week4_hw

Seungjun Lee

2021 3 30

```
#### Exercise 7.3
# data
bluecrab = as.matrix(read.table(url("http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/bluecrab.dat")))
orangecrab = as.matrix(read.table(url("http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/orangecrab.dat")))

# a
# blue crab
n = nrow(bluecrab)
ybar = colMeans(bluecrab)
Mu0 = c(ybar)
Sigma = cov(bluecrab)
S0 = Lambda0 = Sigma
nu0 = 4

# Gibbs Sampler
S = 10000
MU = matrix(NA, nrow =S, ncol= 2)
SIGMA = matrix(NA, nrow = S, ncol = 4)

for(s in 1:S){
  # update MU
  Lambdan = solve(solve(Lambda0) + n*solve(Sigma))
  Mun = Lambdan %*% (solve(Lambda0) %*% Mu0 + n*solve(Sigma) %*% ybar)
  Mu = MASS::mvrnorm(n=1, Mun, Lambdan)

  # update Sigma
  Sn = S0 + (t(bluecrab) - c(Mu)) %*% t(t(bluecrab) - c(Mu))
  # notation -> Sn = S0 = Smu
  # Smu = sum(yi-mu)(y-mu)T
  Sigma = solve(rWishart(1, nu0 +n , solve(Sn))[,,1])

  MU[s,] = Mu
  SIGMA[s,] =c(Sigma)
}

# orange crab

n = nrow(orangecrab)
```

```r
ybar = colMeans(orangecrab)
Mu0 = c(ybar)
Sigma = cov(orangecrab)
S0 = Lambda0 = Sigma
nu0=4

# Gibbs Sampler
S = 10000
MU1 = matrix(NA, nrow =S, ncol= 2)
SIGMA1 = matrix(NA, nrow = S, ncol = 4)

for(s in 1:S){
  # update MU
  Lambdan = solve(solve(Lambda0) + n*solve(Sigma))
  Mun = Lambdan %*% (solve(Lambda0) %*% Mu0 + n*solve(Sigma) %*% ybar)
  Mu = MASS::mvrnorm(n=1, Mun, Lambdan)

  # update Sigma
  Sn = S0 + (t(orangecrab) - c(Mu)) %*% t(t(orangecrab) - c(Mu))
  # notation -> Sn = S0 = Smu
  # Smu = sum(yi-mu)(y-mu)T
  Sigma = solve(rWishart(1, nu0 +n , solve(Sn))[,,1])

  MU1[s,] = Mu
  SIGMA1[s,] =c(Sigma)
}

# b
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: replacing previous import 'vctrs::data_frame' by 'tibble::data_fr
ame'
## when loading 'dplyr'
```

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

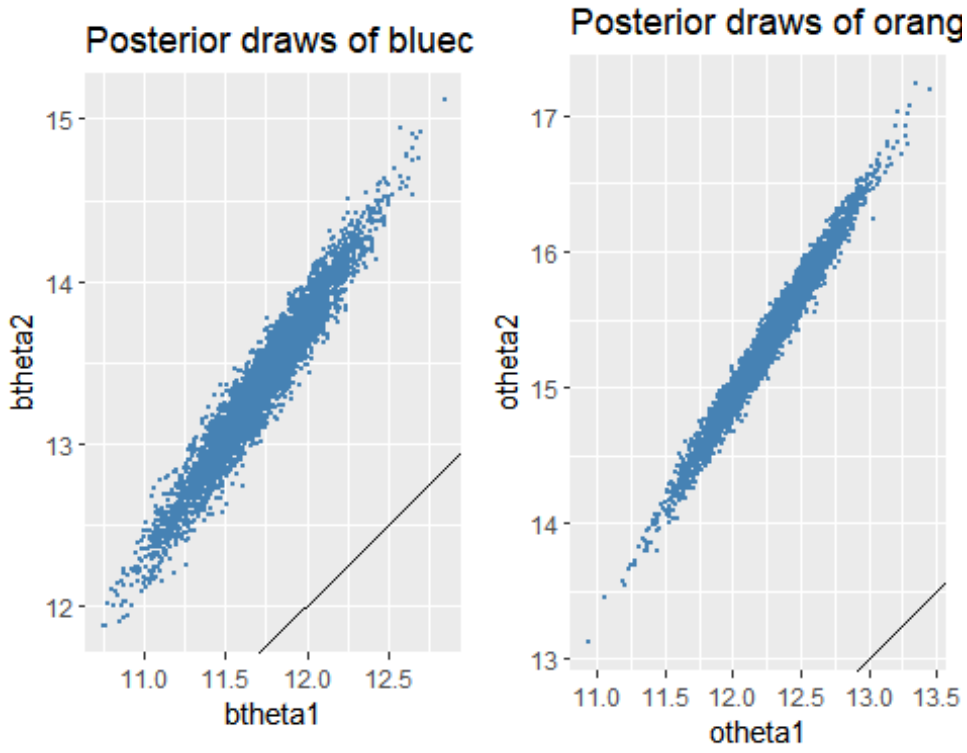
```
## -- Attaching packages ----------------------------- tidyverse 1.3.0 --
```

```
## √ tibble  3.0.3     √ dplyr   1.0.1
## √ tidyr   1.1.3     √ stringr 1.4.0
## √ readr   1.3.1     √ forcats 0.5.0
## √ purrr   0.3.4
```

