

Homework 7

STA 360: Assignment 7, Fall 2020

Due Monday November 2nd, 5 PM Standard Eastern Time

Question 1

Hoff 7. 3

Part A.

```
#Read in the data sets from github
blue_crab_data = as.matrix(read.table(url('https://raw.githubusercontent.com/resteorts/modern-bayes/master/data/BlueCrabData.csv')))
orange_crab_data = as.matrix(read.table(url('https://raw.githubusercontent.com/resteorts/modern-bayes/master/data/OrangeCrabData.csv')))

#We need to sample from the posterior for both data sets, so let's use lapply with a function
#The function is heavily based on Rebecca Steort's code in Module 8
crab.mcmc = lapply(list('blue_crab' = blue_crab_data, 'orange_crab' = orange_crab_data), function(crab) {
  n = nrow(crab)
  # Set prior parameters according to problem
  mu0 = colMeans(crab)
  lambda0 = cov(crab)
  s0 = cov(crab)
  nu0 = 4
  THETA <- SIGMA <- CORR <- NULL
  # Start with sigma sample
  sigma = s0
  set.seed(1);
  for (s in 1:10000) {
    # Update theta
    Ln <- solve(solve(lambda0) + n * solve(sigma))
    mun <- Ln %*% (solve(lambda0) %*% mu0 + n * solve(sigma) %*% mu0)
    theta <- rmvnorm(1, mun, Ln)

    ## update sigma
    Sn <- s0 + (t(crab) - c(theta)) %*% t(t(crab) - c(theta))
    sigma <- solve(rwish(nu0 + n, solve(Sn)))

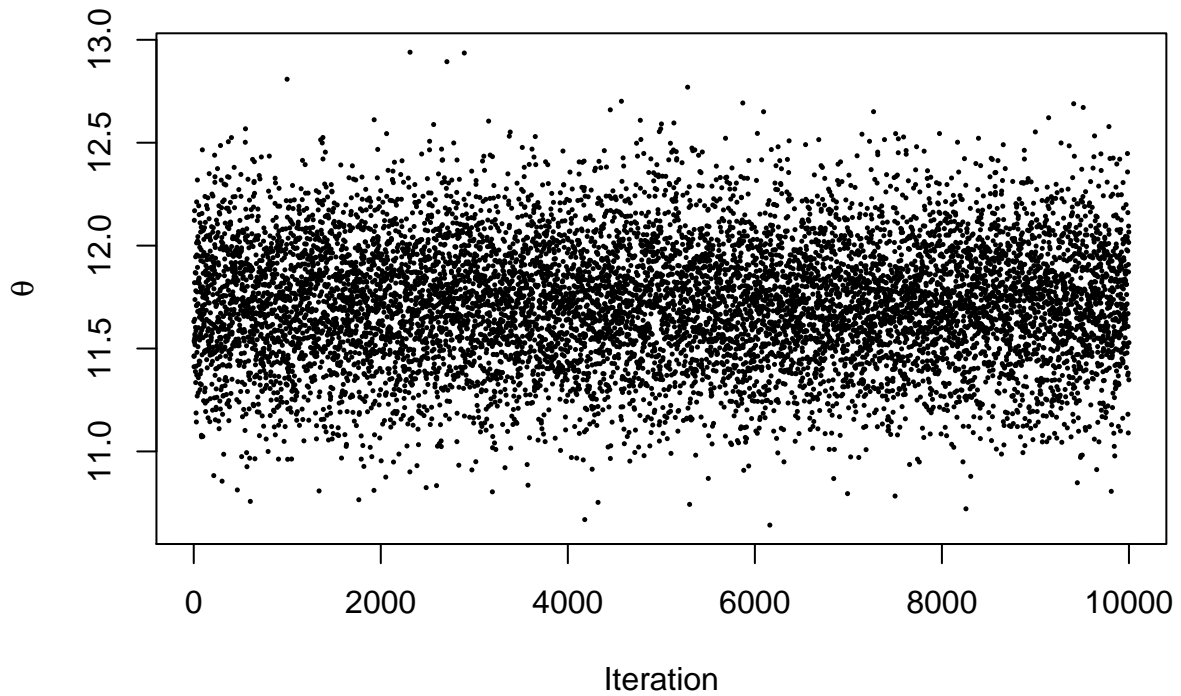
    #Correlation Coefficient
    corr <- sigma[2]/(sqrt(sigma[1] * sigma[4]))

    ## save results
    THETA <- rbind(THETA, theta)
    SIGMA <- rbind(SIGMA, c(sigma))
    CORR <- rbind(CORR, corr)
  }
})
```

