Code

Lab 4 | Jon Lee | BINF 6310 | Spring 2020

1. You walk into the “occasionally dishonest casino” with prior probabilities and likelihoods set to the values in slides 24-25 of lecture #4.

You pick up one die and with it roll:
2 3 2 6 3 5 6 2 6 6 2 6 6 2 3 6 6 6 5 6 6 5 6 6 6 6 6 4 6 3 3 3 6 6 5 6 6

Make a graph of the posterior probability that you have picked up a loaded die as a function of the number of times you have rolled the die.

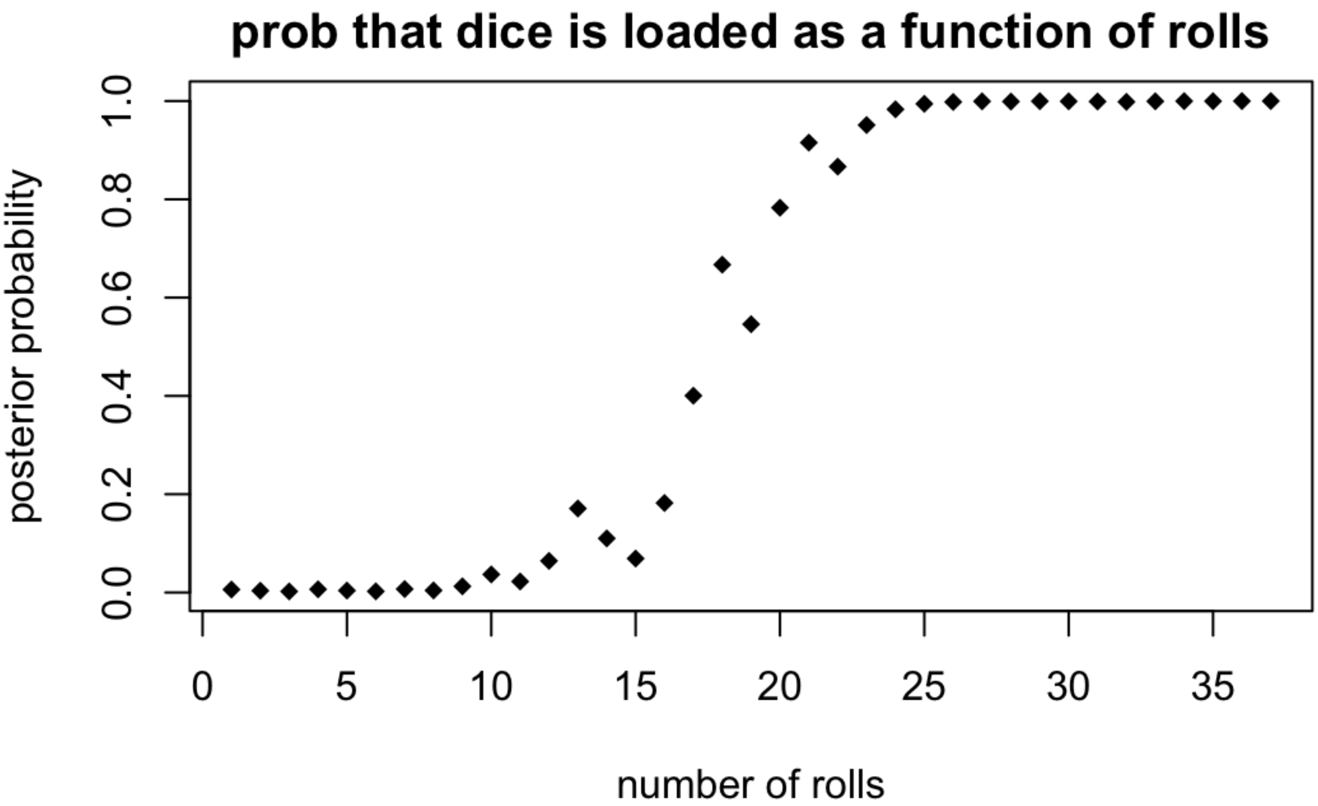
Show your code…

You can represent the rolls as data<-c(2,3,2,6,3,5,6,2,6,6,2,6,6,2,3,6,6,6,5,6,6,5,6,6,6,6,6,4,6,3,3,3,6,6,5,6,6)

