Assignment #4 contains problems about the binomial, Poisson, and normal distributions. Problems which require the use of R have the symbol ⟨R⟩.

Please include **your name** and **the discussion section (day/time) that you attend** on your homework.

This assignment is worth 50 points in total, and points for each problem are indicated.

If you feel challenged by these problems, I encourage you to do additional problems on your own. Many problems have answers in the back of the textbook.

Your assignment must be turned in during lecture or to your TA's mailbox by 5pm on the due date. We will not grade late homework. If there are special circumstances, please speak to Professor Larget, preferably in advance, for consideration.

The first several problems require you to make graphs of binomial distributions using R and to then answer questions. To graph the distributions you will need to load in some R functions as on the last assignment. These functions are in the file `prob.R` that is on the course web page.

If you are using a computer connected to the internet, this command should be able to load the commands.

```
> source("http://www.stat.wisc.edu/courses/st371-larget/prob.R")
```

Otherwise, you can download the file onto your computer by right-clicking (for a Windows computer) on the link to `prob.R` from the schedule on the course homepage. You source this data into R from an option from the File menu (Source File... on a Mac, Source R code... on a Windows machine).

The function `gbinom` will graph the binomial distribution. Here are several examples. Lines with # are comments.

```
# Graph the binomial(500,0.4) distribution over the entire range.
> gbinom(500,0.4)
# Graph the binomial(500,0.4) distribution automatically scaled.
> gbinom(500,0.4,scale=T)
# Graph the binomial(500,0.4) distribution over a specified range.
> gbinom(500,0.4,low=100,high=250)
# Graph a distribution and compute P(190 <= Y <= 210).
> gbinom(500,0.4,scale=T,a=190,b=210)
# Find the 90th percentile.
> gbinom(500,0.4,scale=T,quantile=0.90)
```

1. ⟨R⟩ *[10 points]* Use R to plot the five binomial distributions with $n = 9$ fixed and for $p = 0.1$, 0.3, 0.5, 0.7, and 0.9. Answer these questions by examining graphs, not by using formulas. Please *do not* include graphs of the distributions with your submitted work.

   (a) How does the center of the distribution change as $p$ increases? (Does it stay the same, increase, or decrease? If it changes, how?)

   (b) For which values of $p$ is the distribution most spread out and for which is it most concentrated?

   (c) For which values of $p$ is the distribution skewed to the right and for which is it skewed to the left?

   (d) For which values of $p$ is the skewness the strongest?

2. ⟨R⟩ *[10 points]* Use R to plot the six binomial distributions with $n = 10, 20, 40, 80, 160$ and for $p = 0.1$ fixed. Please *do not* include graphs of the distributions with your submitted work.

   (a) How does the center of the distribution change as $n$ increases? (Does it stay the same, increase, or decrease? If it changes, how?)
   (b) How does the variance change as $n$ increases?
   (c) How does the skewness change as $n$ increases?
   (d) How doe the shape of the distribution change as $n$ increases?

3. ⟨R⟩ *[5 points]* In the U.S. population, 42% of people have blood type A. In a sample of 400 individuals, what is the probability that 150 or fewer people are blood type A? You could use the function `gbinom` to answer this question. Here are two alternative commands.

```
> pbinom(150,400,0.42)
> sum(dbinom(0:150,400,0.42))
```

   The `dbinom` function computes individual binomial probabilities. The `pbinom` function computes the cumulative distribution function, the sum of probabilities from 0 to the first argument here.

4. *[5 points]* A model for a single site in a DNA sequence is that the number of nucleotide substitutions at the site in an evolutionary lineage follows a Poisson process with a rate of 0.08 substitutions per million years. Let $X$ be the random number of substitutions in a ten million year time period.

   (a) What is $E(X)$, the expected value of $X$?
   (b) What $\Pr\{X = 0\}$, the probability of no substitutions at this site?
   (c) What is $\Pr\{X \geq 2\}$, the probability of two or more substitutions at the site?

5. *[5 points]* Do Exercise 4.1.

6. *[5 points]* Do Exercise 4.5.

7. *[5 points]* Do Exercise 4.8.

8. *[5 points]* Do Exercise 4.13 – 4.15.