Homework – Lecture #6

1. Recall the sign test. We said that the test statistic follows a binomial distribution. Why is this? What is n? p? k?

2. Derive the formula for the standard error of the mean. The following facts will be useful in this derivation:
   a. The standard error of the mean is the standard deviation (square root of the variance) of the sampling distribution of the mean. To generate this sampling distribution, you pick n data points from the population, add them together, and then divide by n.
   b. The variance of a sum of random variables is the sum of their variances (for independent random variables – if they are correlated this property goes away. But in this case, we are dealing with independent random variables).
   c. If n is a scalar and x is the random variable, then Var(x/n) = (1/n^2)Var(x). You can check this for yourself using the definition of variance.

3. Let's say you are infecting cells with a virus that contains GFP. You titrate the virus such that there is one virus for every cell on your plate (MOI = 1). We know it is common to describe the probability of *n* viruses infecting a single cell through a Poisson, where $P(n\ viruses\ in\ 1\ cell) = \frac{m^n e^{-m}}{n!}$ and m = MOI. So, the probability of 2 viruses infecting a single cell is $\frac{1^2 e^{-1}}{2!}$ = 0.1839. Given that your plate contains *k* cells, we can then compute the expected number of cells with *n*=2 viruses. Of course, the number of cells infected with 2 viruses that we observe in our experiment (after doing some single-cell PCR on our infected cell cultures) will vary from experiment to experiment. Let's say we do one experiment – we infect cells, and then count the number of cells infected with 2 viruses. The number of cells we observe seems suspiciously low to us, and leads us to think that maybe a cell is less likely to be infected by another virus once it is already infected, which would be a very exciting result. How should we quantify the potential result that the number of doubly-infected cells is much lower than what we would expect? How can we extend this to include triple-y (and higher) cells?

4. What is your favorite probability distribution and why?