Computer Lab 1

Andrea Bruzzone, Thomas Zhang

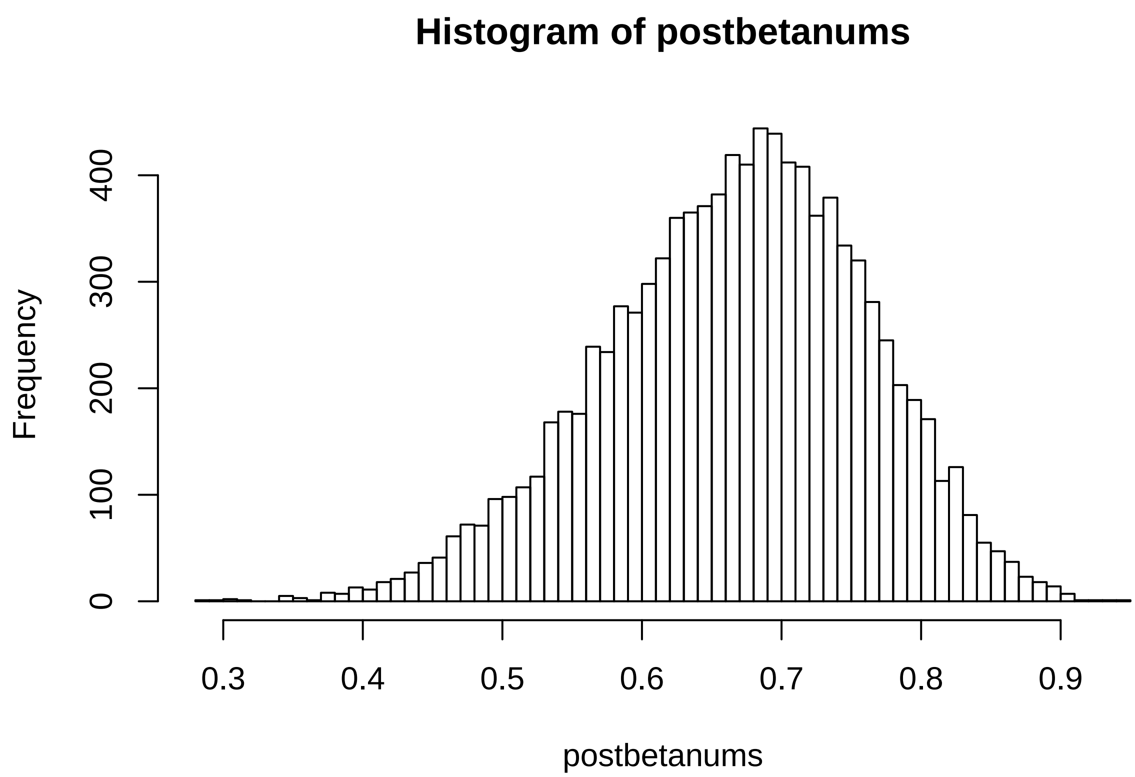
2016 M04 6

Assignment 1

We take the numbers given, put them into the formulae and calculate theoretical mean and std. deviation. Then follows a histogram over the posterior draws of θ, and the sample mean and sample s.d. of θ.

```
## [1] "Theoretical posterior beta distr. mean: 0.667"

## [1] "Theoretical posterior beta distr. sd: 0.094"
```




```
##   numberofdraws     means         sd
## 1            10 0.6478641 0.06646594
## 2            50 0.6712053 0.09923920
## 3           100 0.6699225 0.09252613
## 4          1000 0.6714212 0.09599516
## 5         10000 0.6666817 0.09416953
```

