**(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)

See figure 1. Associated R script is located at [this link](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab04/roll_die_with_emits.R). The general function (**run\_bayes\_sim.R**) that is being used by all other functions is located at [this link](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab04/run_bayes_sim.R).



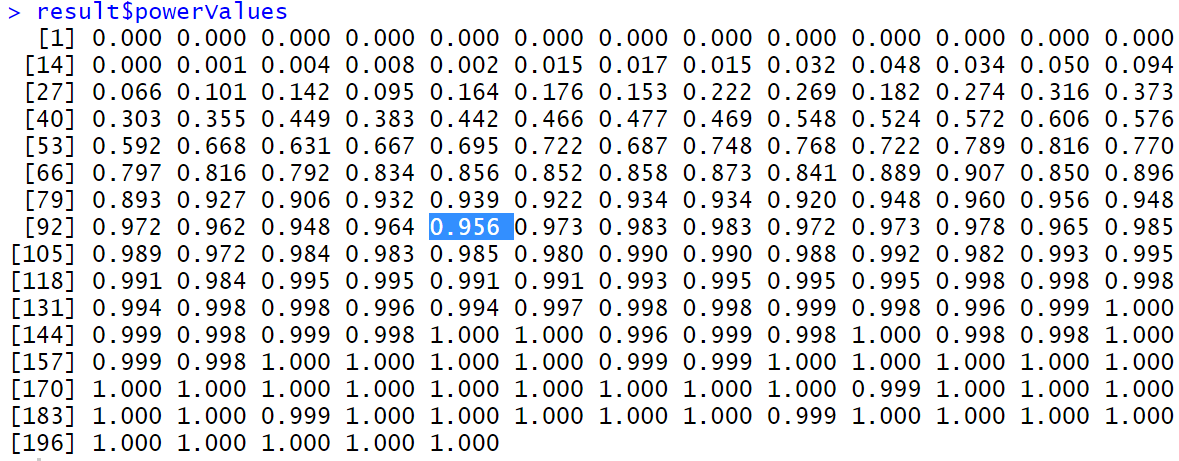
**Figure 1:** Plot of the posterior probability of theses die rolls: 2 3 2 6 3 5 6 2 6 6 2 6 6 2 3 6 6 6 5 6 6 5 6 6 6 6 6 4 6 3 3 3 6 6 5 6 6

**(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)**

It takes about 96 rolls for a loaded die to be 99.99% sure that it was loaded at least 95% of the time (see figures 2 & 3). Associated R script is located at [this link](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab04/run_loaded_die_1.R).



**Figure 2:** Power Value per number of rolls



**Figure 3:** Data for figure 2, showing it takes about 96 rolls

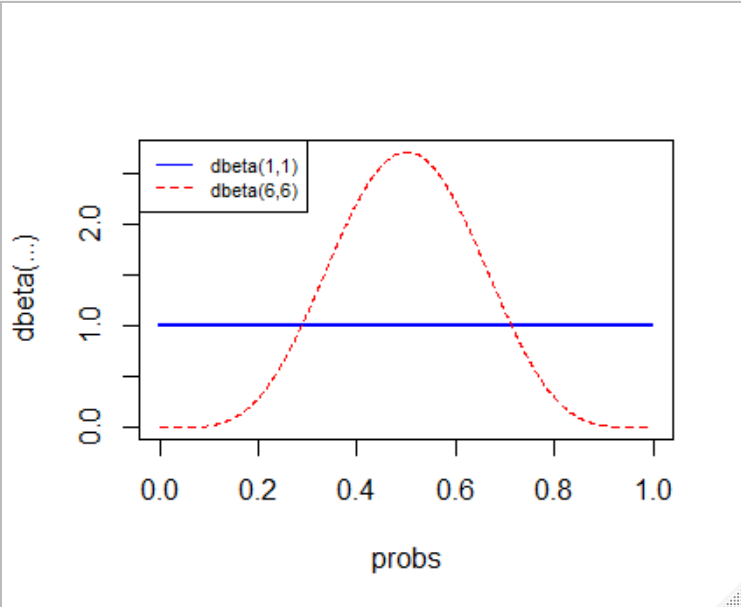
**(3) Consider two priors for our belief about p(heads) for a coin:**

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.**

**See figure 3.**



**Figure 3:** R command - plot2beta(c(1,1), c(6,6))

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

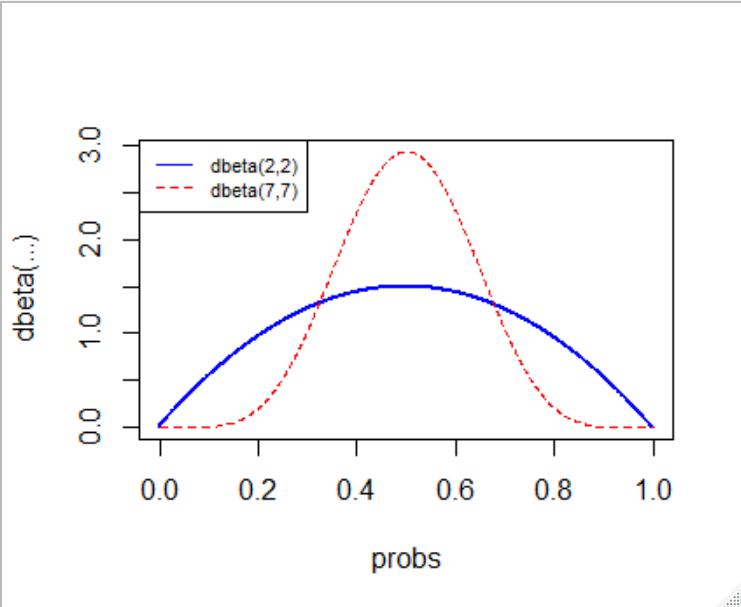
One with 1 heads and 1 tail as additional observations.