Answer Code:

Hide

```
#prior probabilities (fair, loaded)
pDice <- c(0.99, 0.01)

#probability of Dice
pFairDice <- c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)
pLoadedDice <- c(1/10, 1/10, 1/10, 1/10, 1/10, 1/2)

#what you roll
data <- c(2,3,2,6,3,5,6,2,6,6,2,6,6,2,3,6,6,6,5,6,6,5,6,6,6,6,6,4,6,3,3,3,6,6,5,6,6)
numOfRolls <- length(data)

#function of posterior prob based on number of rolls
pLoaded <- vector(length = length(data))

for(n in 1:numOfRolls)
{
  denom <- pLoadedDice[data[n]]*pDice[2]+pFairDice[data[n]]*pDice[1]

  pDice[1] = pDice[1]*pFairDice[data[n]]/denom
  pDice[2] = pDice[2]*pLoadedDice[data[n]]/denom

  pLoaded[n] = pDice[2]
}

plot(1:numOfRolls, pLoaded, pch = 18, xlab = "number of rolls", ylab = "posterior probab
ility", main = "prob that dice is loaded as a function of rolls")
```



2. How many times on average would you need to roll a loaded die to be 99.99% sure that it was loaded at least 95% of the time? (Show your work)

Answer Code:

Hide

```
#sample data of rolling dice
rollDice <- function(x)
{
  rolls <- vector(length=x, mode="double")

  for(i in 1:x)
  {
    rolls[i] <- sample(1:6, 1, prob=c(0.1, 0.1, 0.1, 0.1, 0.1, 0.5))
  }

  rolls
}

#estimated power function
estimatedPower <- function(prob, likelihood_1, likelihood_2, numTests, numSimulationsPer
Cycle)
{
  t <- 1:numTests
  s <- numSimulationsPerCycle

  avgPosterior <- vector(length = length(t))
  estPower <- vector(length = length(t))

  for(i in t)
  {
    posteriorValues <- vector(length = s)

    for(j in 1:s)
    {
      p <- prob

      data <- rollDice(t[i])

      for(k in 1:length(data))
      {
        denom <- likelihood_2[data[k]]*p[2]+likelihood_1[data[k]]*p[1]

        p[1] = p[1]*likelihood_1[data[k]]/denom
        p[2] = p[2]*likelihood_2[data[k]]/denom
      }

      posteriorValues[j] = p[1]
    }

    avgPosterior[i] = mean(posteriorValues)
    estPower[i] <- sum(posteriorValues >= 1-0.0001)/s
  }

  return(estPower)
}
```

Hide

```
#reverse of pDice
pDiceRev <- c(0.01,0.99)
numberOfTests <-1:100
numberOfSimulationsPerCycle <- 10000

#diseased <- estimatedPower(probOfDisease, likelihoodGivenDisease, likelihoodGivenHealth
y, 25, 10000)
loaded <- estimatedPower(pDiceRev, pLoadedDice, pFairDice, 100, 1000)
print(loaded)
```

