Infuse Simulations into Probability and Mathematical Statistics
Melinda Harder

I created the lab “Confidence Intervals for Proportions” for my mathematical statistics students so they could experimentally compare approximate, exact, and “plus four” intervals for a proportion. The lab was inspired by Christopher Lacke’s e-mail “Teaching a method that isn’t used” (12/06/2005 on the isostat list) and by the flurry of responses to it.

The poster shows the results from my class of twenty-four students. Each student was assigned three values for p “the probability of success”, four values for n “the sample size”, and instructed to do 400 simulations for each combination of n and p. I randomly assigned four students to each of the following groups

		p			p			p	
	.1	.4	.7	.2	.5	.9	.3	.6	.8
5									
n 15									
30									
70									
10									
n 20									
50									
100									

The group assignments were done in class using a deck of cards containing A, K, Q, J, 10, 9 in all four suits. Students selecting an ace were assigned to group 1, p = .1, .4, .7 and n = 5, 15, 30, 70, and so on. So there are1600 simulations per combination, and if we assume a success rate of approximately 95% (success occurs when the true value of p is contained in the interval) then the error is approximately 0.5% for the estimated coverage probabilities.

This is just one example of the labs I’ve created. Many of the students who take mathematical statistics at Bates College are computer savvy and enjoy using simulations to do experiments. They were surprised that such a strange method (adding four observations, two successes and two failures) would give better coverage probabilities (closer to 95%) than the exact or approximate methods.

Reference A. Agresti and B.A. Coull “Approximate is better than ‘exact’ for interval estimation of binomial proportions”, *The American Statistician*, 52 (1998), pp. 119-126

From your text, to find a $100(1-\alpha)\%$ confidence interval for a binomial proportion p compute

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

where $\hat{p}$ is the sample proportion of "successes" and $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. This interval is an approximate interval, and the accuracy of the approximation depends both on the sample size n and the value of p. To avoid approximation, some texts recommend exact intervals. (These are the intervals that Minitab computes by default.) In practice these intervals are conservative because the coverage is always $100(1-\alpha)\%$ or more, the "more" due to the fact that the binomial distribution is a discrete distribution. Another method suggested by Agresti and Coull (1998) is to add "two successes" and "two failures" to the sample, and then construct a confidence interval as described above for the approximate interval. That is, instead of using $\hat{p} = y/n$ where y is the number of "successes" in a sample of size n, use $\tilde{p} = (y+2)/(n+4)$ as the estimate of the sample proportion, and use $\tilde{n} = n+4$ as the adjusted sample size. The $100(1-\alpha)\%$ interval for p is

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}}\,.$$

To illustrate the three methods, here is an example. Suppose you want to estimate the average proportion of blue M&Ms that are put into every bag of plain M&Ms. You have a sample bag containing 56 candies, and 7 of them are blue. Find a 95% confidence interval for the theoretical proportion of blue M&Ms.

Method 1 (approximate normal interval) The sample proportion is $\hat{p} = 7/56$.

$$\frac{7}{56} \pm 1.96\sqrt{\frac{(7/56)(49/56)}{56}}$$

The 95% confidence interval for p is (.03838, .21162).

Method 2 (exact)

$$\sum_{k=7}^{56}\binom{56}{k}p_L^k(1-p_L)^{56-k} = .025$$

Solving for the lower endpoint, $p_L = .051765$.

$$\sum_{k=0}^{7}\binom{56}{k}p_U^k(1-p_U)^{56-k} = .025$$

Solving for the upper endpoint, $p_U = .240733$.

The 95% confidence interval for p is (.051765, .240733).

Method 3 ("plus four") The adjusted sample proportion is $\tilde{p} = 9/60$.

$$\frac{9}{60} \pm 1.96\sqrt{\frac{(9/60)(51/60)}{60}}$$

The 95% confidence interval for p is (.05965, .24035).

Which of these methods for constructing 95% confidence intervals works best in practice for finding intervals that actually contain p ninety-five percent of the time? You will investigate this question in Minitab using the attached macro. Before running the macro, you will need to specify the sample size (i.e. the number of trials, n), the probability of "success" p, and the number of experiments. The macro will construct the three types of intervals for each experiment and check to see whether p is in the interval. If the intervals are 95% confidence intervals, this should happen in approximately 95% of the experiments.

A copy of the macro is saved in webct. To copy it into your Paris folder do the following. From the internet or netscape type in the web address
http://webct.bates.edu
Log in with your user id and password. Click on our class and click on class files. Click on **macro for ci lab**, and select the tab marked **View** to see the file. To copy the file into notepad click on the screen. Select **Edit** and select **Select All**. The macro file should be highlighted. Then click on **Edit** and select **Copy**. Click on **Start**, select **Programs>Accessories>Notepad**. Click on **Edit** in notepad and select **Paste**. The file should appear in notepad. Press **File** and click on **Save As**. Select **Network Neighborhood**, **Paris**, and **your folder**. Type in a filename **ci**.

In Minitab type
let k1=n
let k2=p
let k3=400
where n and p are numbers you assign (for example 5 and .2). Then run the macro.
%**Paris****username****ci.txt**

What you will hand in
Hand in a table of results for all the combinations of n and p (12 combinations per person) containing the observed proportion of times that each of the three intervals contained p. I will assign sample sizes and values of p in class, and show you how I would like the tables so that I can compile your results in a larger experiment. Write a paragraph describing your results. Which type of interval contained p a proportion of the time closest to the target of 95%? Do you notice any trends with sample size n or with the value of p?

```
gmacro
ci
random k3 c1;
binomial k1 k2.
erase c2 c3
do k4=1:k3
let k5=c1(k4)
let k6=k5+1
let k7=k1-k5+1
let k8=k1-k5
if k5=0
  let k9=0
else
invcdf .025 k9;
beta k5 k7.
endif
let c2(k4)=k9
if k5=k1
  let k10=1
else
invcdf .975 k10;
beta k6 k8.
endif
let c3(k4)=k10
enddo
let c4=c1/k1
let c5=1.96*sqrt(c4*(1-c4)/k1)
let c6=c4-c5
let c7=c4+c5
let c8=(c1+2)/(k1+4)
let c9=1.96*sqrt(c8*(1-c8)/(k1+4))
let c10=c8-c9
let c11=c8+c9
let c12=(k2>=c2&k2<=c3)
let c13=(k2>=c6&k2<=c7)
let c14=(k2>=c10&k2<=c11)
let k4=sum(c12)/k3
let k5=sum(c13)/k3
let k6=sum(c14)/k3
name k4 "exact"
name k5 "approx"
name k6 "plus4"
print k4 k5 k6
endmacro
```