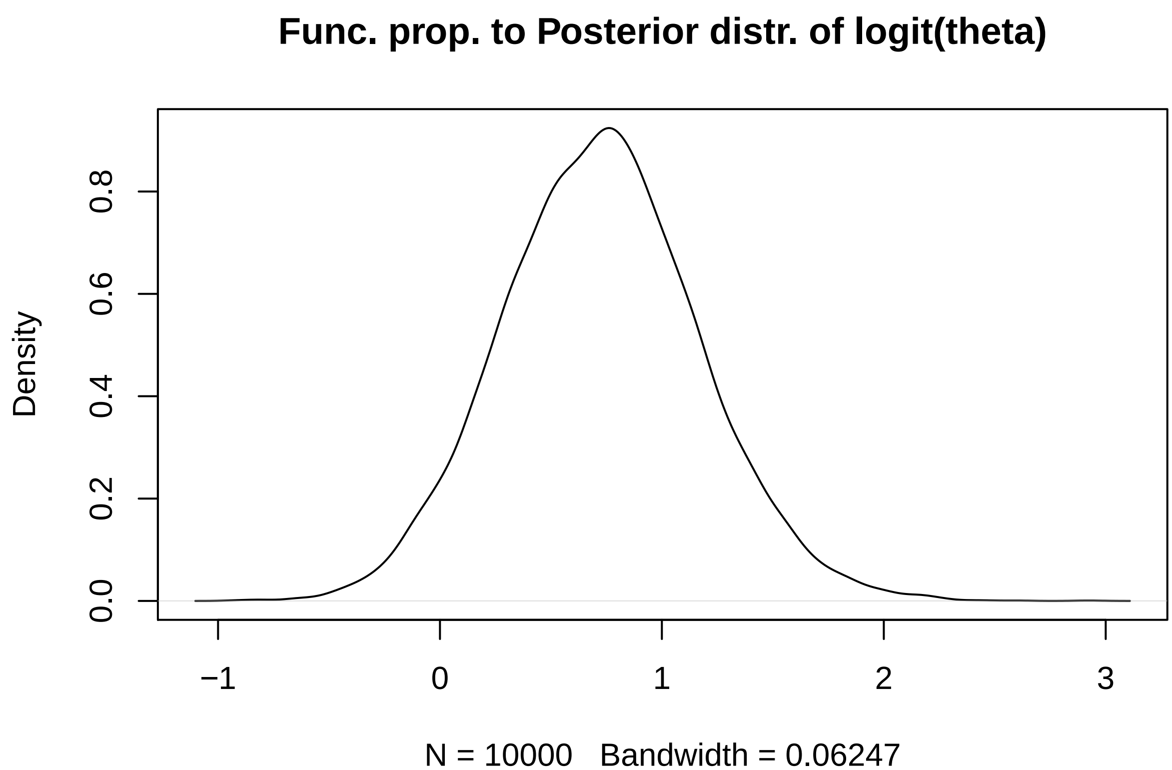
We see that the posterior draws do look like a beta density, and the sample mean and sample s.d. are almost the same as the theoretical values when the number of draws becomes large. Next, we calculate the proportion of sample draws of $\theta < 0.4$ and compare with the theoretical probability.

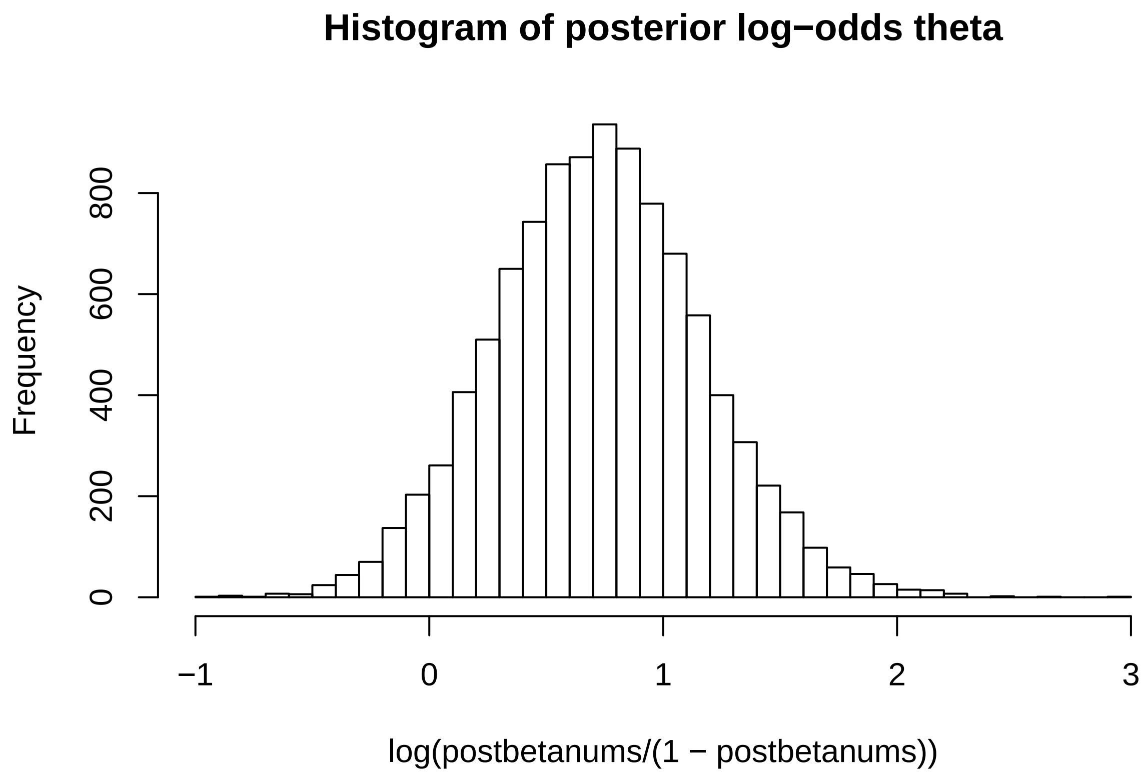
```
## [1] "Theoretical prob. posterior theta less than 0.4: 0.00397"
```

```
## [1] "Posterior beta sample prob. theta less than 0.4: 0.0042"
```

