COMPUTATIONAL STATISTICS & PROBABILITY

FALL 2021

PROBLEM SET 1

DUE DATE: 23:59:59 **November 3**, 2021
FORMATTING: Acceptable formats are `.pdf`, `.R`, or `.Rmd` only.
SUBMITTING: Upload your completed assignment to Canvas.
Submissions must comply with the homework policy on the course page.

---

1. Use the following R code to generate a specific set of samples to get an exact answer to the questions in this section.

```
p_grid <- seq(  from=0 , to=1 , length.out=1000 )
prior <- rep( 1 , 1000 )
likelihood <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- likelihood * prior
posterior <- posterior / sum(posterior)
set.seed(215)
samples <- sample( p_grid , prob=posterior , size=1e4 ,
 replace=TRUE )
```

   a) How much posterior probability lies below $p = 0.2$?
   b) How much posterior probability lies above $p = 0.8$?
   c) How much posterior probability lies between $p = 0.2$ and $p = 0.8$?
   d) 20% of the posterior probability lies below which value of p?

2. Suppose the globe tossing experiment yielded the following sequence of 15 observations:

$$[W, L, W, W, L, L, W, L, W, L, L, W, L, W, W]$$

   a) Using grid approximation, construct the posterior distribution with the same flat prior as before.
   b) Using grid approximation, construct the posterior distribution with a prior that is 0 below $p = 0.5$ and constant above $p = 0.5$.

· What is the difference between these two models?

· How does each posterior distribution compare to the true value of $p = 0.7$?

· Which prior is better and why?

3. Suppose you want a very precise estimate of the proportion of the Earth's surface covered in water. Specifically, you want the 99% percentile interval of the posterior distribution of p to be only 0.05 wide, that is, the distance between the lower and upper bound on p should be 0.05. How many times will you have to toss the globe to do this? A precise answer is unnecessary. I am primarily interested in your approach.