This lesson uses a demo that needs to be downloaded. Do the following.

1. Load `pmg`

2. At the command line type the following exactly (including capitalization)

```
> source("http://www.math.csi.cuny.edu/verzani/classes/MTH113/binomialDemo.R")
```

If there is an error message it didn't work. If there is no error message it worked.

3. Run the command

```
> binomialDemo()
```

A new window should popup

What does this "demo" do? Err, it demonstrates the binomial. Before seeing this though, lets review.

1 The Binomial Distribution

The binomial distribution is the random description of X when X counts the **number of successes** in n **identical trials**. The terms "success" and "trials" are generic, each application requires one to identify what they mean in the particular case.

For instance, If you toss a coin 25 times and count the number of heads using X, then X is binomial with $n = 25$ and $p = .5$. Why? Each trial is a tossing of a coin once. A "success" is a head. The success probability p is 0.5 – the probability of tossing a head – and X counts the number of successes. The term "identical trials" is taken to mean that each time a trial is done the same success probability applies *regardless* of the previous trials.

1.1 Problems

Identify if each of the following should have a binomial distribution. Specify what the trial is, a success, the number of trials and the success probability.

1. We roll a pair of dice 24 times. Let X be the number of double sixes.

2. A hospital performs 100 surgeries. Historically 3% have complications, on average. Let X be the number with complications in the 100 surgeries performed.

3. We randomly select 100 students from CSI. Let X be the number who are interested in the upcoming dormitories on campus.

2 Visualizing the binomial distribution

Once you see that the binomial distribution applies, you can take advantage of the known formulas for the distribution, the mean and standard deviation:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mu = E(X) = np, \quad \sigma = \sqrt{np(1-p)}.$$

How can we visualize the distribution. The figure shows the binomial distribution for $n = 10, p = .5$; $n = 100, p = .5$; $n = 10, p = .1$; and $n = 100, p = .1$. The shape depends on the value of n and p.

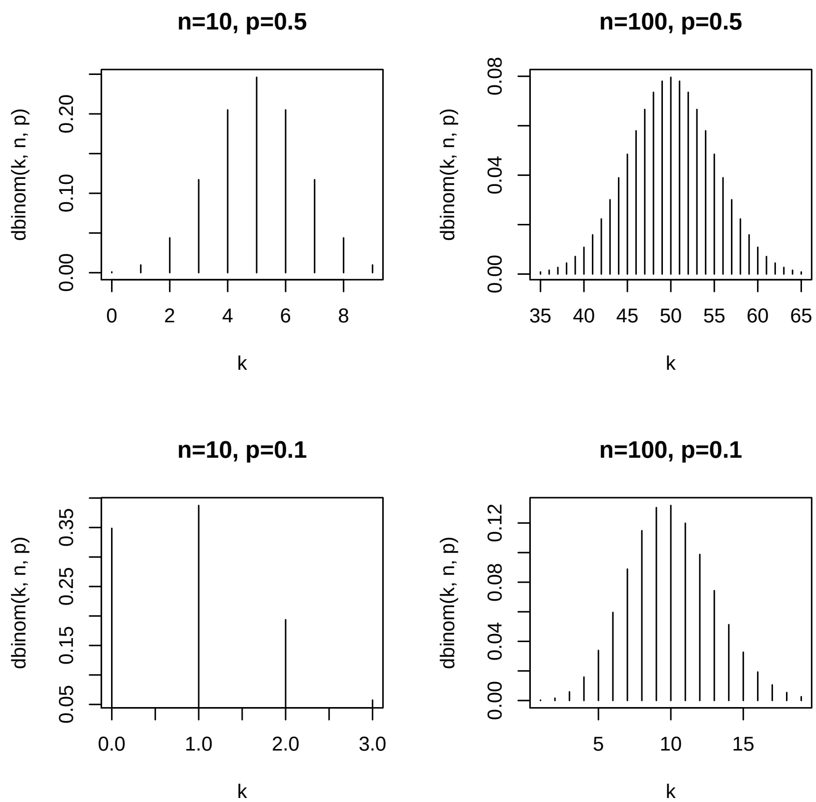

Figure 1: Binomial distribution for different values of n and p.

2.1 Problems

1. Which of the graphics, if any, in the Figure show a “bell shape?”

3 The demo

Now to return to the demo. First click on "Add one more." You should see a graphic appear. The lower part shows n spots, of which some random number are colored blue. Count the number of blue ones, and you should get the number under the bar in the upper graph.

Click "Add ten more." You should now see a barplot of 11 samples of a binomially distributed number with n and p specified by 10 and 0.5 (from the top left).

What is a sample? Each sample consists of n-trials with success probability p. The successes are colored "blue." The barplot simply shows the data set given by all of the samples. The last sample produces the bar colored blue.

If you click "Add ten more" a bunch of times, you will see that the distribution of the random samples, is very similar to that of probabilities in the upper left corner of the Figure. (It likely won't be identical as one is theoretical, and one is random.)

1. Estimate the mean of the barplot using the idea of balancing. Does your value make sense? Explain why or why not.

2. The length of 2 standard deviations is about the distance of the middle 2/3rds of the data. Can you estimate this? Compare to $\sqrt{np(1-p)}$.

3. For bell shaped data, a rule of thumb is that the standard deviation is about $1/6$th the range of the data. For you data estimate the standard deviation this way. Compare to the last problem.

4. Change n to 25. If you keep clicking "add ten more" does the barplot change dramatically? Click "add 10 more 3 times" so that there are 30 samples represented. What value is the most common? Theoretically, the most likely value is the mean or an integer adjacent to the mean if np is not an integer. Is that the case for your random sample?

5. Change $n = 25$ and $p = 0.1$. Now generate some random samples. Describe any differences and similarities between the barplot now and that before when $p = 0.5$.

6. For $n = 25$ and $p = 0.1$ describe the shape of the distribution (how many modes, symmetric or skewed, ...)

7. Repeat the previous when $n = 25$ and $p = .5$.

8. Repeat the previous when $n = 100$ and $p = .05$

9. A rule of thumb we will learn is that if np and $n(1-p)$ are both more than 5 then the shape of the binomial distribution will be approximately bell shaped. What is the mean and standard deviation of this bell shape?