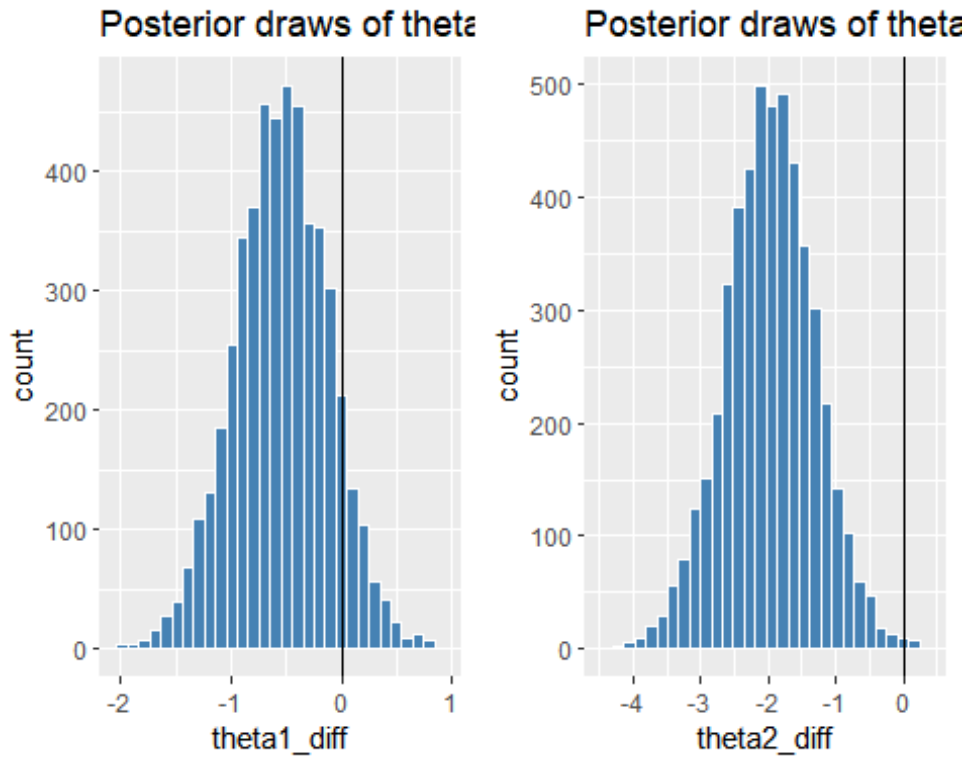
```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

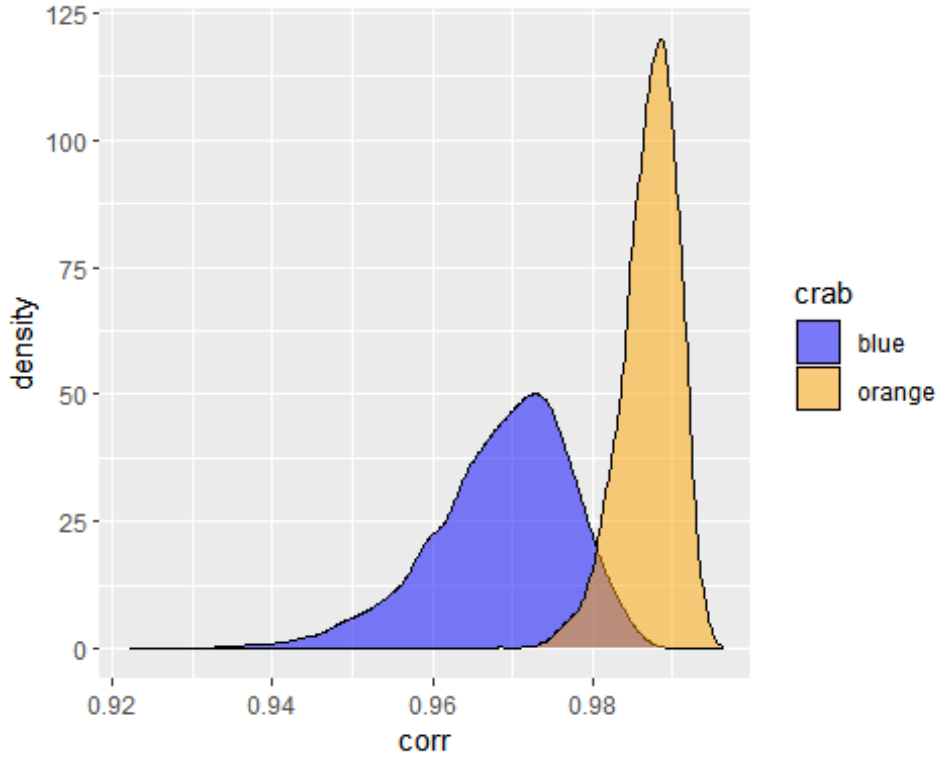
```
## Warning: package 'readr' was built under R version 3.6.3