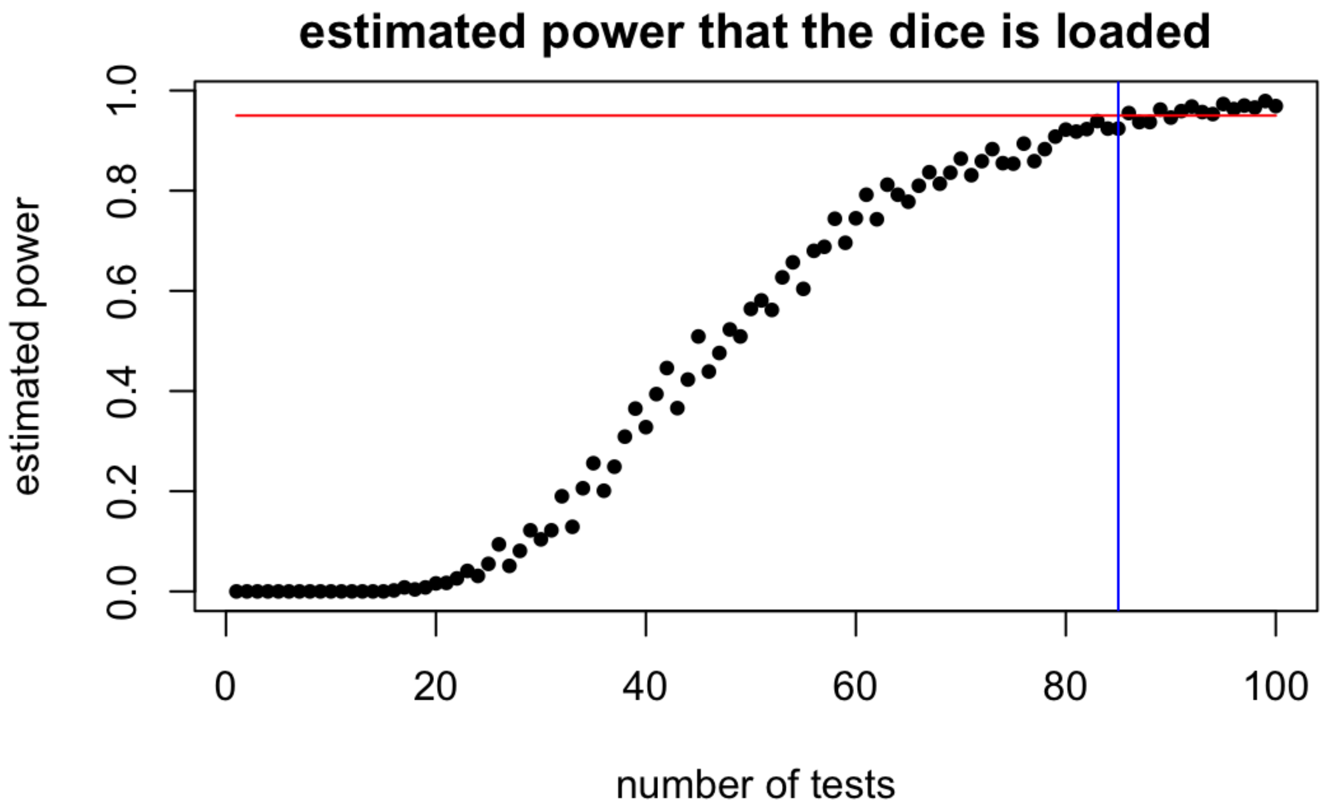
```
  [1] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
 [12] 0.000 0.000 0.000 0.000 0.002 0.008 0.004 0.008 0.016 0.017 0.026
 [23] 0.041 0.031 0.055 0.094 0.051 0.081 0.122 0.104 0.122 0.190 0.129
 [34] 0.206 0.256 0.201 0.249 0.309 0.365 0.328 0.394 0.446 0.366 0.423
 [45] 0.509 0.439 0.476 0.523 0.509 0.564 0.581 0.562 0.627 0.657 0.604
 [56] 0.680 0.688 0.744 0.696 0.745 0.792 0.743 0.812 0.792 0.778 0.810
 [67] 0.837 0.814 0.836 0.864 0.831 0.859 0.883 0.855 0.854 0.894 0.859
 [78] 0.883 0.908 0.922 0.918 0.923 0.939 0.924 0.924 0.955 0.937 0.937
 [89] 0.962 0.946 0.959 0.968 0.957 0.953 0.973 0.963 0.970 0.966 0.979
[100] 0.969
```

Hide

```
plot(numberOfTests, loaded, main = "estimated power that the dice is loaded", xlab = "nu
mber of tests", ylab = "estimated power", pch = 20)
lines(numberOfTests, rep(0.95, length(numberOfTests)), col = "red")
```

Hide

```
abline(v = 85, col = "blue")
```



Answer: In order to be 95% confident that the dice you chose was loaded, you would need to roll the dice approximately 85 times.

3. Consider two priors for our belief about p(heads) for a coin:

A uniform prior (for example dbeta(1,1)). A prior of 5 heads and tails (dbeta(6,6)).

(3A) superimpose visualizations of these two priors (using different colors for each prior) ranging from 0 to 1.