We can see that the numbers are similar, and any inaccuracy is probably due to the low probability of this outcome.

Finally, we compute a function proportinal to the posterior probability distribution for the log-odds of θ by simulation from posterior draws of θ.

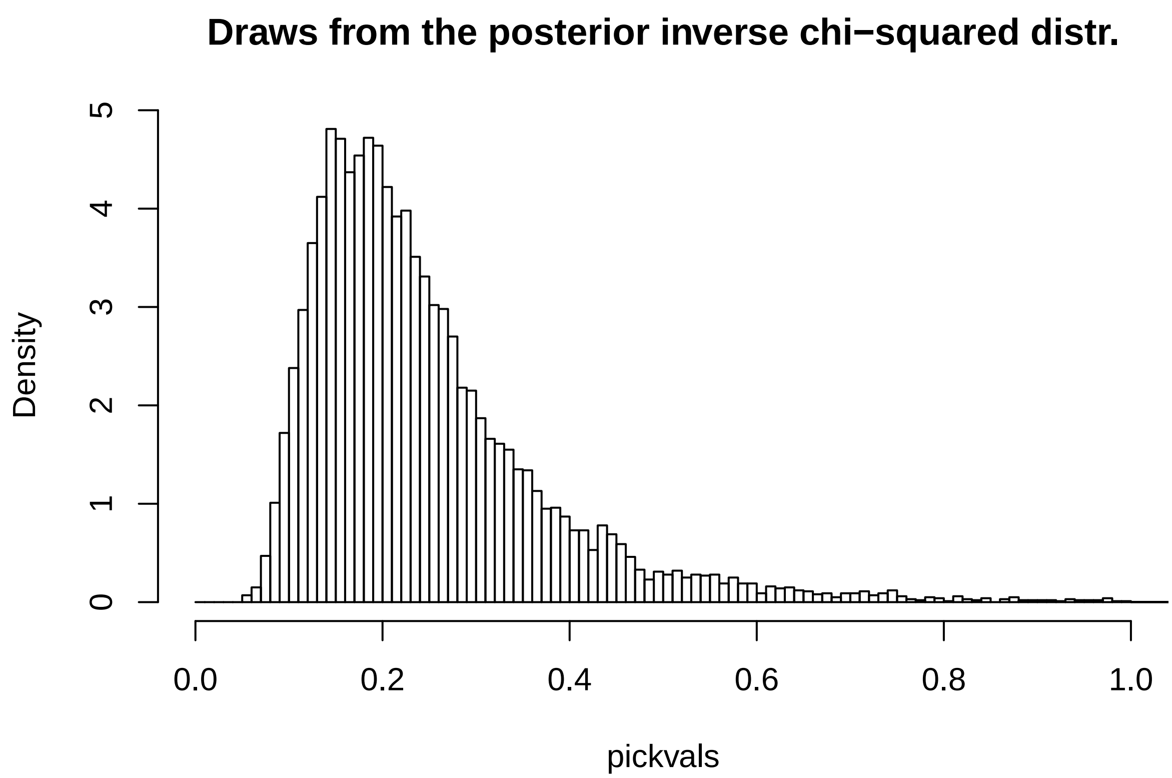


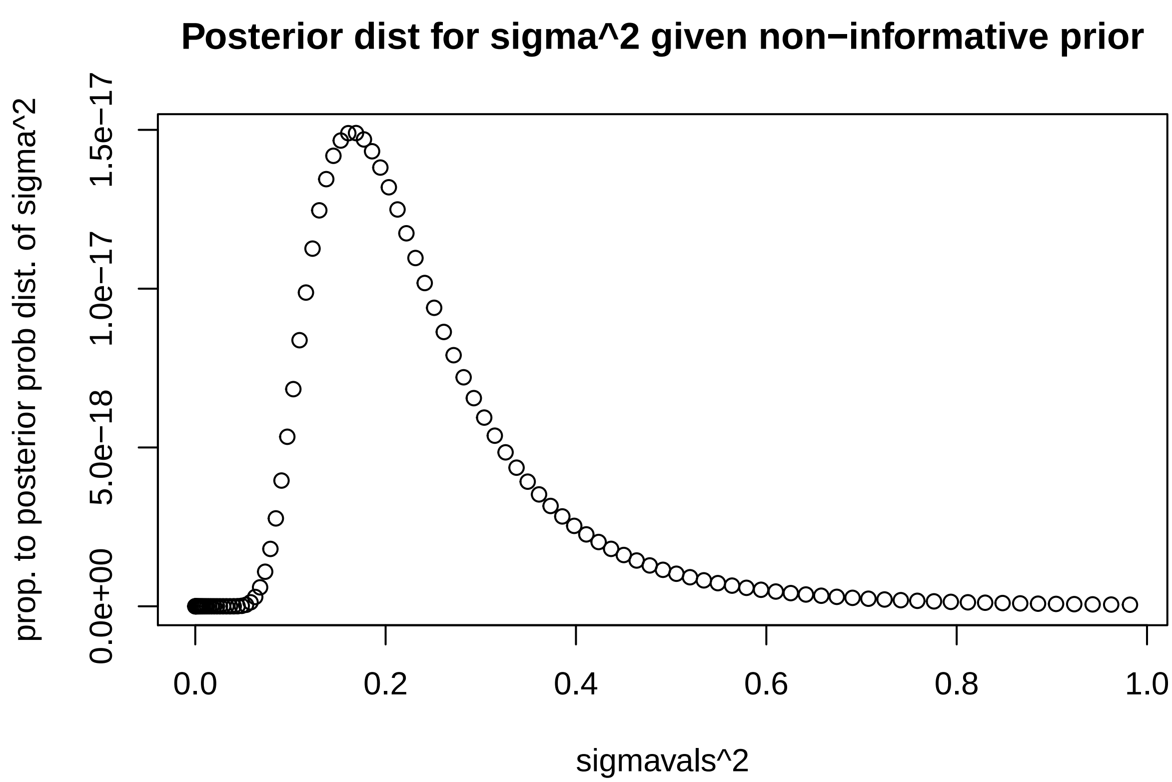


We see that the function proportional to the posterior probability density of log-odds of θ is directly made from the histogram of log-odds of θ. It looks like a smooth function. Please note that the range of log-odds, and thus support for the posterior p.d.f. of log-odds, is actually all real numbers. However, it seems that the most probable outcomes, empirically, fall in the range of the histogram shown.