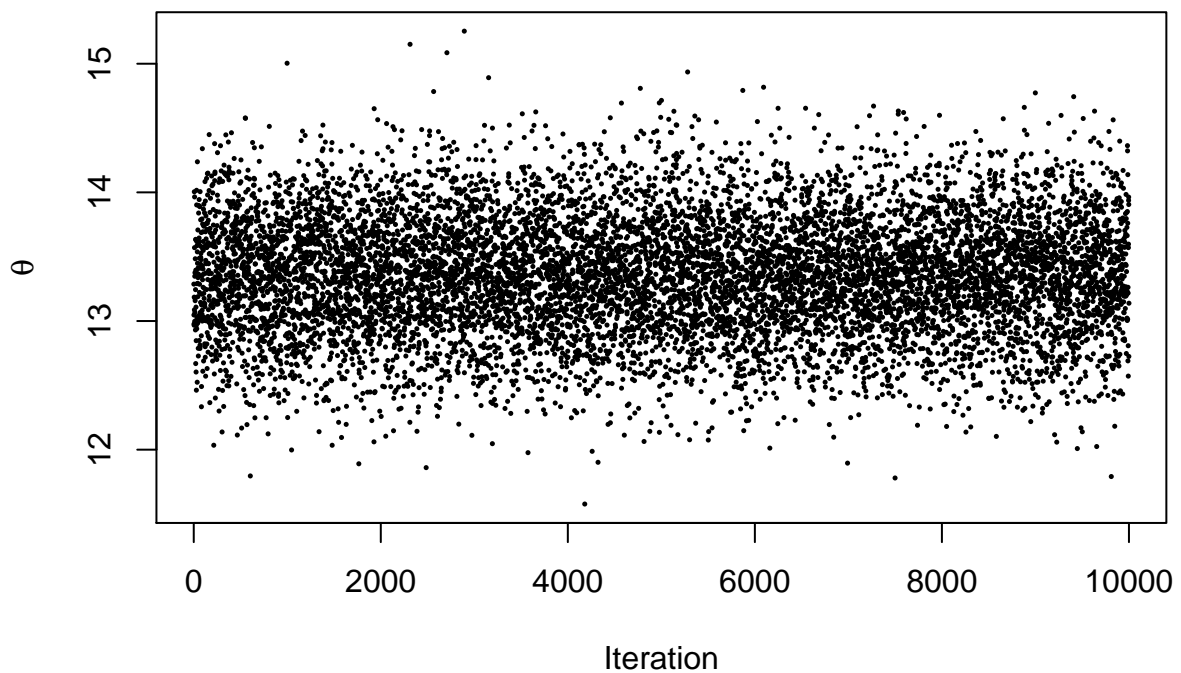
```
list(theta = THETA, sigma = SIGMA, corr = CORR)
})
```

Part B.

