Introduction to Statistics

As we go through our introductory lessons into statistics (single variable descriptive statistics), you will be collecting your own data and analyzing it using every method discussed in class. If you are interested in a particular kind of data, perhaps related to a sport, hobby or work, you can use this for the project, after it is approved by the teacher. Otherwise, you can use one of these three experiments outlined below.

1. Count the number of coin flips until you get the third head. Repeat this experiment 60 times. For example, if you flipped THTTHH, you would record 6 flips.

2. Place numbers (for example, 1-40) in a hat, and draw several (say, 4) at a time. Record the largest (or smallest) number drawn.

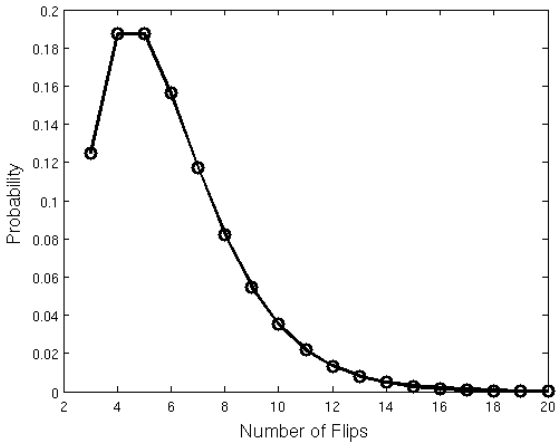
3. Find any large data set of measured values, that span several orders of magnitude (e.g., longest rivers in the world, tallest structures, the populations of various counrties). For each entry in the data set, record only the first digit (i.e., if a river is 6400 km, use a 6).

Begin the project by collecting the data, and finding the mean, median and mode of your data. As we learn more statistics, we will add more types of analysis for you to do with your data.

For the Teacher

A little bit of discussion of each project:

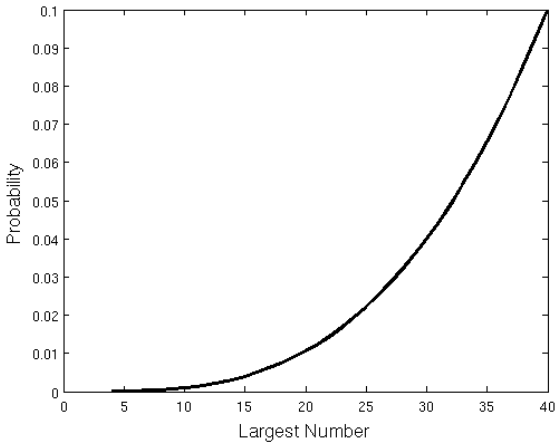
1. This is a simple example of a negative binomial distribution. The probability of getting the third head after k flips ($k \geq 3$) is $\binom{k-1}{k-3}\left(\frac{1}{2}\right)^k$. The probability density graph is:



2. This is a sample maximum without replacement. The probability that the largest number drawn (using the example numbers) is k ($k \geq 4$) is

$$\frac{\binom{k-1}{3}}{\binom{40}{4}}$$

The probability density graph is



3. The distribution of the first digit of a measured number is usually not uniform! It generally follows a distribution known as Benford's Law. This law is sometimes applied to financial data, or records of voter turnout, in an effort to detect fraud. Benford's Law says that the probability that the first digit is k, $k = 1, 2, 3, \ldots, 9$, is $\log_{10}\left(\frac{k+1}{k}\right)$. The probability density is