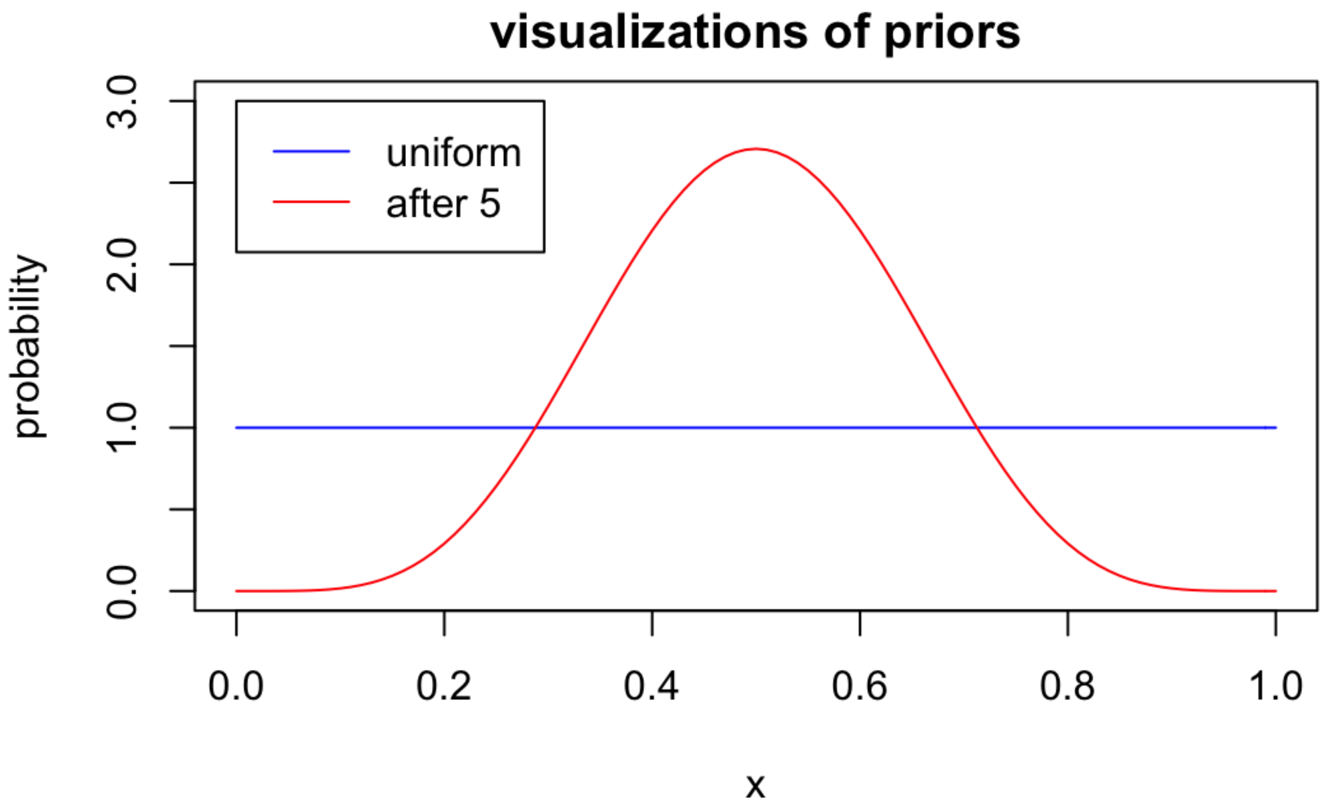
Ansewr Code:

Hide

```
#probabilities
probs <- seq (0, 1, 0.01)

#uniform prior
unif <- dbeta(probs, 1, 1)

#pior of 5 heads and tails)
after5 <- dbeta(probs, 6, 6)

#plotting
plot(probs, unif, type = "l", ylim = c(0,3), xlab = "x", ylab = "probability", main = "v
isualizations of priors", col = "blue")
lines(probs, after5, col = "red")
```

Hide

```
legend(0, 3, legend = c("uniform", "after 5"), col = c("blue", "red"), lty = c(1, 1))
```



(3B) Make posterior graphs for two experiments with new data:

One with 1 heads and 1 tail as additional observations. One with 400 heads and 400 tails as additional observations.