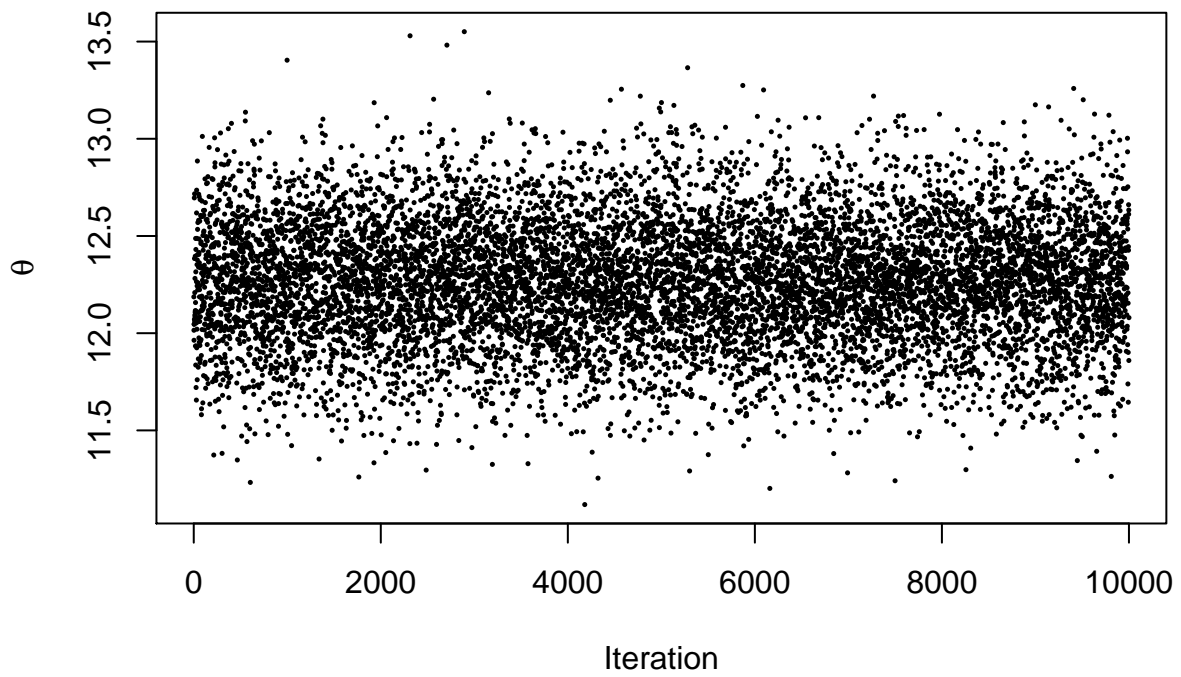
Traceplot of θ_1 for Blue Crab



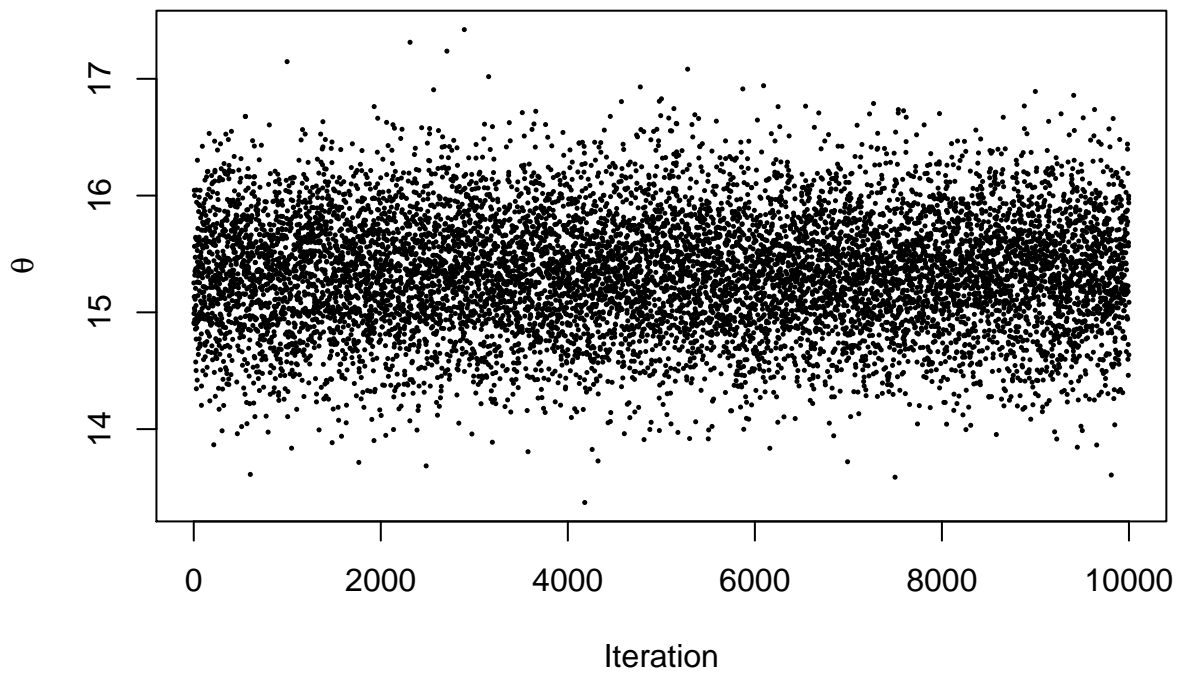
Traceplot of θ_2 for Blue Crab



