Math 4397/6397, Fall 2009
Problem Set 6, due Thursday, Oct 8

Problem 1. In this problem we will verify that standardized means of iid normal data follow Gossett's t distribution. Randomly generate $1,000 \times 20$ normals with mean 5 and variance 2. Place these results in a matrix with $1,000$ rows. Using two `apply` statements on the matrix, create two vectors, one of the sample mean from each row and one of the sample standard deviation from each row. From these $1,000$ means and standard deviations, create $1,000$ outcomes which come from a t distribution. Now use R's `rt` function to directly generate $1,000$ outcomes of t random variables with 19 df. Compare the pth quantiles for $p = 0.1$, $p = 0.2$, $\ldots$, and $p = 0.9$ for both the constructed t random variables and R's t random variables. Do the quantiles seem to agree?

Problem 2. Here we will verify the chi-squared result. Simulate $1,000$ sample variances of 20 observations from a normal distribution with mean 5 and variance 2. Convert these sample variances so that they should be chi-squared random variables with 19 degrees of freedom. Now simulate $1,000$ random chi-squared variables with 19 degrees of freedom using R's `rchisq` function. Compare the pth quantiles for $p = 0.1$, $p = 0.2$, $\ldots$, and $p = 0.9$ for both the constructed t random variables and R's t random variables. Do the quantiles seem to agree? Describe why they should.

Problem 3. If $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ then we know that $(n-1)S^2/\sigma^2$ is chi-squared with $n-1$ degrees of freedom. Use the fact that $E[S^2] = \sigma^2$ to derive the expected value of the chi-squared distribution.

Problem 4. You repeat Mendel's 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 5. 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 have the coverage that they are supposed to?

b. Repeat the calculation only now use the interval

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

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

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