Assignment #5 (revised) — Due Wednesday, February 25, 2009, by 5:00 P.M.

Turn in homework in lecture, discussion, or your TA's mailbox. Indicate the discussion section in which you expect to attend to pick up this assignment on the assignment.

311: Monday 1:20–2:10 **312:** Monday 12:05–12:55

1. In Example 6.3.7 beginning on page 310, the textbook gives a formula for the confidence interval for a Bernoulli model with the form
$$\bar{x} \pm z\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$
where z is a standard normal quantile. In the previous homework assignment, you used the `binomialSimulation()` to approximate the coverage probability for $n = 20$ and $\theta = 0.3$. This problem asks you to find exact coverage probabilities for a wide range of n and θ.

   (a) Notice that the confidence interval contain θ if and only if
   $$|\bar{X} - \theta| \le z\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$
   where $\bar{X} = X/n$ and $X \sim \text{Binomial}(n, \theta)$. Square both sides of this inequality and reexpress it as an interval for X. In other words, find expressions a and b such that
   $$|\bar{X} - \theta| \le z\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \qquad \text{if and only if} \qquad a \le X \le b$$

   (b) For $n = 20$ and $\theta = 0.3$, find the end points a and of a 95% confidence interval and compute the exact coverage probability using `pbinom()` in R. (For example, `pbinom(15,20,0.3) - pbinom(4,20,0.3)` calculates $\mathsf{P}(X \le 15) - \mathsf{P}(X \le 4) = \mathsf{P}(5 \le X \le 15)$.)

   (c) For $n = 21$ and $\theta = 0.3$, find the end points of this interval and compute the exact coverage probability using `pbinom()` in R. Notice the difference!

   (d) The newly updated file `stat310.R` has a function `exact.binomial.coverage()` that implements the equation you should have found in this problem. Plot the true coverage probability for $n = 100$ as θ varies from 0.05 to 0.95 by 0.001. Comment on the accuracy of 95% confidence intervals using this traditional formula. For a "typical" θ between 0.05 and 0.95, what is the true coverage probability of a traditional 95% confidence interval?

   Here is an example on how to make such a plot with an added line at the desired coverage probability.

```
> source("stat310.R")
> theta = seq(0.05,0.95,0.001)
> coverage = exact.binomial.coverage(100,theta)
> fig = xyplot(coverage ~ theta, type="l",
               panel = function(x,y,...) {
                 panel.xyplot(x,y,...)
                 panel.abline(h=0.95,col="red",lty=2)
               }
              )
> print( fig )
```

2. Other textbooks recommend a modification to the traditional approach by constructing confidence intervals of the form

$$\tilde{\theta} \pm z\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n+4}}$$

where z is a standard normal quantile and $\tilde{\theta} = (x+2)/(n+4)$. Notice that this is the traditional formula applied to a modified data set where four "phantom observations", two successes and two failures, are added to the original sample for a modified sample with $x+2$ successes in $n+4$ trials. More generally, the adjustment could be of the form $\tilde{\theta} = (x+a)/(n+2a)$ and n is replaced by $n+2a$ in the denominator of the SE.

   (a) Find exact coverage probabilities using the this adjustment with $a = 2$ when $n = 20$ and $n = 21$ when $\theta = 0.3$. Compare to the previous problem.
   (b) Plot the true coverage probabilities for $n = 100$ and θ varying from 0.05 to 0.95 as in the previous problem. Use `adj=2` in `exact.binomial.coverage()`.
   (c) Comment on the salient differences between this plot and the plot from the previous problem.
   (d) Explore differences in these methods for other n. Which method is better for more accurate confidence intervals in general?

3. Do Exercise 6.3.7.

4. Do Exercise 6.3.11.

5. The file `wi-tornado.txt` contains the total number of tornadoes in Wisconsin for each year from 1950 to 1995. Assume that the number of tornadoes in a given year is a Poisson random variable. Find a 95% confidence interval for the average number of tornadoes in Wisconsin per year during a time period with comparable climate.

6. Measurements in centimeters are assumed to be an i.i.d. normal sample. The sample ($n = 10$) is 4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, and 5.3.

   (a) If we assume that the distribution for a single measurement is $\mathrm{N}(\mu, \sigma^2)$ where σ^2 is unknown, find a 95% confidence interval for μ.
   (b) Find the p-value for the two-sided t-test with null hypothesis $H_0\colon \mu = 4$.
   (c) If a number μ_0 is in the confidence interval you found in part (a), what is true about the p-value for the two-sided t-test with null hypothesis $H_0\colon \mu = \mu_0$?
   (d) If a number μ_0 is not in the confidence interval you found in part (a), what is true about the p-value for the two-sided t-test with null hypothesis $H_0\colon \mu = \mu_0$?

Work to do, but not turn in.

- Read sections 7.1–7.3.