Lab 7 – Solution

1. Go to the following website: `http://bcs.whfreeman.com/ips5e/default.asp` → `Statistical Applets` → `Normal Approximation to Binomial`
   Choose different values of n and p and observe what happens to the binomial probability histogram. Then answer the following questions:

   (a) For $n = 10$, what happens to the histogram as p moves away from 0.5 (to the left or to the right).

   When p get closer to 1, the histogram becomes left skewed. When p gets closed to 0, the histogram becomes right skewed.

   (b) When $n = 100$, do you observe the same behavior of the histogram as p moves away from 0.5?

   When $n = 100$, the histogram looks like the normal curve (approximately) not matter what hte value of p is.

2. Go to the following website:
   `http://onlinestatbook.com/stat_sim/sampling_dist/index.html`, read the instructions and click `Begin`. Choose different distributions from the pop-up menu and for different sample sizes (5, 1000, 10000) look at the histogram of the mean in the third panel. (Check also the fit normal box). Then, answer the following questions:

   (a) What happens to the distribution of the mean ($\bar{X}$) when the sample size increases? According to the Central Limit Theorem, what is the approximate distribution of $\bar{X}$?

   As the sample size increases the distribution of the mean looks approximately like the normal curve. According to the CLT the approximate distribution of $\bar{X}$ is $N(\mu, \sigma/\sqrt{samplesize})$, where μ and σ are the mean and the standard deviation of each of the X_i's.

   (b) In this applet the examples of parent distributions are normal and uniform. Can you use the CLT for the binomial or another discrete distribution? Which are the assumptions that need to be satisfied in order to apply the CLT?

   In order to use the CLT, we need to have independent and identically distributed random variables and a sufficiently large sample size (Rule of thumb: $n > 30$).

   (c) Assume that $X_1, X_2, \ldots, X_{200}$ (i.e. $n = 200$) all follow a normal distributions with mean 1.6 and standard deviation 1.2. Use the Central Limit Theorem to find the probability that $\bar{X}$ is greater than 2.

The CLT tells us that the approximate distribution of the sum of iid random variable is approximately normal $N(1.6,\ \frac{1.2}{\sqrt{200}})$.

$$
\begin{aligned}
\mathbb{P}(\bar{X} > 2) &\approx \mathbb{P}\left(Z > \frac{2 - 1.6}{0.085}\right) \\
&= \mathbb{P}(Z > 4.71) = 0
\end{aligned}
$$