Math4310/Biol6317, Fall 2011
Problem Set 6, due Thursday, Oct 6

Problem 1. You repeat Mendel's pea experiment at home. Let p denote the unknown proportion of peas in your garden that are wrinkled. Suppose that $X = 12$ of a sample of $n = 20$ peas collected at random locations are found to be wrinkled.

a. Use the CLT to create a 95% confidence interval for the true proportion of peas that are wrinkled. Interpret your results.

b. You plan a much larger study. How large should n be to have a margin of error (half the width of the confidence interval) no larger than $.01$ for estimating the proportion of wrinkled peas with a 95% confidence interval? Use the fact that $p(1-p) \leq 1/4$. Also try the calculation with the estimate of p from the current study.

Problem 2. This problem investigates the performance of the Wald confidence interval.

a. Using R, generate 1000 observations $x_1, x_2, \ldots x_{1000}$ of a Binomial random variable X for $n = 10$ and $p = .3$ Calculate the percentage of times that

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$$

contains the true value of p. Here for each i, $\hat{p} = x_i/n$ where x_i is the observation of the binomial variable. Do the intervals appear to contain the true value of p as frequently as they are supposed to?

b. Repeat the calculation only now use the interval

$$\tilde{p} \pm 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$$

where $\tilde{p} = (x_i + 2)/(n+4)$. Does the coverage appear to be closer to .95?

c. Repeat this comparison (parts a. and b.) for $p = .1$ and for $p = .5$. Which of the two intervals appears to perform better?