## Warning: package 'purrr' was built under R version 3.6.3

## Warning: package 'dplyr' was built under R version 3.6.3

## Warning: package 'forcats' was built under R version 3.6.3

## -- Conflicts --------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 3.6.3

disp = tail(1:S, S/2)
title1 = "Posterior draws of bluecrab size"
p1 = data.frame(btheta1 = MU[disp,1], btheta2 = MU[disp,2]) %>%
  ggplot(aes(x=btheta1, y=btheta2)) + geom_point(size= 0.5, color ="steelblue
") +
  geom_abline(slope = 1, intercept = 0 ) +
  coord_fixed(ratio = 1 ) +
  labs(title = title1)

title2 = "Posterior draws of orangecrab size"
p2 = data.frame(otheta1 = MU1[disp,1], otheta2 = MU1[disp,2]) %>%
  ggplot(aes(x=otheta1, y=otheta2)) + geom_point(size= 0.5, color ="steelblue
") +
  geom_abline(slope = 1, intercept = 0 ) +
  coord_fixed(ratio = 1 ) +
  labs(title = title2)

ggarrange(p1, p2)
```




```
title3 = "Posterior draws of theta1 difference"
theta1_diff = MU[disp, 1] - MU1[disp, 1]
p3 = data.frame(theata1_diff = theta1_diff) %>%
  ggplot(aes(x=theta1_diff)) +
  geom_histogram(color ="white", fill = "steelblue", bins = 30) +
  geom_vline(xintercept = 0 ) +
  labs(title=title3)

title4 = "Posterior draws of theta2 difference"
theta2_diff = MU[disp, 2] - MU1[disp, 2]
p4 = data.frame(theata2_diff = theta2_diff) %>%
  ggplot(aes(x=theta2_diff)) +
  geom_histogram(color ="white", fill = "steelblue", bins = 30) +
  geom_vline(xintercept = 0 ) +
  labs(title=title4)

ggarrange(p3, p4)
```




```
mean(MU[disp, 1] > MU1[disp, 1])

## [1] 0.097

mean(MU[disp, 2] > MU1[disp, 2])

## [1] 0.0022

# c
bcorr = apply(SIGMA, MARGIN = 1, FUN = function(SIGMA){
  SIGMA[2] / sqrt(SIGMA[1]*SIGMA[4])
})

ocorr = apply(SIGMA1, MARGIN = 1, FUN = function(SIGMA){
  SIGMA[2] / sqrt(SIGMA[1]*SIGMA[4])
})

p5 = data.frame(crab = c(rep('blue',length(bcorr)/2), rep("orange",length(oco
rr)/2)),
                corr = c(bcorr[disp], ocorr[disp])) %>%
  ggplot(aes(x = corr, fill = crab)) +
  geom_density(alpha =0.5) +
  scale_fill_manual(values = c('blue', 'orange'))

ggarrange(p5)
```




```
mean(bcorr<ocorr)
```

```
## [1] 0.9883
```