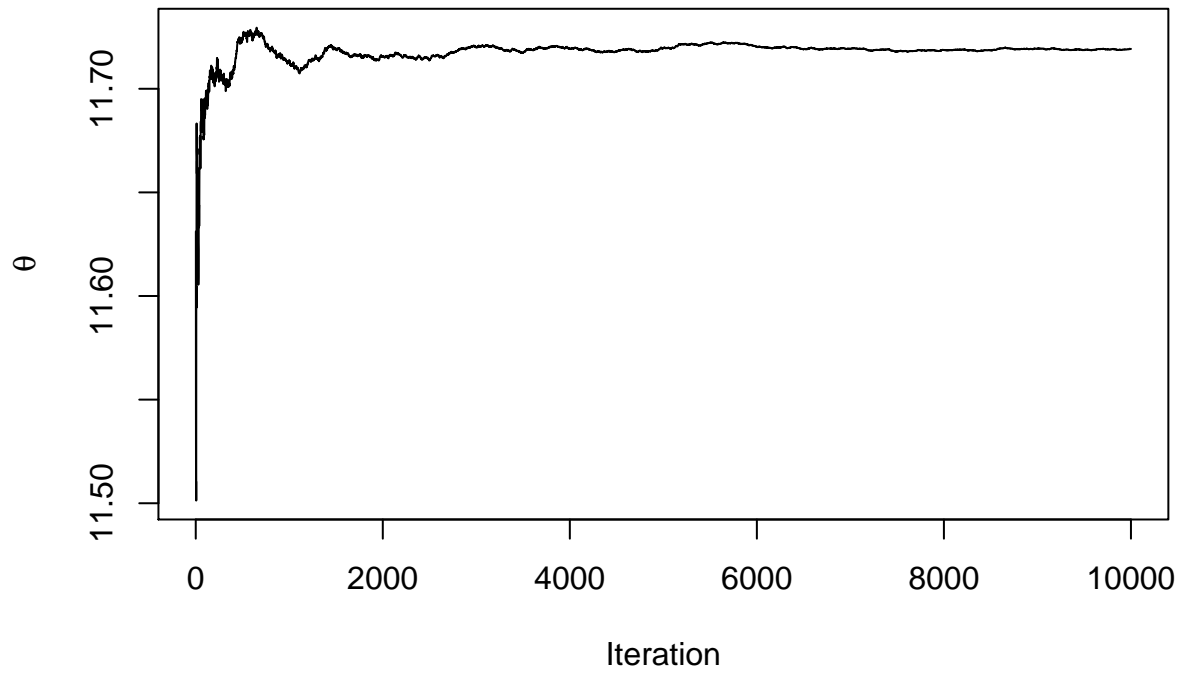
Traceplot of θ_1 for Orange Crab



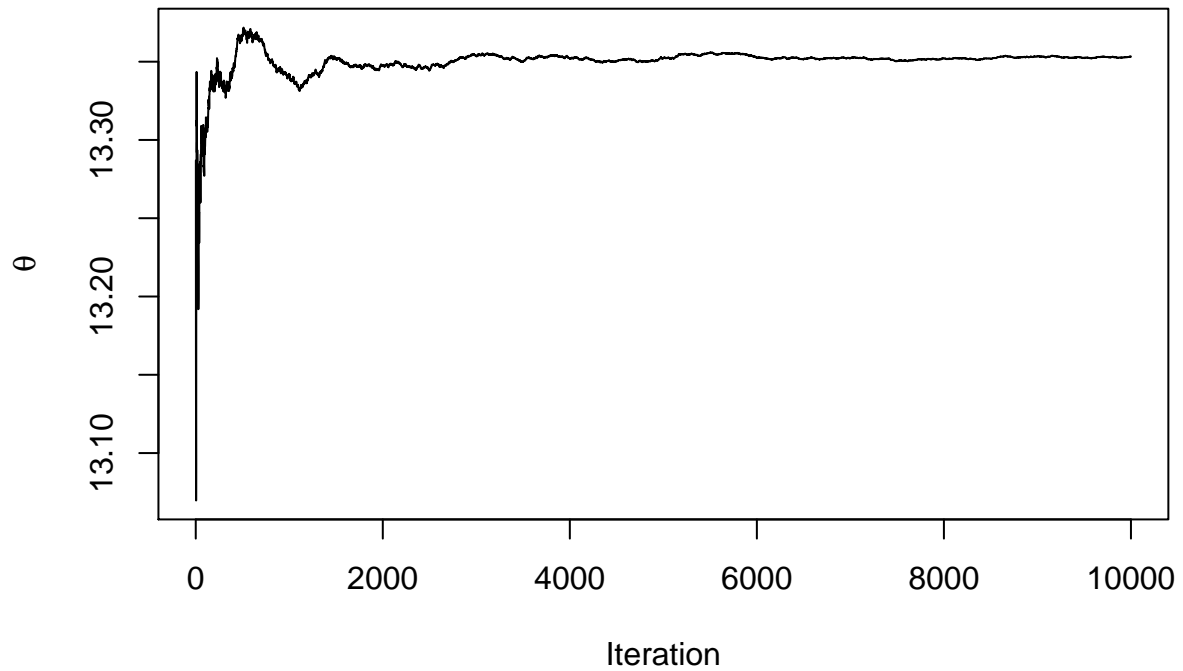
Traceplot of θ_2 for Orange Crab