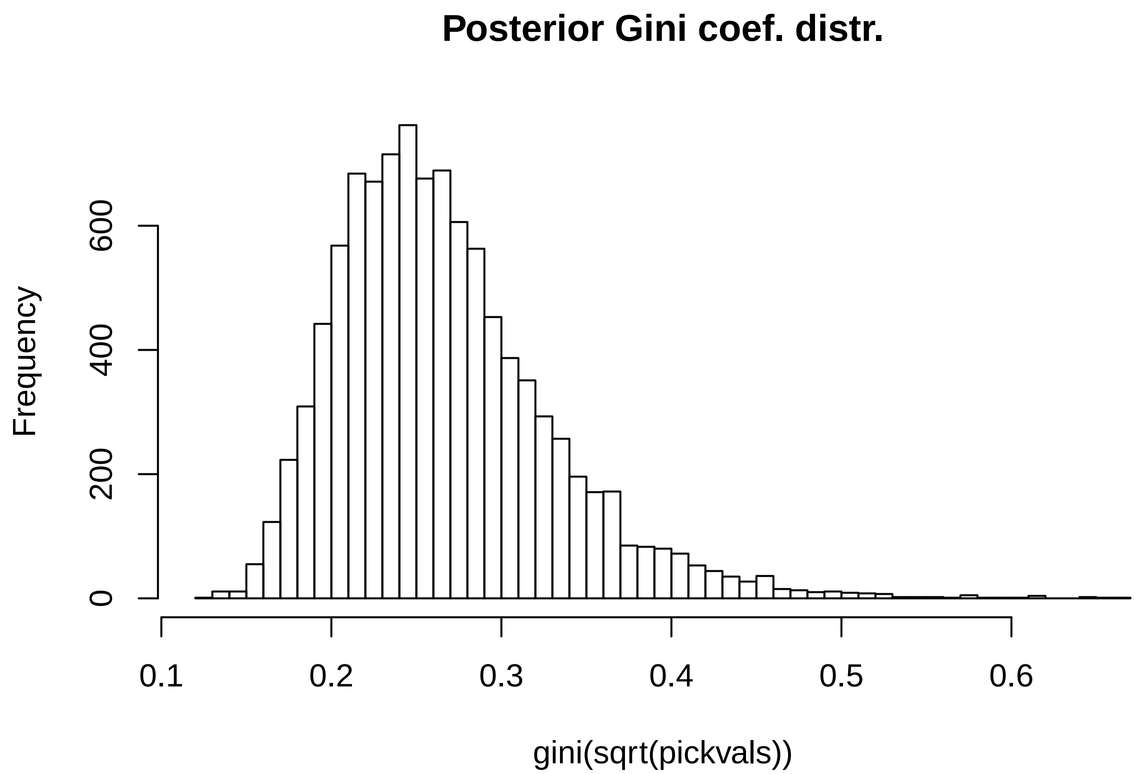
Assignment 2

We simulate 10000 draws from the $inv - \chi^2(\nu_n, s_n^2)$ distribution with the parameter values derived in 2.a. Then we compare that histogram of sample draws to a graph proportional to the posterior p.d.f. of σ^2, given non-informative prior distribution $p(\theta) = \frac{1}{\sigma^2}$.


Draws from the posterior inverse chi−squared distr.
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
pickvals