So you should end up with 4 posterior plots: (2 datasets * 2 priors).

Plot the two distributions involving the 2 new coin flips on one graph and the two distributions involving the 800 new coin flips on a separate graph.

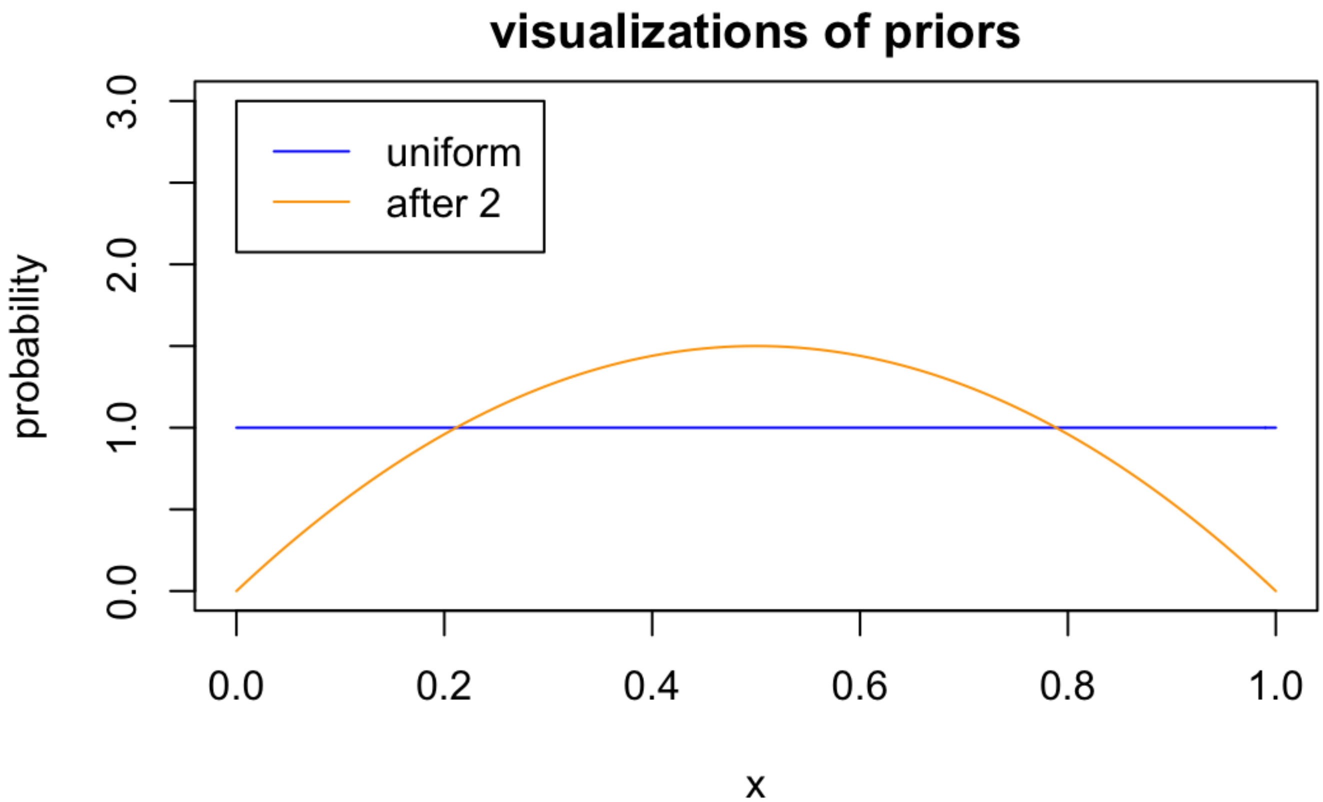
Answer Code:

Hide

```
#data of uniform prior after 1 head and 1 tail (total of 2)
after2 <- dbeta(probs, 2, 2)

plot(probs, unif, type = "l", ylim = c(0, 3), xlab = "x", ylab = "probability", main =
"visualizations of priors", col = "blue")
lines(probs, after2, col = "orange")
```

Hide

```
legend(0, 3, legend = c("uniform", "after 2"), col = c("blue", "orange"), lty = c(1, 1))
```

