Coin Toss Exercise

Day 3 Learning Objectives covered:

- Compute summary statistics from a dataset
- Use computer simulations to build intuition about random variables
- Simulate different kinds of random distributions
- Simulate sampling distributions

1) We have a coin whose probability of landing on heads is unknown (i.e. We do not know if it is a “fair” 50/50 coin.). We flip this coin 50 times and obtain 31 heads. What is our best estimate of the probability of obtaining heads, p(H), on one flip?

2) If we flip a different coin 5000 times and obtain 3100 heads, what is our best estimate of p(H)?

3) As you should be aware from the Rattus binomialis exercise, even a completely fair coin [p(H) = 0.5] will occasionally give 31 (or more) heads from 50 tosses. Thus we can never be completely sure of the true p(H) of any coin based on a limited number of tosses. But if we compare the coins from questions 1 and 2, for which coin are we more confident in our estimate of p(H)?

4) This is a common problem when dealing with summary statistics derived from data. We have some measure of the data, such as the mean, and we would like to know how confident we should be in our estimate of the true mean. That is, we want a <u>confidence interval</u> around our estimate.

You’re probably already familiar with one such confidence interval for the mean, which is the standard error of the mean (SEM). Say you want to estimate the true mean of some quantity for a large population. The best approach to do this is to take a sample and compute the sample mean. In this case, the mean ± SEM [mean-SEM to mean+SEM] is the 67% confidence interval. This means that, if we repeated our sampling experiment many, many times, and calculated a confidence interval around our sample mean each time, then 67% of the time the true population mean would lie within this confidence interval.

In our case, we are interested in the 95% Confidence Interval for the intrinsic head probability of a coin that landed on "heads" 31 out of 50. Let's say we found the Confidence Interval. Which of the following statements is true?

- If we flip the same coin 100 times, its true "heads" probability will lie within the Confidence Interval 95 times.

- If we flip 100 coins and estimate their probability of landing on “heads”, then the estimated “heads” probability will lie within the Confidence Interval 95 times.
- For 100 coins that landed on heads 31 out of 50 times, 95 will have a true "heads" probability that lies within the Confidence Interval.
- We are about 95% sure we got it right.

5) Imagine you had a large collection of coins, all of which have landed on "heads" 31 out of 50 times in a recent trial. Imagine also that you know the intrinsic "heads" probability of all of them (maybe because you are all-knowing). How would you create an interval that contains 95% of all the intrinsic "heads" probabilities in your collection?

6) Imagine you have a vast number of coins and know all their intrinsic probabilities of landing on "heads". Imagine also you have tons of time. How would you create a (smaller) collection of coins that landed on "heads" 31 times out of 50 in a recent trial?

7) How, then, would we tackle the problem of finding the confidence interval for the intrinsic "heads" probability of a coin that landed on heads 31 times out of 50? Assume we were all-knowing master coin-flippers with plenty of time on our hands.

8) Write a MATLAB code that simulates this process. What is your estimate for the intrinsic heads probability of the coin?

9) What is the lower limit of your confidence interval?

10) What is the upper limit of your confidence interval?

Versions:

MIS wrote initial exercise on Learning Catalytics, 07 Sep 2013

RTB wrote introductory questions, 13 Aug 2014

MIS merged both versions, added learning objectives, 17 Aug 2014