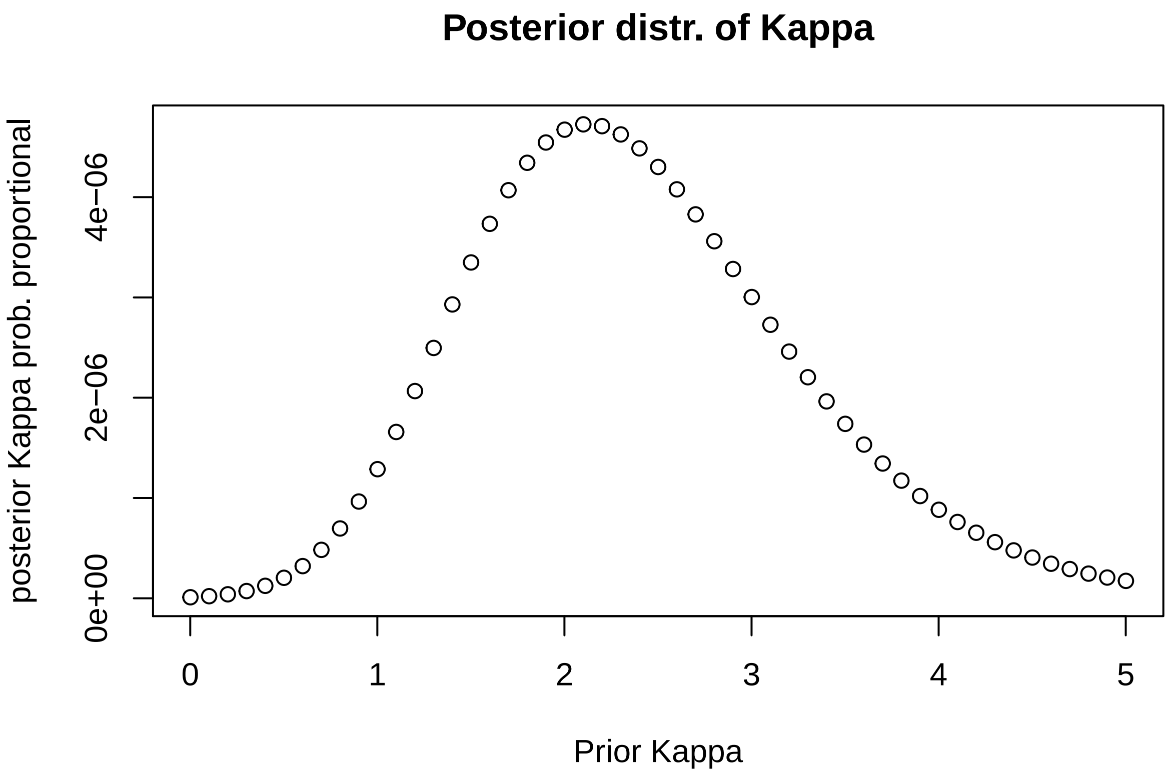

Posterior dist for sigma^2 given non−informative prior
prop. to posterior prob dist. of sigma^2
0.0e+00
5.0e−18
1.0e−17
1.5e−17
0.0
0.2
0.4
0.6
0.8
1.0
sigmavals^2

We see that the two shapes are very similar in shape, in accordance with the theory. Now, using these draws, the gini coefficient has posterior distribution proportional to the histogram below.