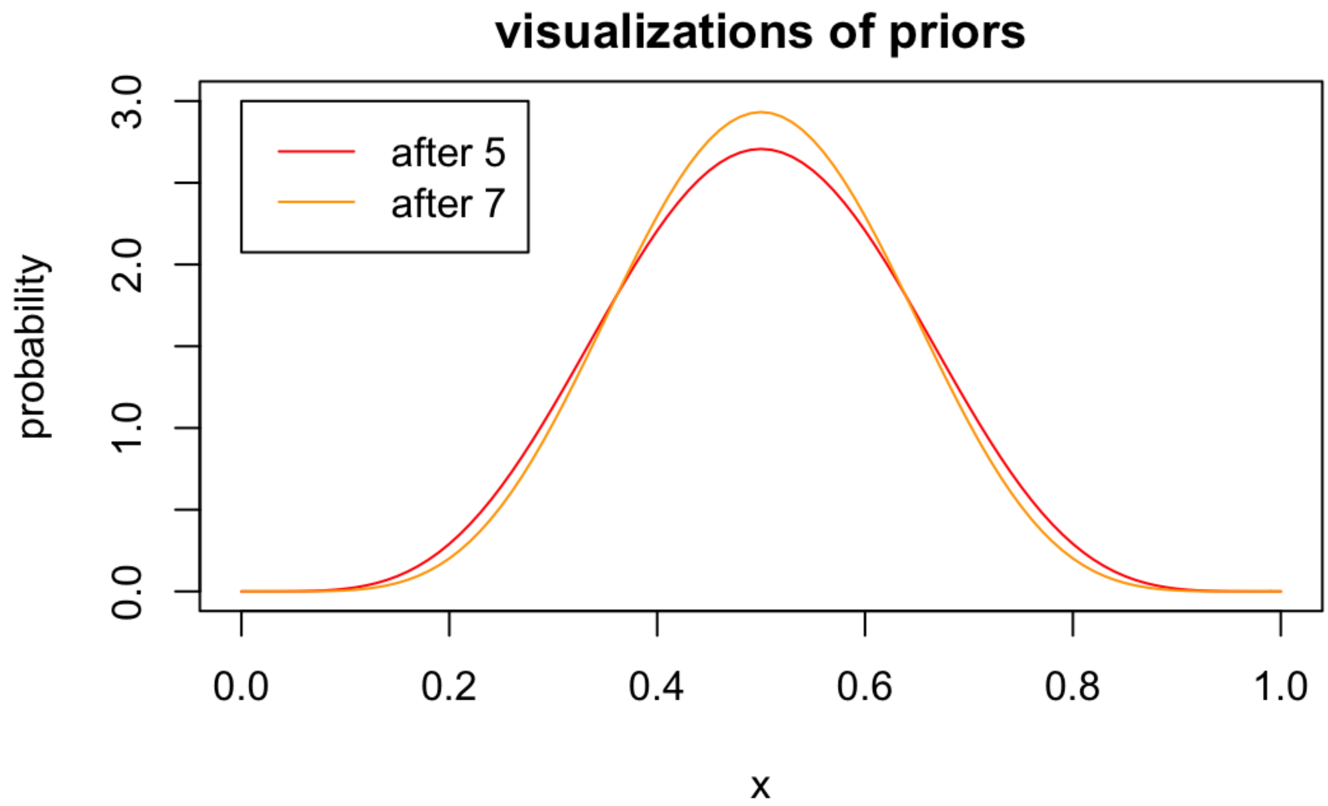


Hide

```
#data of after5 prior and after 1 heads and 1 tails (total of 7)
after7 <- dbeta(probs, 7, 7)

plot(probs, after5, type = "l", ylim = c(0, 3), xlab = "x", ylab = "probability", main =
"visualizations of priors", col = "red")
lines(probs, after7, col = "orange")
```

Hide

```
legend(0, 3, legend = c("after 5", "after 7"), col = c("red", "orange"), lty = c(1, 1))
```