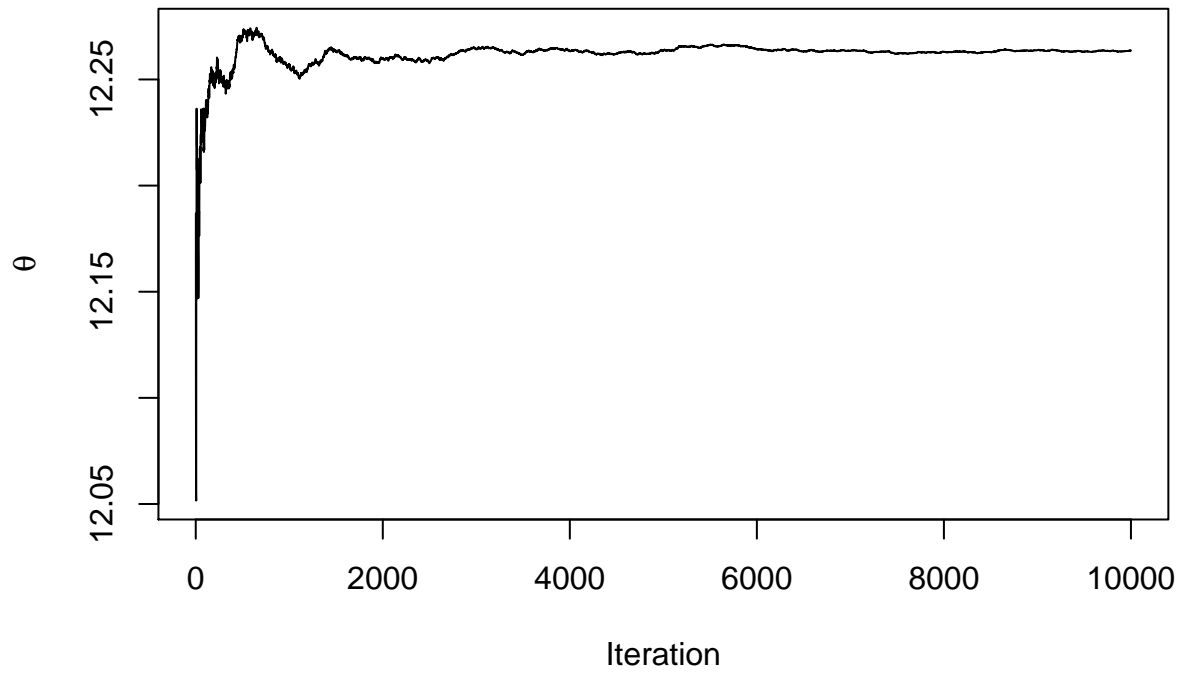
Running Average Plot of θ_1 for Blue Crab



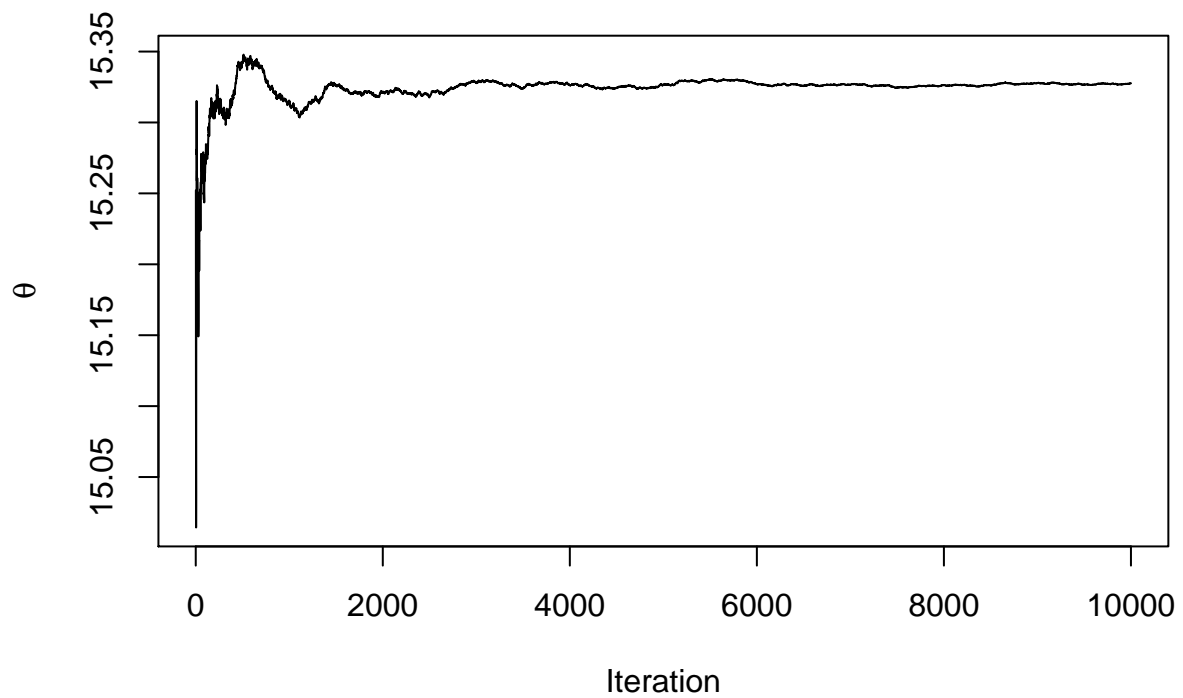
Running Average Plot of θ_2 for Blue Crab



