Second homework assignment, 9/24/2020 (due 10/1/2020).

Part One: Applying probability laws, conditional probability

Suppose you plan to roll a fair die twice. Recall that the probability of any particular simple event (e.g., roll=1, roll=4) is 1/6. Use your knowledge of the probability laws to answer the following questions. For the more complicated questions, the trick is to figure out how to rephrase the question so that it fits one of the probability rules.

- What is the probability of obtaining a 2 on the first roll and a 4 on the second roll?
- What is the probability of obtaining a 2 and a 4 (in any order)?
- What is the probability of obtaining at least one 2 or at least one five?
- What is the probability of obtaining a sum of 8?

Now, for each of those questions, use simulation in R to validate the results by simulating die rolls and analyzing them appropriately. The R transcript from 9/18/2018 will help you see how to do this.

Part Two: Bayes' theorem

Your doctor has just called you and reported that he has some good news and some bad news. The bad news is that you have tested positive for an exotic disease that will make your head explode. The good news is that the disease is relatively rare, occurring in the general population with a probability of only .001 (i.e., only one in a thousand persons will get the disease).

You do some further investigation, and find that, according to the company marketing the test, it's a pretty accurate test. Specifically, the sensitivity of the test is 95 percent (that is, if you have the disease, the probability is .95 that you will test positive). Furthermore, the false positive rate is only five percent (i.e., P(test positive | don't have disease) = .05). The false negative rate is also just five percent (i.e., P(test negative | have disease) = .05).

How worried are you? Your answer should involve Bayes' theorem.

Part Three: Using R to understand probability distributions

Exponential random variables are sometimes used to model time until an event occurs or time between events. The exponential distribution is governed by a single *rate* parameter, which must be greater than zero. So we might speak of an "exponential(2.5)" distribution, meaning an exponential distribution with rate 2.5.

In R, it is easy to sample from the exponential distribution. The syntax is: rexp(size of sample, rate parameter). For example, to create a variable called "x" that contains half a million draws from the exponential distribution with rate parameter 1.0, we would enter

```
x <- rexp(500000, 1.0)
```

We could then do things like plot a histogram of x to see the shape of the distribution, find the mean and variance of x, and so on.

Your first task is to set up a table to help you think about the relationships between a number and common functions of that number. For each z value, find the squared value and reciprocal value (1/z) and include it in the table below:

z	z^2	$1/z$
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Your second task in Part Three is to use simulations in R to determine the relation between the rate parameter and the mean and variance of the distribution. Do this by simulating several large samples of exponential random variables with different rate parameters, calculating the mean and variance of each sample, and looking for the pattern. (Note: "var(x)" will give the sample variance of x.)

Your third task is to describe the shape of the exponential distribution. Specifically, how does the shape of the distribution depend on the rate parameter? You already know something about the mean and variance now, and you should use that information in your description. But you should also talk about other aspects of shape, particularly symmetry. You can use histograms and Pearson's index of skew to discuss symmetry. Recall that in R, you can calculate Pearson's index for a variable called "x" by typing

```
3*(mean(x)-median(x))/sd(x)
```

or by using the "pskew()" function that we wrote.

Does the symmetry of the distribution appear to change when you change the rate parameter?