Homework 2

STA 360: Assignment 2, Fall 2020

Due Friday August 28, 5 PM Standard Eastern Time

```r
library(tidyverse) #load in tidyverse package
```

Lab Component

a. Task 3

Write a function that takes as its inputs that data you simulated (or any data of the same type) and a sequence of θ values of length 1000 and produces Likelihood values based on the Binomial Likelihood. Plot your sequence and its corresponding Likelihood function.

The likelihood function is given below. Since this is a probability and is only valid over the interval from $[0, 1]$ we generate a sequence over that interval of length 1000.

```r
set.seed(123)
### Bernoulli LH Function ###
# Input: obs.data, theta
# Output: bernoulli likelihood
obs.data <- rbinom(n = 100, size = 1, prob = 0.01)
bernLH <- function(obs.data, theta){
  N <- length(obs.data)
  x <- sum(obs.data)
  LH <- (theta ^x) *((1-theta)^{N-x})
  return (LH)
}

### Plot LH for a grid of theta values ###
# Create the grid #
theta.sim <- seq(from = 0, to = 1, length.out = 1000)
# Store the LH values
sim.LH <- bernLH(obs.data, theta = theta.sim)
# Create the Plot
plot(theta.sim, sim.LH, type = "l", main = "Likelihood Profile",
     xlab = "Simulated Support",
     ylab = "Likelihood")
```

Likelihood Profile
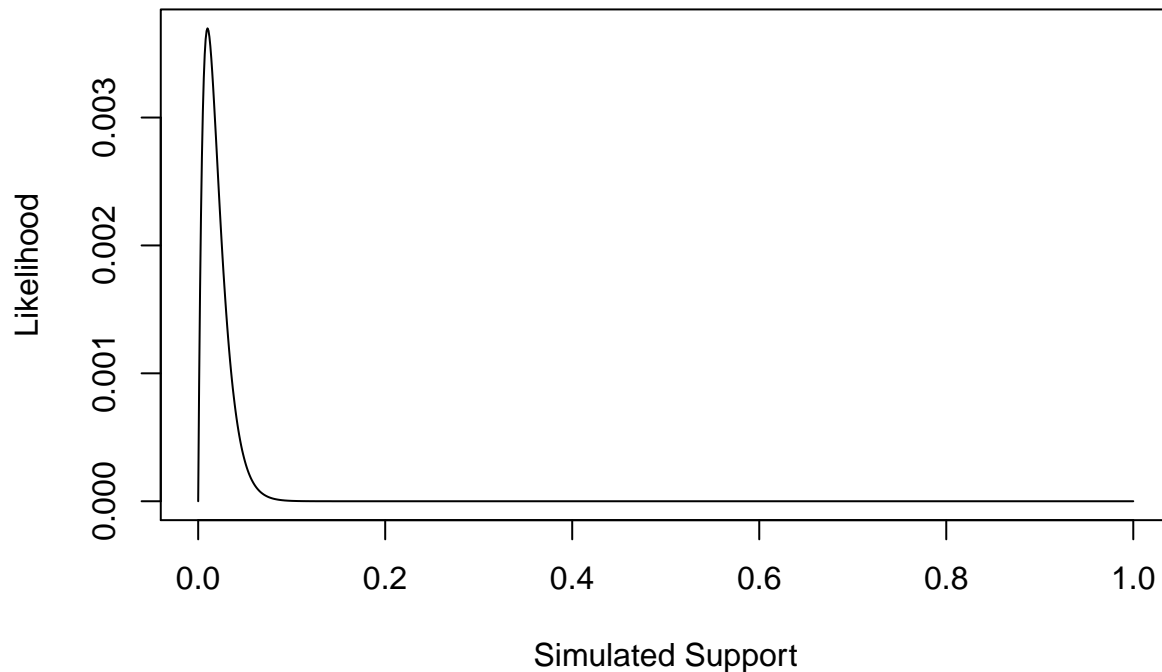
b. Task 4 Write a function with, input: prior parameters a, b, and the observed data. output: parameters of the Beta posterior distribution of theta takes as its inputs prior parameters a and b for the Beta-Bernoulli model and the observed data, and produces the posterior parameters you need for the model

```r
myBetaBernoulli <- function(obs.data, theta, a, b){
  N <- length(obs.data)
  x <- sum(obs.data)
  BeBern <- dbeta(theta, x + a, N-x+b)
  param1= x+a
  param2 = N-x+b
  print(param1)
  print(param2)
  return(BeBern)
}

non.informative.prior <- myBetaBernoulli(obs.data, theta.sim,1,1)
```

```
## [1] 2
## [1] 100
```

```r
informative.prior <- myBetaBernoulli(obs.data, theta.sim,3,1)
```

```
## [1] 4
## [1] 100
```

