Math 362 Lab 2 Spring 2006

29th January 2006

Suppose that a geographically isolated area has a population of 200 wolves. Suppose that 20 of these wolves are humanely captured and tagged. Now suppose that a random sample of 15 wolves from the same population is examined. What is the probability that 2 of the wolves in this random sample are tagged? Here is one way to analyse this problem. The number of ways to choose 13 untagged animals is $\binom{180}{13}$. The number of ways to choose 2 tagged animals is $\binom{20}{2}$. Thus the probability of finding two tagged animals in a sample of size 15 is,

$$\frac{\binom{180}{13}\binom{20}{2}}{\binom{200}{15}}.$$

It is possible to use R to evaluate this rather simply. Just type in,

$$choose(180, 13) * choose(20, 2)/choose(200, 15),$$

and press enter. You should get about .279.

Actually this is not the most interesting form for the problem. Suppose 20 animals are tagged as above but the total number of wolves in the population is not known. Now suppose that 20 wolves are randomly captured and it is found that 5 of them are tagged. If we let n denote the number of untagged wolves then the probability that 5 in a random sample of 20 are tagged is,

$$p(n) = \frac{\binom{n}{15}\binom{20}{5}}{\binom{n+20}{20}}.$$

We could try to estimate n by choosing it so that $p(n)$ is maximized. This is called a maximum likelyhood estimate (note that this is not maximizing the probability that n occurs–it is merely the choice of n that maximizes the probability that 5 out of 20 animals are tagged in a random sample). This lab is devoted to showing you how to use R to find this value for n. First we create a vector x with entries $(p(1), p(2), p(3), \ldots, p(N))$ and then we plot x to see where it has its maximum. How big one needs to pick N is maybe not clear to begin with–it just needs to be big enough so that it is bigger than the place where $p(n)$ takes its maximum value. I will choose N=120. First we create two vectors of length 120 with the commands, $i = 1 : 120$ and $x = 1 : 120$. The vector 1:5 is, for example, (1,2,3,4,5). Both the vectors i and x have entries 1 to 120. We will use i as an index vector and we will now change x to give the vector we desire with the command,

$$for\,(n\,in\,i)\,x[n] = \frac{choose(n,15) * choose(20,5)}{choose(n+20,20)},$$

For $n = 1$ to $n = 120$ this sets the value of the n^{th}component $x[n]$ of the vector x equal to $p(n)$. Incidentally, for reasons that I don't completely understand this command will not create a vector x on the fly–x needs to be a vector defined beforehand with 120 entries. To see the graph of $p(n)$ just issue the command,

$$plot(x).$$

Problem 1. Use the plot of p(n) to find the value of n which maximizes p(n). What estimate for the total population, n+20, of wolves does this give? Is there a simpler way to come up with this estimate? Hint: 5/20 is the fraction of the sample which is tagged–imagine that this is a good estimate of the total fraction of the population of wolves which is tagged.

Note that the graph will give you a good idea where to look for the maximum. You can more clearly check what the values are near the maximum by issuing the commands $x[k]$ for k near the value at which the maximum is achieved.

In the preceeding problem you found a simple estimate for $n + 20$which did not involve using the maximum likelyhood calculations at all. You might think that this makes those calculations superfluous. However, the maximum likelyhood calculations can give you some feel for how much confidence you want to invest in this estimate for $n + 20$.

Problem 2. Look at the plot of p(n) you found for problem 1 to determine how far on either side of the maximum likelyhood estimate you need to move in order for the value of p to fall to 1/2 its value at the maximum. This gives you a kind of confidence interval about the maximum likelyhood estimator. Admittedly it is a little hard to give a precise meaning to this confidence interval but it does give you some feeling for the accuracy of the estimate.