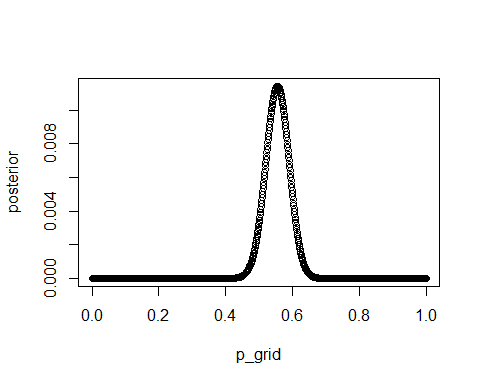
Assignment1(B04704016林家毅)

# Question 1

library(tidyverse)  
library(tidybayes)  
library(rethinking)

birth1 <- c(1,0,0,0,1,1,0,1,0,1,0,0,1,1,0,1,1,0,0,0,1,0,0,0,1,  
 0,0,0,0,1,1,1,0,1,0,1,1,1,0,1,0,1,1,0,1,0,0,1,1,0,  
 1,0,0,0,0,0,0,0,1,1,0,1,0,0,1,0,0,0,1,0,0,1,1,1,1,  
 0,1,0,1,1,1,1,1,0,0,1,0,1,1,0,1,0,1,1,1,0,1,1,1,1)  
  
birth2 <- c(0,1,0,1,0,1,1,1,0,0,1,1,1,1,1,0,0,1,1,1,0,0,1,1,1,  
 0,1,1,1,0,1,1,1,0,1,0,0,1,1,1,1,0,0,1,0,1,1,1,1,1,  
 1,1,1,1,1,1,1,1,1,1,1,0,1,1,0,1,1,0,1,1,1,0,0,0,0,  
 0,0,1,0,0,0,1,1,0,0,1,0,0,1,1,0,0,0,1,1,1,0,0,0,0)  
  
p\_grid <- seq( from = 0, to = 1, length.out = 1000 )  
prior <- rep( 1, 1000 )  
  
prob\_data <- dbinom(( sum( birth1 ) + sum( birth2 )), size = 200, prob = p\_grid )  
  
posterior <- prob\_data \* prior  
posterior <- posterior / sum( posterior )  
  
answer\_1 <- round( p\_grid[ which.max( posterior ) ], 3)  
plot( posterior ~ p\_grid )



## The parameter value which maximizes the posterior probability: p = 0.555

# Question 2

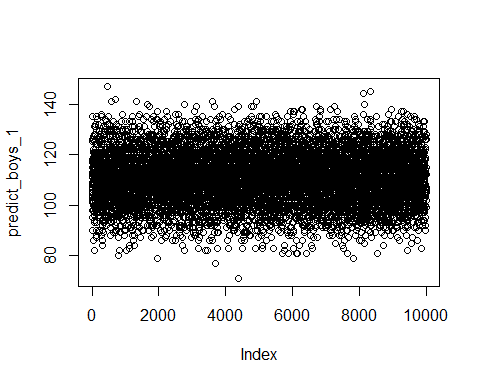
set.seed( 100 )  
samples <- sample( p\_grid, prob = posterior, size = 10000, replace = TRUE )  
answer\_2 <- mode\_hdi( samples, .width = c( 0.5, 0.89, 0.97 ))

## Posterior density intervals:

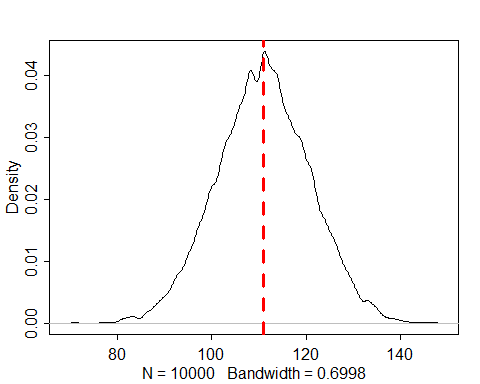
## y ymin ymax .width .point .interval  
## 1 0.5561376 0.5315315 0.5765766 0.50 mode hdi  
## 2 0.5561376 0.4974975 0.6076076 0.89 mode hdi  
## 3 0.5561376 0.4774775 0.6276276 0.97 mode hdi