Assignment 3

We plot the posterior distribution of κ up to a constant of proportionality over some possible values of κ, and find the mode of posterior κ.

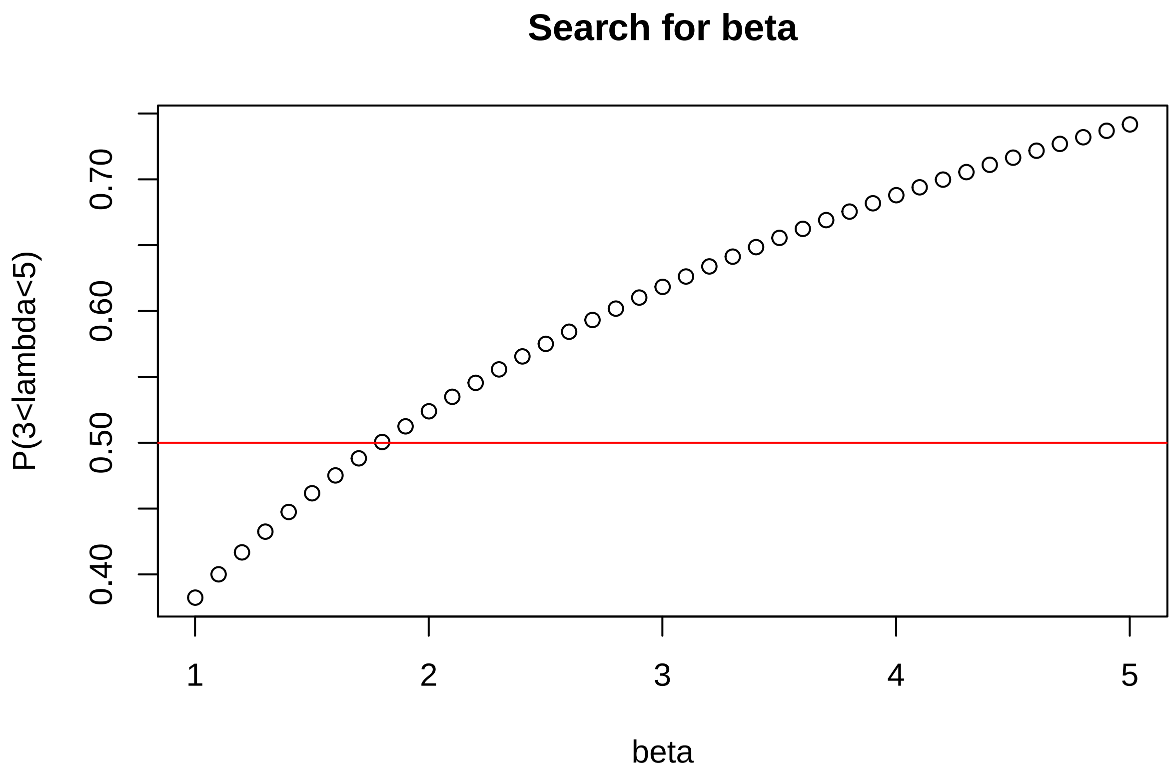



```
## [1] "Posterior mode: 2.101"
```

Since κ is a measure of how “spread out” the wind directions are, it is possible that the posterior mode of κ is dependent on data.