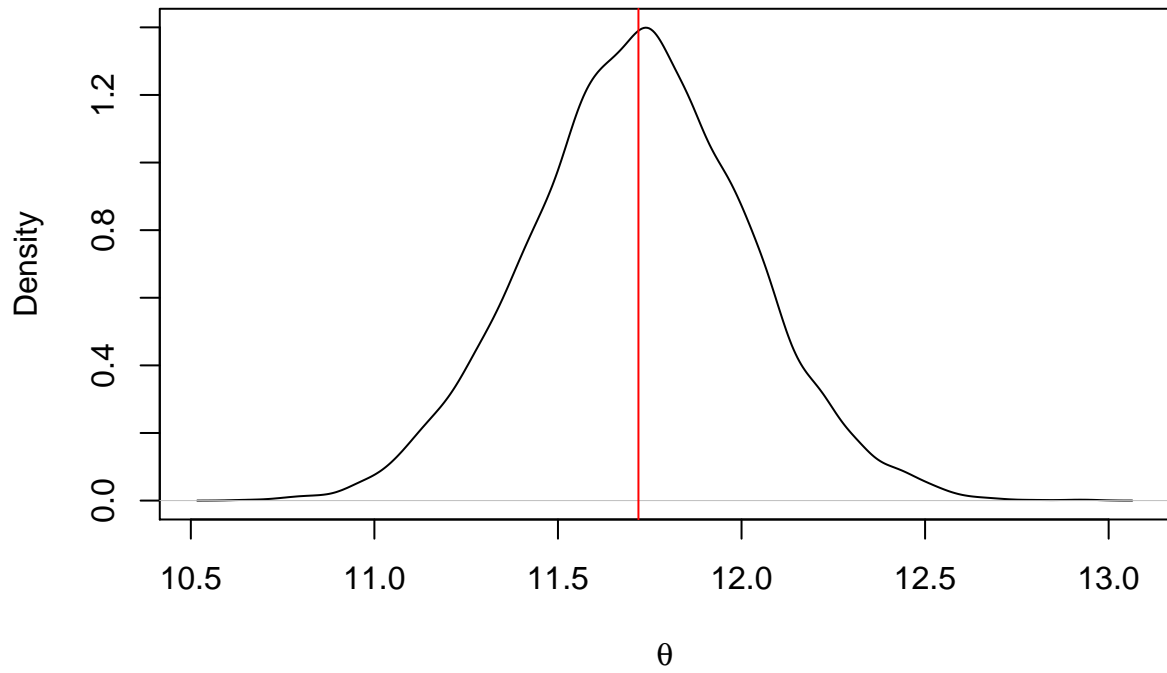
Running Average Plot of θ_1 for Orange Crab



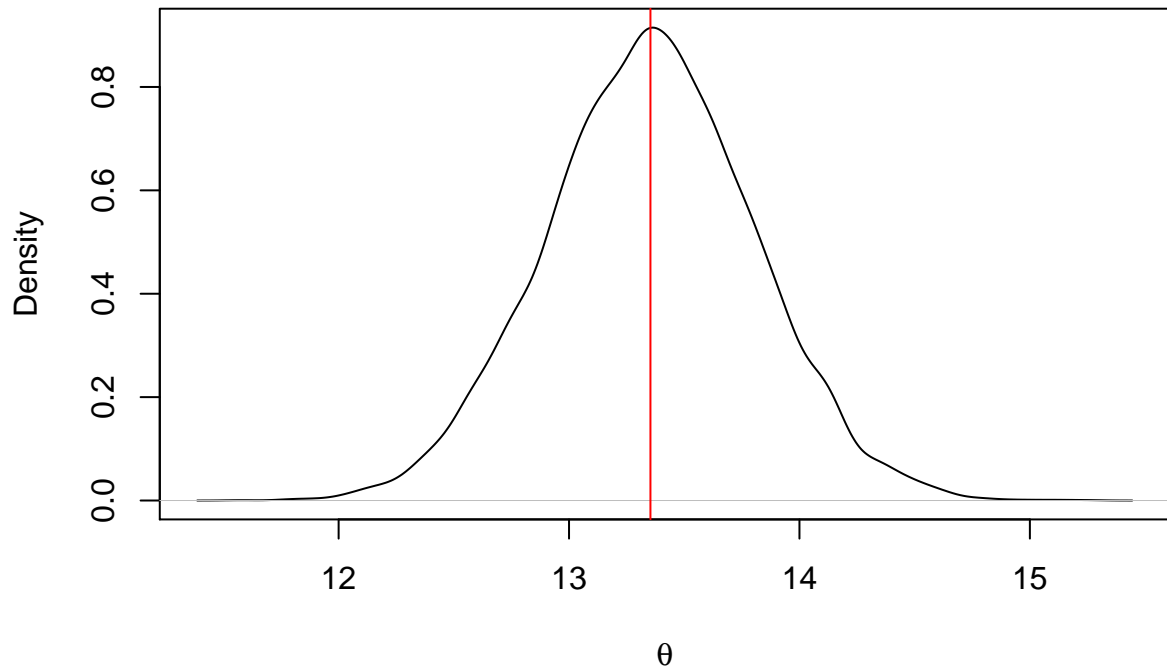
Running Average Plot of θ_2 for Orange Crab