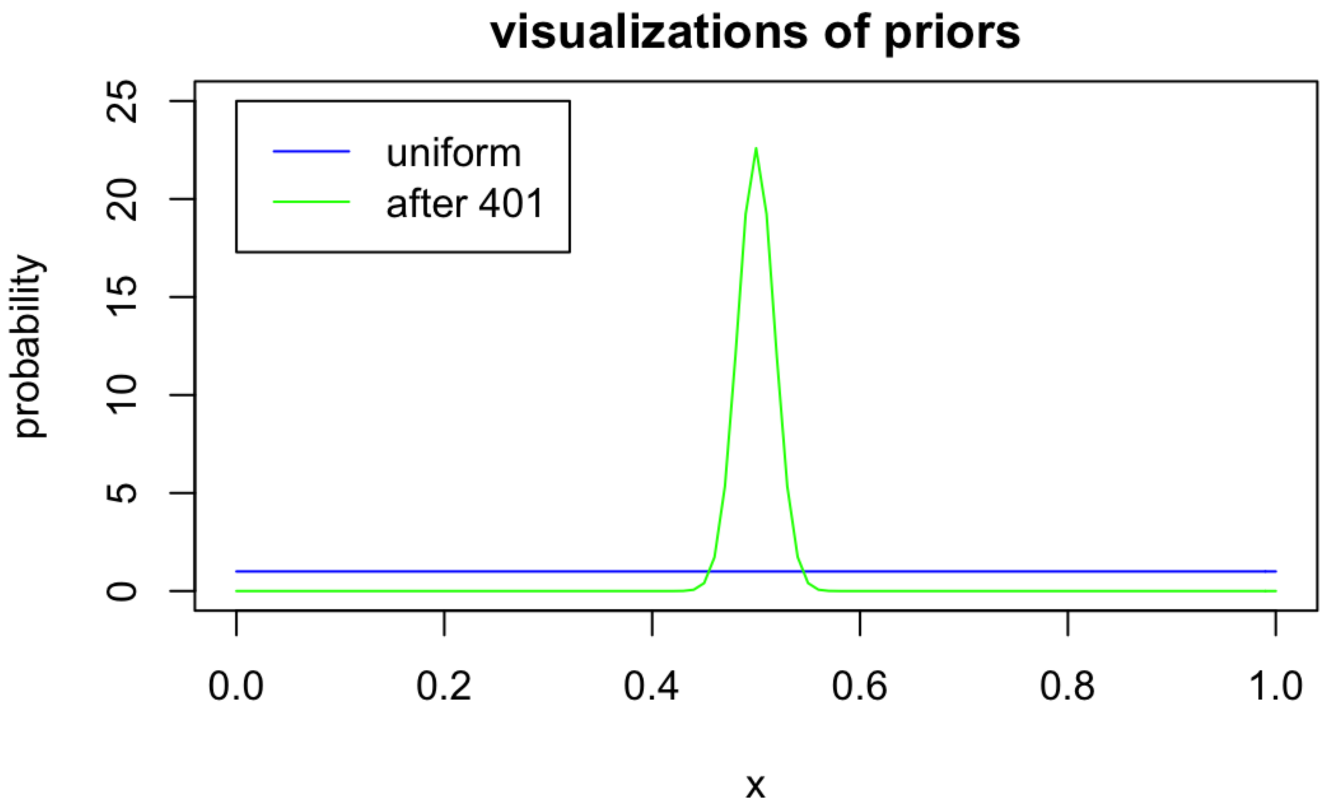


Hide

```
#data of uniform prior and after 400 heads and 400 tails
after401 <- dbeta(probs, 401, 401)

plot(probs, unif, type = "l", ylim = c(0, 25), xlab = "x", ylab = "probability", main =
"visualizations of priors", col = "blue")
lines(probs, after401, col = "green")
```

Hide

```
legend(0, 25, legend = c("uniform", "after 401"), col = c("blue", "green"), lty = c(1, 1
))
```

