GROUP TESTING: AN OPTIMIZATION PROBLEM FOR CALCULUS

by: Dan Teague
The North Carolina School of Science and Mathematics

Introduction

For the past several years, I have posed the following problem to my elementary calculus and mathematical modeling classes:

> Suppose that you have a large population (N) that you wish to test for a certain characteristic in their urine. Each test will be either positive or negative. Since the number of individuals to be tested is quite large, we can expect that the cost of testing will also be large. How can one reduce the number of test needed to screen everyone and thereby reduce the costs? If the urine could be pooled by putting G samples together and then testing the pooled sample, the number of tests required might be reduced. What is the relationship between the probability of an individual testing positive (p) and the group size (G) that minimizes the total number of tests requested?
>
> Use your solution to determine the number of tests required to find 100 individuals who will test positive in a population of 1,000,000.

The students spend approximately a week working in small groups on the problem. My role is to offer encouragement and suggestions as they develop their solutions. Over the year, the students have used many different approaches in solving this problem and developed several different models in the process. This article presents one of the more interesting solution paths. Although solutions need to be integral, it uses a continuous model and develops a theoretial optimum by assuming a simplifying, worst case situation; specifically, if a group tests positive, exactly one person in the group is positive.

The Full Lab Solution

The following are the variables that will be used in the solution:

N	Total number of persons
M	Maximum number of persons that can be tested at one time
P	Probability of person being positive
G_k	Size of kth group

The solution begins with the argument,

> The lab will likely have only a limited amount of equipment to use in the testing. Suppose the maximum number that can be tested at any one time is M. What would happen if, at each stage of the testing, we use all the available equipment, that is, test M groups each time.

Recall that we assume the worst case, that is, in any group that tests positive, exactly one person in the group is positive. This means that since we expect $N \bullet p$ individuals to test positive, we will have $N \bullet p$ groups testing positive as well.

If there are N persons to be tested, then the first round of tests would have a group size $G_1 = \frac{N}{M}$. Since we expect $N \bullet p$ individuals to test positive, under the worst case assumption, we would have $N \bullet p$ groups testing positive and therefore $N \bullet p \bullet G_1$ persons to be retested. The second testing would be in groups of size $G_2 = \frac{NpG_1}{M} = \frac{N^2 p}{M^2}$, resulting in $N \bullet p \bullet G_2$ persons needing retesting. Continuing, we find that the group size for the kth group is $G_k = \frac{N^k p^{k-1}}{M^k}$. We will have finished the testing when the group size is one, so the number of iterations of the procedure is determined by the value of k when $G_k = 1$. Solving $\frac{N^k p^{k-1}}{M} = 1$ for k, we find that

$$k = \frac{\ln(p)}{\ln\left(\frac{Np}{M}\right)}.$$

In this scenario, we have k iterations of M tests, so the total number of tests is given by $T(M) = M\ k = M\left[\frac{\ln(p)}{\ln\left(\frac{Np}{M}\right)}\right]$. We have assumed that we should use all of the testing facilities available, but this assumption is contradicted by the model. The function T(M) defined above has a minimum value, indicating an optimal value for M. What value of M minimizes the total number of tests? To minimize T, we differentiate with respect to M to find

$$\frac{dT}{dM} = \frac{\ln\left(\frac{Np}{M}\right)\ln(p) - M\ln(p)\left(\frac{M}{Np}\right)\left(-\frac{Np}{M^2}\right)}{\ln^2\left(\frac{Np}{M}\right)}.$$

The derivative is defined for all M>0, and setting the derivative equal to zero and simplifying yields the equation

$$\ln\left(\frac{Np}{M}\right)+1=0$$

Solving for M we find that

$$M=Npe.$$

If possible, to minimize the number of tests, perform M=Npe tests on each iteration with a group size on the kth iteration of

$$G_k=\frac{N^k p^{k-1}}{M^k}=\frac{N^k p^{k-1}}{(Npe)^k}=\frac{1}{pe^k}.$$

The total number of tests required using this procedure is given by

$$T=(Npe)\left(\frac{\ln(p)}{\ln\left(\frac{Np}{Npe}\right)}\right)=(Npe)(-\ln(p))\text{ total tests.}$$

In our problem, with 1,000,000 people and 100 testing positive, we would need only $(Npe)(-\ln(p)=(1{,}000{,}000)(.0001)(e)(-\ln(.0001))\approx 2{,}500$ tests!

Students are always surprised to see e show up in the solution. Of course, while this theoretical result is pleasing, it may not be realizable, for $G_k=\frac{1}{pe^k}$ may be too many specimen to handle in a single group. ✍