The parameters for the posterior with a non-informative prior are Beta(2, 100) and for the informative, the parameters are Beta(4, 100).

c. Task 5 Create two plots, one for the informative and one for the non-informative case to show the posterior distribution and superimpose the prior distributions on each along with the likelihood. What do you see? Remember to turn the y-axis ticks off since superimposing may make the scale non-sense.

```
df<- data.frame(theta.sim, non.informative.prior, informative.prior, sim.LH)
ggplot(df, aes(theta.sim)) +
  geom_line(aes(y=sim.LH, color="Likelihood")) +
  geom_line(aes(y=non.informative.prior,
                color= "Non-Informative")) +
  geom_line(aes(y=informative.prior, color="Informative"))+
  scale_color_manual(name = "Distributions",
     breaks = c("Likelihood","Non-Informative",
                "Informative"),
     values = c("red", "green","blue"))+
labs(title = "Likelihood, Posterior given Non-Informative/Informative Prior",
     y = "Density",
     x = "Theta" ) +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
  )
```

Likelihood, Posterior given Non–Informative/Informative Prior
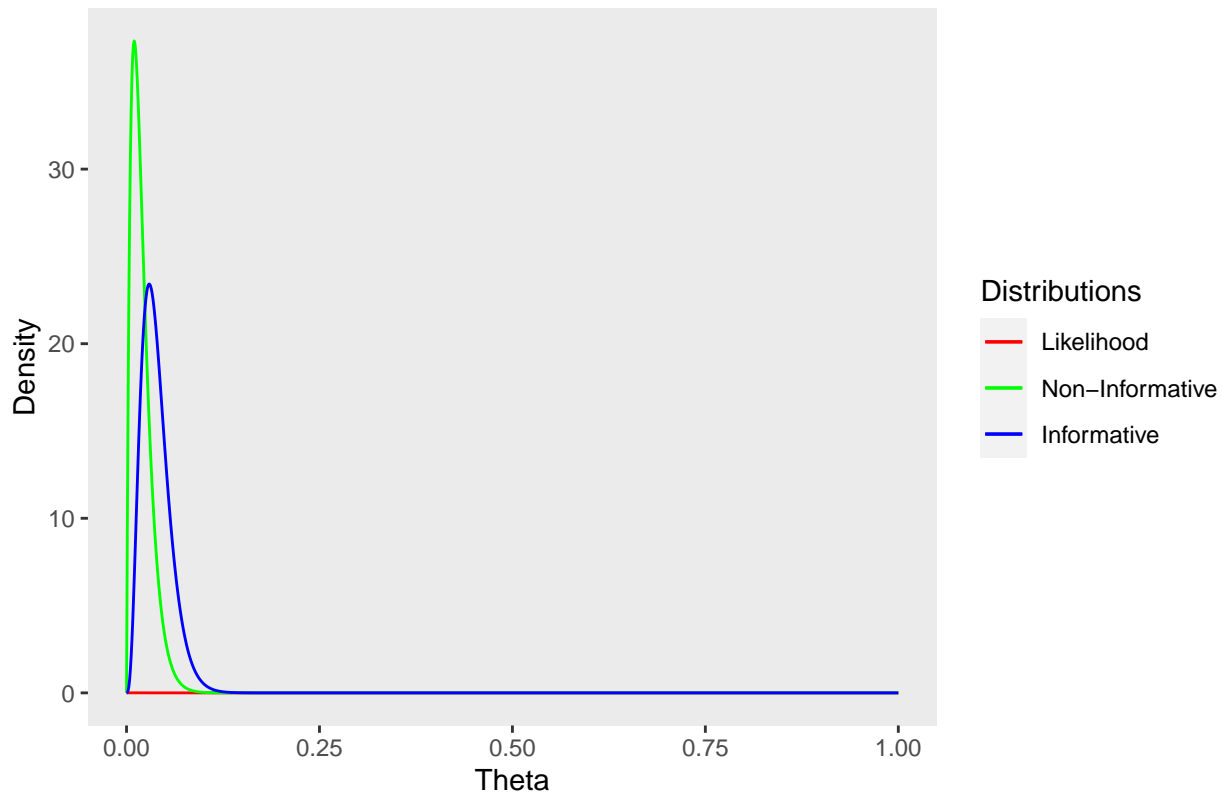
We can see that the informative prior is slightly more to the right than the non-informative prior. Having an informative prior shifted the posterior distribution so the value of theta it's centered about is greater than the non-informative prior posterior distribution. We can also see that the non-informative distribution is taller than the informative,