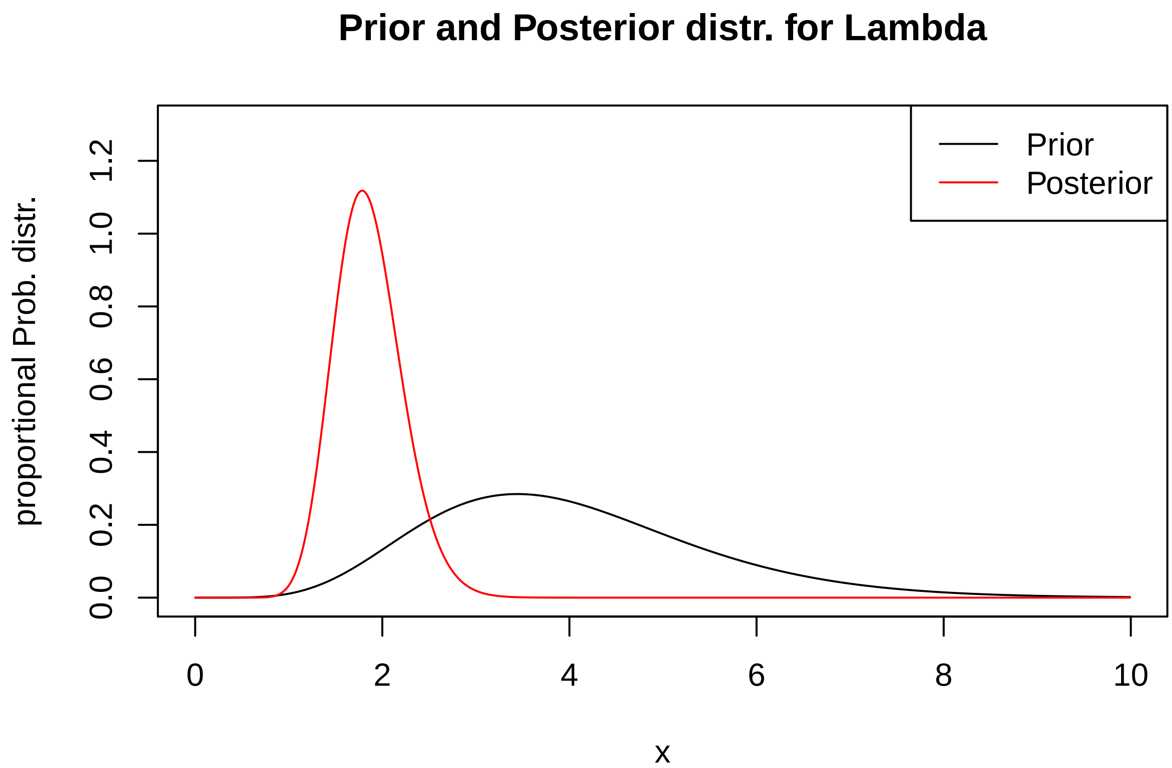
Assignment 4

we first plot a graph in order to find the optimal value of β (and thus also of α)




```
## [1] "optimal beta value: 1.8"
```

Using this optimal β value, we update the prior information using data and the formula derived in 4.a. and graph the posterior distribution (proportionally). The prior and posterior do not necessarily have the same constant of proportionality.




```
## [1] "posterior probability lambda between 3 and 5: 0.00313"
```

We see that the posterior distribution more strongly suggests a value of λ closer to two than does the prior distribution. The posterior probability for $3 < \lambda < 5$ is actually fairly small.

