sum(dnorm(x=randomSample,
            mean = grid$mu[x], sd = grid$sigma[x], log = T))
})
prod <- lik + dnorm(grid$mu, mean = 0, sd = 5, log = T)+
  dexp(grid$sigma, 1, log=T)
prob <- exp(prod-max(prod))
postSample <- sample(1:nrow(grid), size=1e3, prob=prob)
plot(grid$mu[postSample], grid$sigma[postSample], xlab="Mu", ylab="Sigma",
     pch=16, col=rgb(0,0,0,0.2))
abline(v=trueMu, h=trueSig, col="red", lty=2)
```



Question2

2a

```
swim=read.table("swim.dat")
```

```
library(MASS)
    library(dplyr)
S = 5000
X = cbind(rep(1, 6), seq(1, 11, by = 2))
n = dim(X)[1]
p = dim(X)[2]
# Prior
beta0 = c(23, 0)
sigma0 = rbind(c(0.25, 0), c(0, 0.1))
nu0 = 1
s20 = 0.25
set.seed(1)
inv = solve

swim_pred = apply(swim, MARGIN = 1, function(y) {

BETA = matrix(nrow = S, ncol = length(beta0))
```

```
SIGMA = numeric(S)

beta = c(23, 0)
s2 = 0.7^2

for (s in 1:S) {

V = inv(inv(sigma0) + (t(X) %*% X) / s2)
m = V %*% (inv(sigma0) %*% beta0 + (t(X) %*% y) / s2)

beta = mvrnorm(1, m, V)

ssr = (t(y) %*% y) - (2 * t(beta) %*% t(X) %*% y) + (t(beta) %*% t(X) %*% X %*% beta)

s2 = 1 / rgamma(1, (nu0 + n) / 2, (nu0 * s20 + ssr) / 2)
BETA[s, ] = beta
SIGMA[s] = s2
}

xpred = c(1, 13)
YPRED = rnorm(S, BETA %*% xpred, sqrt(SIGMA))
YPRED
})
```

2b

```
fastest_times = apply(swim_pred, MARGIN = 1, FUN = which.min)
table(fastest_times) / length(fastest_times)
```

```
## fastest_times
##      1      2      3      4
## 0.6524 0.0134 0.3060 0.0282
```

2c

We notice with our posterior predictive dataset that Swimmer 1 is the fastest about 65% of the time by week 13, so we recommend Swimmer 1 for the race.