# Question 3

predict\_boys\_1 <- rbinom( 1e4, size = 200, prob = samples )  
plot( predict\_boys\_1 )



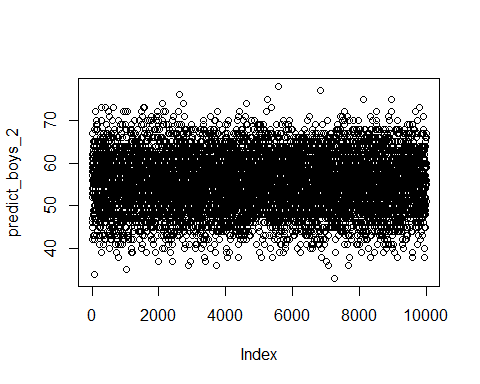
dens( predict\_boys\_1 )  
abline( v = sum( birth1 ) + sum( birth2 ) , col = "red", lwd = 3, lty = 2 )



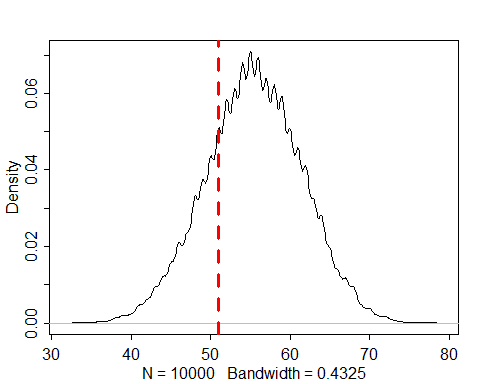
## The posterior probability model seems to fit the data well, since the distribution of predictions include the actual observation as a central.

# Question 4

predict\_boys\_2 <- rbinom( 1e4, size = 100, prob = samples )  
plot( predict\_boys\_2 )



dens( predict\_boys\_2 )  
abline( v = sum( birth1 ) , col = "red", lwd = 3, lty = 2 )



## The posterior probability model seems to fit the data badly, since the distribution of predictions do not include the actual observation as a central.

# Question 5

female\_first\_borns <- c( which( birth1 == 0 ))  
cat( "The following families have a girl as the first child:", female\_first\_borns)

## The following families have a girl as the first child: 2 3 4 7 9 11 12 15 18 19 20 22 23 24 26 27 28 29 33 35 39 41 44 46 47 50 52 53 54 55 56 57 58 61 63 64 66 67 68 70 71 76 78 84 85 87 90 92 96

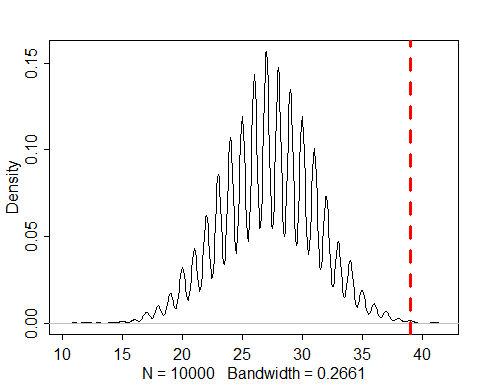
cat( "Number of families:", length(female\_first\_borns) )

## Number of families: 49

second\_birth\_following\_female <- birth2[ female\_first\_borns ]  
cat( "The second child of these families have the following genders:", second\_birth\_following\_female)

## The second child of these families have the following genders: 1 0 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 0 0 0 1 0 1

second\_male\_simulation <- rbinom(n = 1e4, size = length(second\_birth\_following\_female), prob = samples)  
dens(second\_male\_simulation)  
abline(v = sum( second\_birth\_following\_female ), col = "red", lwd = 3, lty = 2 )



## The posterior probability model do not fit the data well, one of the reasons might be that the sample size of data is too small.