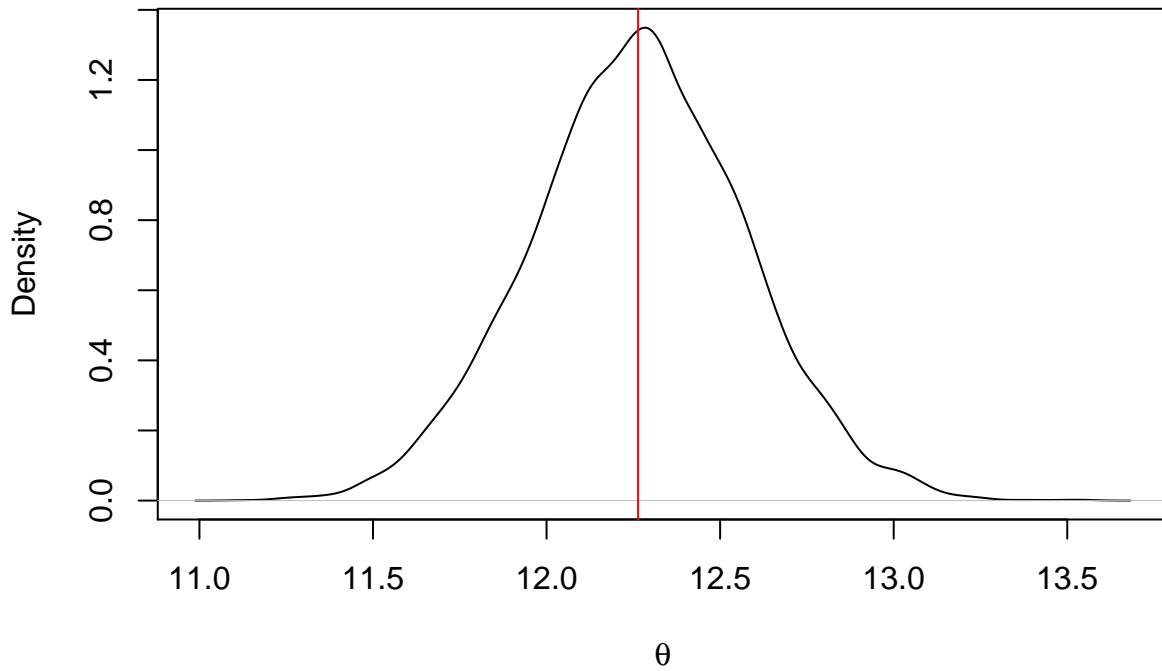
Density of θ_1 for Blue Crab



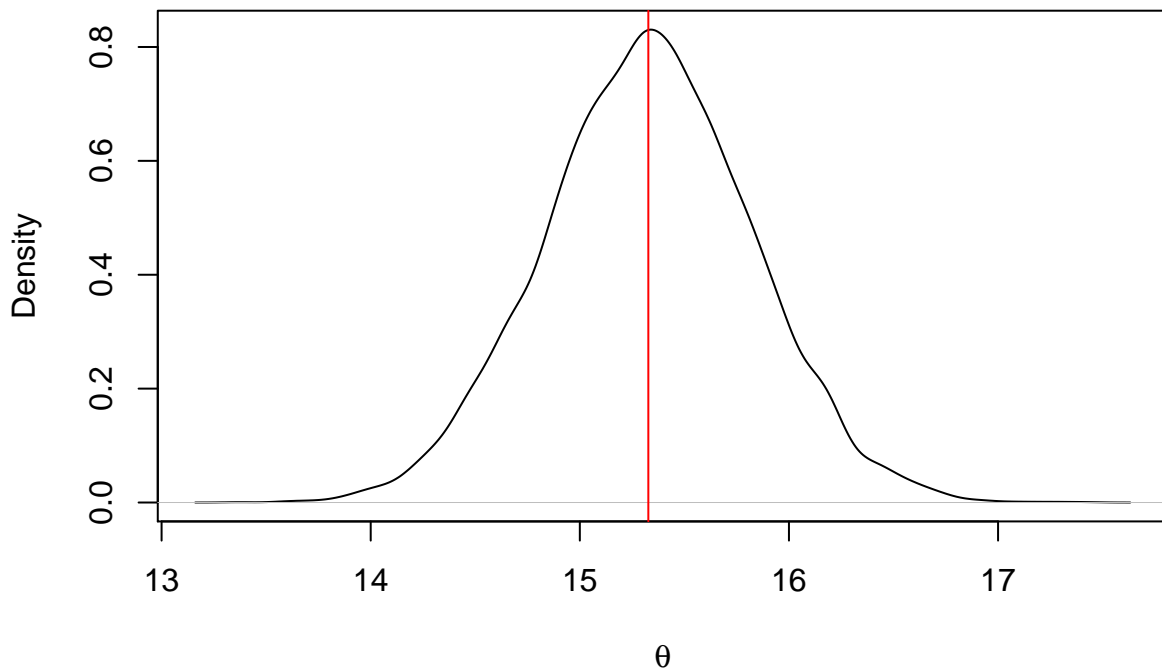
Density of θ_2 for Blue Crab



Density of θ_1 for Orange Crab



Density of θ_2 for Orange Crab

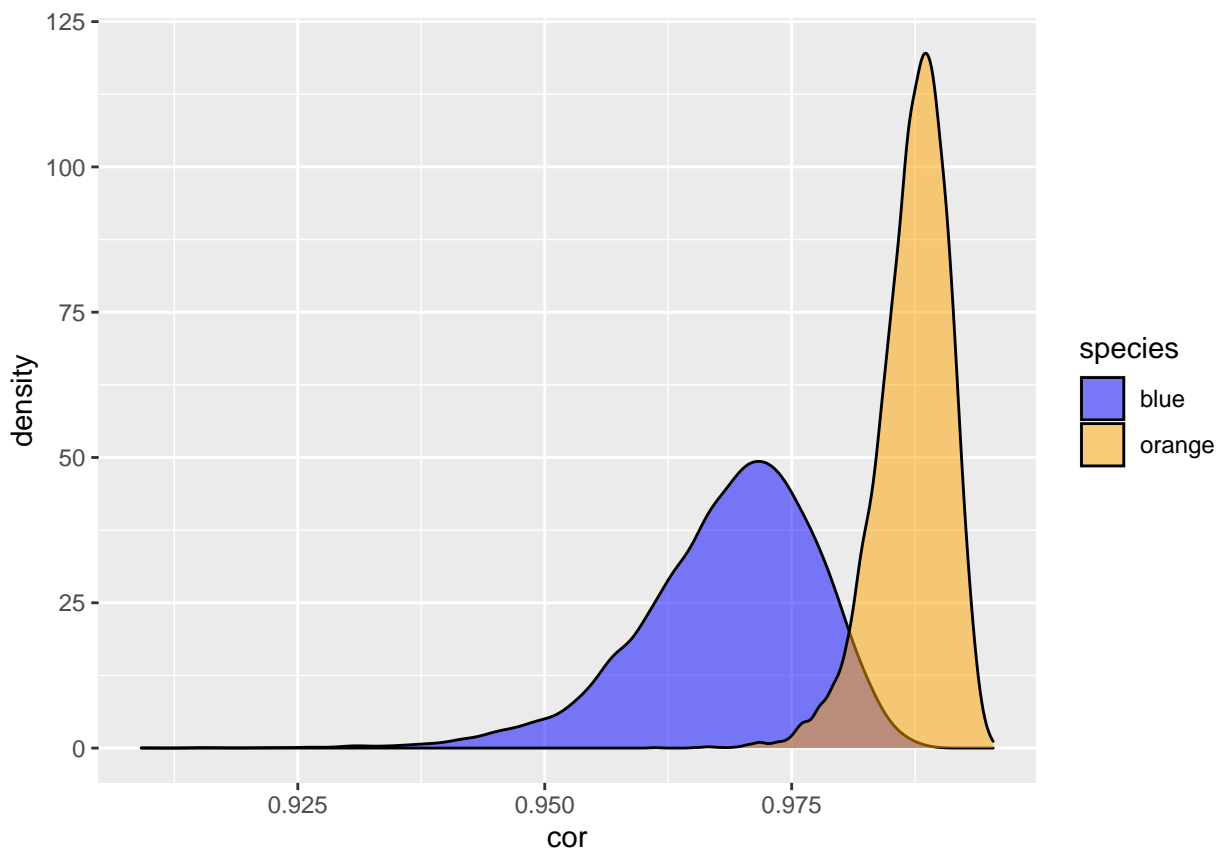


From the plots above, we can see that the chain has seemingly not failed to converge. And we can see that for body width and for rear width, the orange crabs seem to be bigger. The density and running average plots both reveal, that the orange crabs have larger values for body with and rear width.

Part C.

```
cor.df = data.frame(species = c(rep('blue', length(crab.mcmc$blue_crab$corr)), rep('orange', length(crab.mcmc$orange_crab$corr))),  
  cor = c(crab.mcmc$blue_crab$corr, crab.mcmc$orange_crab$corr))
```

```
ggplot(cor.df, aes(x = cor, fill = species)) +  
  geom_density(alpha = 0.5) +  
  scale_fill_manual(values = c('blue', 'orange'))
```



```
mean(crab.mcmc$blue_crab$corr < crab.mcmc$orange_crab$corr)
```

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