Appendix

R code

```
library(emdbook)

#1
s = 14
n  = 20
f = n - s
alph = 2
bet = 2

postbetanums <- rbeta(10000, alph + s, bet + f)
paste("Theoretical posterior beta distr. mean:",
      round((alph + s) / (alph + s + bet + f),3))
paste("Theoretical posterior beta distr. sd:",
      round(sqrt((alph + s) * (bet + f) /
```

```r
                   ((alph + s + bet + f)^2 * (alph + s + bet + f + 1))),3))
hist(postbetanums,breaks = 50)
paste("Posterior beta sample mean:",round(mean(postbetanums),3))
paste("Posterior beta sample sd:",round(sd(postbetanums),3))
#It checks out to 3 decimal places

postbetalessthan <- postbetanums < 0.4
ratio <- sum(postbetalessthan) / 10000
paste("Theoretical prob. posterior theta less than 0.4:",
      round(pbeta(0.4,alph + s,bet + f),5))
paste("Posterior beta sample prob. theta less than 0.4:",round(ratio,5))
# similar numbers

postlogit <- density(log(postbetanums / (1 - postbetanums)))
postlogit
plot(postlogit, main="Posterior distr. of logit(theta)")
hist(log(postbetanums / (1 - postbetanums)), breaks = 50)

#2
lognormmu <- 3.5

y <- c(14,
       25, 45, 25, 30, 33, 19, 50, 34 , 67)

gini <- function(sigma){
  return( 2*pnorm(sigma / sqrt(2)) - 1)
}
#sigma must be a number
lognormPosterior <- function(sigma,y){
  n <- length(y)

  liklihood <- prod( dlnorm(y,meanlog = lognormmu,sdlog = sigma))
  posterior <- liklihood * 1/sigma^2
  return(posterior)
}

sigmavals <- seq(0.001,1,0.01)

lognormpostvals <-c()
for( i in 1:length(sigmavals)){
  lognormpostvals[i] <- lognormPosterior(sigmavals[i],y)
}
plot(sigmavals,lognormpostvals,
     main = "Posterior dist for sigmas given non-informative prior",
     ylab = "prop. to posterior prob dist. of sigma")


#This is a draw from Inv-chisq(n-1,ssquared)
invchisqaccordingtoslide <- function(n,ssquared){
  eks <- rchisq(1, df = n - 1)
  sigmasquared <- (n - 1) * ssquared / eks
  return(sigmasquared)
}
```

```r
pickvals <-c()
for(i in 1:10000){
  pickvals[i] <- invchisqaccordingtoslide(n + 1,tausquared)
}
hist(pickvals, breaks = 50,main = "Draws from the posterior inverse chi-squared distr.")
hist(gini(sqrt(pickvals)),breaks = 50, main = "Posterior Gini coef. distr.")

#3
radianobs <- c(-2.44,2.14,2.54,1.83,2.02,2.33,-2.79,2.23,2.07,2.02)
#kappa must be a constant
vonMisesPosterior <- function(kappa,y){
  n <- length(y)

  liklihood <- prod( exp( kappa * cos(y - 2.39)) /
                     (2 * pi * besselI(x = kappa, nu = 0)))
  posterior <- liklihood * dexp(x = kappa,rate = 1)
  return(posterior)
}
kappavals <- seq(0.001,5.001,0.1)
#kappavals <- lseq(0.001,5,100)
vonMisespostvals <-c()
for( i in 1:length(kappavals)){
  vonMisespostvals[i] <- vonMisesPosterior(kappavals[i],radianobs)
}

plot(kappavals,vonMisespostvals,main="Posterior distr. of Kappa",
     xlab = "Prior Kappa",
     ylab = "posterior Kappa prob. proportion")
paste("Posterior mode:",signif(kappavals[which.max(vonMisespostvals)]))

#4
diseasedata <- data.frame(pop = c(120342, 235967,
           243745,
           197452,
           276935,
           157222))
diseasedata <- cbind(diseasedata,cases = c(2,5,3,5,3,1))
#searchforbeta
beta <- seq(1,5,0.1) + 0.001
#gammanums <- rgamma(10000, 4*beta, beta)

plot(beta,pgamma(5,4*beta, beta) - pgamma(3,4*beta, beta),
     main = "Search for beta",ylab="P(3<lambda<5)")
abline(h = 0.5,col ="red")
paste("optimal beta value:",beta[9] - 0.001)

x <- seq(0.001,10,0.01)
plot(x, dgamma(x,1.8 * 4,1.8), ,type ="l",main="Prior and Posterior distr. for Lambda",
     ylim = c(0,1.3),ylab="Prob. distr.")

newshape <- 1.8 * 4 + sum(diseasedata$cases)
newrate <- 1.8 + sum(diseasedata$pop) / 100000
lines(x, dgamma(x,newshape,newrate), col = "red")
```

```r
legend("topright",legend = c("Prior","Posterior"),lty = c(1,1), col =c("black","red"))
paste("posterior probability lambda between 3 and 5:",
      signif(pgamma(5,newshape,newrate) - pgamma(3,newshape,newrate),3))
```