

NCAAPMT Calculus Challenge 2010-2011

Calculus Challenge #6

Solutions due January 19, 2011

Since schools have differing schedules, this challenge will be due after the winter break. School that have exams before their break should wait until they return to do the problem. Schools that have exams after the break, should do the problem in the next two weeks. Hopefully, everyone will have an opportunity to work through this challenge without compromising their exam study time. This problem is a bit more extensive than some earlier problems and combines classical calculus and mathematical modeling techniques. If you have questions or something is unclear, please send me a note (teague@ncssm.edu) and I will give some clarification.

The Group Testing Problem

Suppose that you have a large population (N) that you wish to test for a certain characteristic in their urine (all NCAA athletes for steroid use, for example). You will take a sample from all N individuals and test each sample, with either a positive or a negative result. In this problem we will ignore all false positives and false negatives, since the tests are quite good. Since the number of individuals to be tested is very large, we can expect that the cost of testing will also be large. How can we reduce the number of tests needed to screen everyone and thereby reduce the costs? If the urine samples could be combined or pooled by putting a portion of several samples together and then testing this pooled sample, the number of tests required could be reduced.

Suppose we pool the samples into groups of size G . That is, we have one test tube (perhaps large) with a portion of the sample from G individuals in it. 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 required to find all positives in a population of size N ?

Use your model to determine the number of tests needed to find 100 individuals who are positive in a population of 1,000,000.

Before getting too involved with writing equations, consider the following thought questions:

- i) If a pooled sample of size G tests negative, what do we know about the individuals in the group?
- ii) If a pooled sample of size G tests positive, what do we know about the individuals in the group?
- iii) If p is large (close to 0.5), what should be true about G ?
- iv) If p is very small, what should be true about G ?

Creating the Model

One essential aspect in developing any model is to first consider the simplest case that embodies the essence of the problem. If we cannot solve the simplest version of the problem, then we certainly will not be able to solve a more involved and sophisticated model that is perhaps more realistic. Further, the solution to the simplest situation often is helpful in arriving at a more general solution as we will soon see.

In this setting, the simplest form of the group testing problem is to pool the samples into groups only once. Then test each pooled sample and if it tests negative, remove all the individuals in the group from consideration. For any pooled sample that tests positive, we systematically test every individual in the group to determine if they were positive or negative.

Let's create a model for this simplest setting. Later we will improve upon the model by allowing more than one round of testing in groups.

To determine the number of tests needed, we need to consider the two testing settings. First, everyone is put in a group and test the group. Then, anyone in a group that tested positive must be retested individually. The number of tests in total is the sum of the number of tests in the two settings.

1) Explain why the number of groups tested is $\frac{N}{G}$. (In this model, we will ignore any remainders in this division, since it would at most one test to the total).

2) We need to determine many people need to be retested. Recall that the probability of an individual testing positive is p . Explain why the expected number of tests is modeled by

$$T(G) = \frac{N}{G} + \left(1 - (1-p)^G\right) \left(\frac{N}{G}\right) \cdot G = \frac{N}{G} + \left(1 - (1-p)^G\right) N.$$

3) Given the model, $T(G) = \frac{N}{G} + \left(1 - (1-p)^G\right) N$, explain why the techniques of calculus are not suitable for finding the value of G that minimizes T .

Linear Approximations

In practice, we often replace a messy or difficult expression with its linear approximation. In this case, we want to replace the expression $\left(1 - (1-p)^G\right)$ in the equation above with its linear approximation. We will then have a function for which we can indeed find a solution using calculus.

4) a) Find the linear approximation for $f(G) = \left(1 - (1-p)^G\right)$ as a function of G at $G = 0$.

b) Find the linear approximation for $f(p) = \left(1 - (1-p)^G\right)$ as a function of p at $p = 0$.

5) Substitute the simpler approximation from 4) into the function $T(G) = \frac{N}{G} + N\left(1 - (1-p)^G\right)$ and find the value of G that minimizes T .

6) Use the solution from 5) to determine the number of tests needed if $N = 1,000,000$ and $p = 0.0001$. What group sizes are required and how many total tests are needed?

7) Use your solution to 5) to determine which values of p would testing in groups require more total tests than just testing all N individuals separately? For example, if $p = 0.75$, it would be foolish to test in groups.

Improving the Model

In developing the model, we have assumed that we would test only once in groups and then test everyone remaining individually. Clearly, we could regroup those remaining and do a second round of group tests, before testing everyone remaining individually.

8) Use the relationship between G and p from 5) that minimizes the number of tests required and find the group size in terms of p for a second round of group tests. That is, instead of testing all members of groups that tested positive in the first test, we form new groups and retest the groups. What group size would you use in the second group tests? Why is the value of p different in the second grouping than in the first?

9) If you repeatedly used the solution to 5) and the information gained in 8) in retesting, determine the total number of tests required to find 100 individuals who are positive in a population of 1,000,000 by filling in the table below. Remember to use your solution to 7) as a stopping criterion.

Round	Number to be Tested or Retested	Probability of Testing Positive	Size of Optimal Group	Number of Group Tests
1	1,000,000	0.0001		
2				
3				
4				
5				
6				