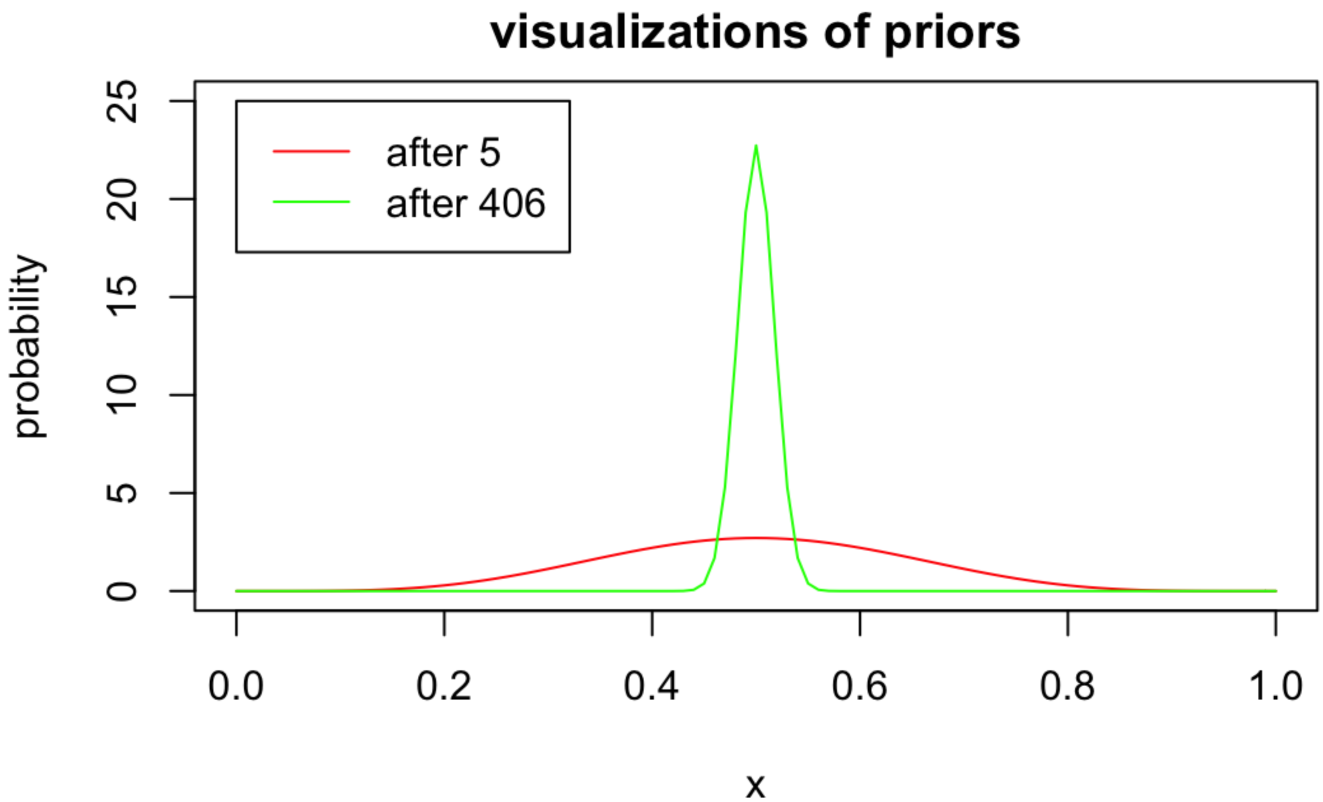


Hide

```
#data of after5 and after 400 heads and 400 tails
after406 <- dbeta(probs, 406, 406)

plot(probs, after5, type = "l", ylim = c(0, 25), xlab = "x", ylab = "probability", main
 = "visualizations of priors", col = "red")
lines(probs, after406, col = "green")
```

Hide

```
legend(0, 25, legend = c("after 5", "after 406"), col = c("red", "green"), lty = c(1, 1
))
```



Why are the two posterior plots involving the 800 coin flips so similar? Why are the two posterior plots involving the 2 coin flips so different?

Answer (to both above):

The addition of two coin flips after a prior uniform distribution has a greater impact on the change in our posterior (because initially we assum that the all probailities are equally probably, aka uniform), as oppsed to the addition after a prior of having seen 5 heads and 5 tails. In the later we already have an idea of the distribution where as in the uniform prior we have no knowledge whatsoever. Therefore, steps (of addition of coin flips) is more impactful to our predicted distribution.

On the opposite side of the spectrum, we are adding a total of 800 coin flips which takes an extremely large step in out data (as opposed to the small steps at the beginning) so we will see a small difference between our two distribution of probailities (aka they look more similar).