HW5, for MATH441, STAT461, STAT561, due October 17th

1. To estimate the number of fish in a lake, some biologists capture some at random in nets, tag one of their fins and release them back to the lake. This is repeated for 20 tagged animals. Sometime later a random sample of 15 animals is obtained and 8 of those re-captured are found to have been tagged previously. Given these numbers, estimate N by plugging in different values of N to the formula

$$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

and finding the value of N that maximizes this probability.

You could write a loop to calculate the probability as a function of N, and plot the probability versus N. A reasonable estimate of N is then the maximum of this function. (This method of estimating a parameter is called *maximum likelihood* and will be explored much more in Statistical Inference in the second semester.)

2. I drive an unreliable car with electrical problems. The number of days, X, until the next breakdown for my car follows a geometric distribution with $p = 0.1$.

(a) Find the probability that my car doesn't break down over the next 10 days. I.e., find $P(X > 10)$.

(b) Given that my car doesn't break down in 10 days, find the probability that it doesn't break down the following day. I.e., find $P(X > 11|X > 10)$.

3. My wife drives a different unreliable car with electrical problems. The number of days, Y, until the next breakdown for my wife's car follows a Poisson distribution with rate $\lambda = 0.1$.

(a) Find the probability that her car doesn't break down over the next 10 days. I.e., find $P(Y > 10)$.

(b) Given that her car doesn't break down in 10 days, find the probability that it doesn't break down the following day. I.e., find $P(Y > 11|Y > 10)$.

(c) Would you rather my car or my wife's car? Why?

4. The hypergeometric is related to the binomial in that if X is hypergeometric by taking the limit as $N \to \infty$. The intuitive explanation for this is that for the hypergeometric, a certain proportion of the population is tagged, and sampling a tagged individual is similar to "success". However, the sampling is done without replacement, so that the same individual cannot be sampled twice. This leads to separate samples as not being independent. As $N \to \infty$, however, sampling with replacement is unlikely to result in the same individual being sampled twice, as long as $n \ll N$, and sampling is similar to sampling with replacement. This is similar to the situation for a binomial where you have n trials and success probability $p = m/N$.

Show that

$$\lim_{\substack{N\to\infty \\ m/N\to p}} \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} = \binom{n}{p} p^x(1-p)^{n-x}$$

5. A computer program generates a (pseudo)random number, U, which can be modeled as coming from a Uniform distribution with minimum 3 and maximum 5. Find

(a) $F(U)$. Give both an equation and a plot of the graph. Be sure to define F over all of $\mathbb{R}$

(b) $P(U < 3.5)$

(c) $E(1/U)$

6. An unreliable car breaks down according to an exponential distribution with rate $\lambda = 1/1000$ (meaning an average of one breakdown per 1000 miles). Find

(a) The probability that it goes 2000 miles without breaking down, i.e., $P(X > 2000)$.

(b) The probability that it goes another 2000 miles without breaking down given that it has already gone 2000 miles without breaking down, i.e., $P(X > 4000|X > 2000)$.

7. Let X be a random variable with density

$$f(x) = \begin{cases} \frac{k}{x^4} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find

(a) $P(X > 2)$

(b) $E(X)$

(c) $E(X^2)$

(d) $E(X^3)$

(e) $F(x)$.

(f) $P(X > 3|X > 2)$