One with 400 heads and 400 tails as additional observations.

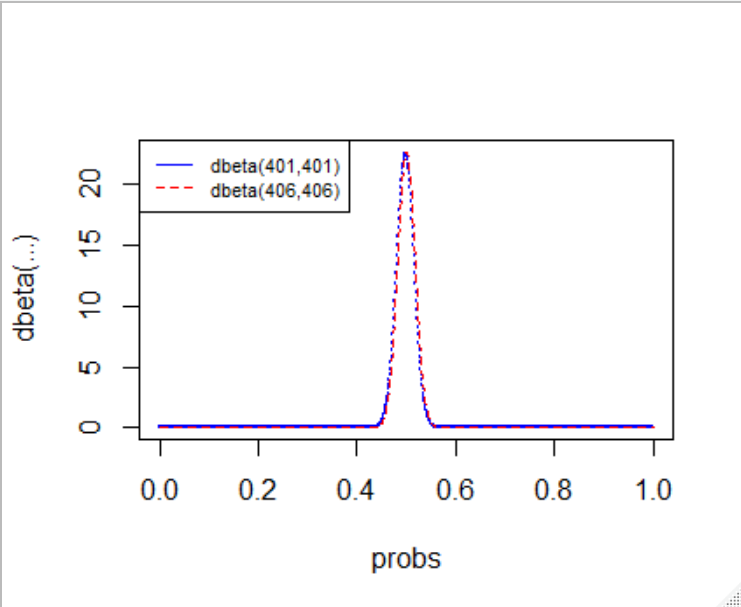
So you should end up with 4 posterior plots: (2 datasets \* 2 priors).

**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.**

See figures 4 & 5.



**Figure 4:** Posterior plots for two additional coin flips (1 head & 1 tail). R command - plot2beta(c(1+1,1+1), c(6+1,6+1)) and [function source file](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab04/plot2beta.R).



**Figure 5** Posterior plots for 800 additional coin flips (400 heads & 400 tails). R command - plot2beta(c(1+400,1+400), c(6+400,6+400)) and [function source file](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab04/plot2beta.R).

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

The two posterior plots (figure 3) involving 800 coin flips are very similar, because the variance between dbeta(401,401) and dbeta(406,406) only differ by 1.2%. Therefore, the data would be dispersed similarly about the expected value. I calculated variances using equation 1 below. I created [this R function](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab04/dbetavar.R) to calculate the variance for each case.

\*\*\*\*\*\*\*\*\*\*\*R function call to calculate variance difference\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

var\_difference\_401\_406 <- (dbetavar(406,406) - dbetavar(401,401))/dbetavar(401,401); var\_difference\_401\_406

# [1] -0.01230012

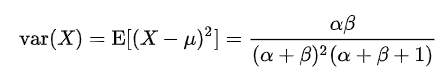
**Why are the two posterior plots involving the 2 coin flips so different?**

The two posterior plots (figure 2) involving two coin flips are so different, because the variance between dbeta(2,2) and dbeta(7,7) differ significantly by 66.7%. Therefore, I the dispersion of data between these two cases vary significantly about the expected value. In addition, figure 6 shows how the variance of the beta distribution, using equation 1, rapidly decreases as alpha = beta increases.

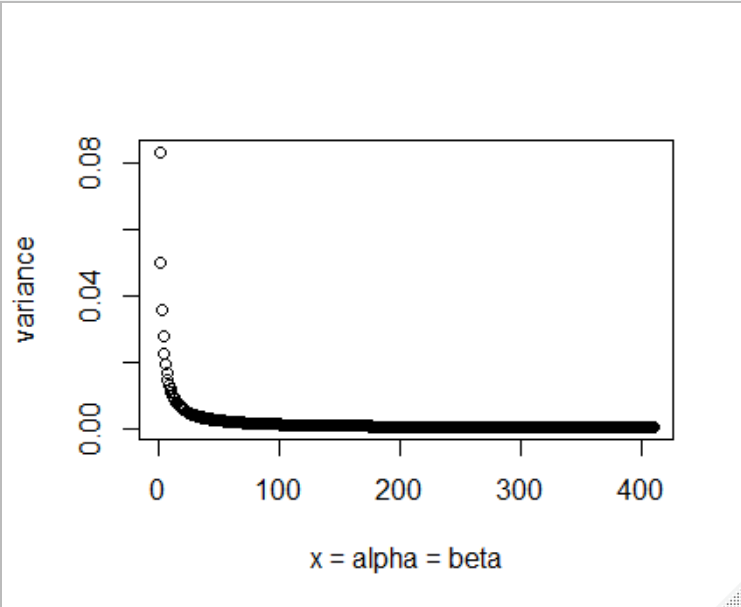
\*\*\*\*\*\*\*\*\*\*\*R function call to calculate variance difference\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

var\_2\_7 <- (dbetavar(7,7) - dbetavar(2,2))/dbetavar(2,2); var\_2\_7

# [1] -0.6666667



**Equation 1:** Variance for beta distribution (Wikipedia: <https://en.wikipedia.org/wiki/Beta_distribution#Variance>)



**Figure 6:** Plot of variances using equation 1, where alpha = beta.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*R code to plot figure 6\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

x <- seq(1, 410, 1)

> plot(x, dbetavar(x,x))

> x <- seq(1, 410, 1)

> plot(x, dbetavar(x,x), xlab = "x = alpha = beta", ylab="variance")