Math 425, Winter, 2003, Sec. 7.

First Group Problems: Solutions

Problem 1. Testing Blood.

It is sometimes necessary to give blood tests to a large population, say, of N people. Here are two possible strategies for testing. (We are assuming that the test is 100% effective and efficient.) (i) The first strategy is to simply administer the test to each of the N individuals. (ii) The second strategy is to group the individuals into groups of k individuals, where $1 < k \leq N$, and to pool their blood samples together and test this pooled sample. If this pooled sample tests *negative*, this one test suffices for the k people. If the test of the pooled blood is *positive*, then each of those k individuals is tested separately, and altogether $k+1$ tests are required for the k people. (It is reported that adoption of strategy (ii) by the military around the time of the Second World War resulted in a savings of approximately 80% in the cost of testing.)

Assume the probability that a person will test positive is p, and that each of the N individuals is "stochastically independent", *i.e.*, their tests are independent of one another.

(a) What is the probability that a pooled sample of k people will test positive?

***** We are assuming that each of the people tested is an independent Bernoulli trial (coin flip) with probability p of being positive for the test. Therefore, the probabilty of the batch of k testing positive is the probability of getting one positive out of k tests, that is, $1-(1-p)^k$.

(b) What is the expected value of the number, $\mathbf{X}$, of tests necessary under strategy (ii)?

***** We will have $\frac{N}{k}$ groups of k individuals. For each of those groups, if somebody is positive, we need $k+1$ tests, and other wise, only 1 test. Weighting these two possibilities with the appropriate probabilities, we get that the expected number of tests for each group of k is just

$$(k+1)\cdot(1-(1-p)^k)+1\cdot(1-p)^k,$$

and so, for the whole group of N, we get the expected number of tests to be

$$\frac{N}{k}\{(k+1)\cdot(1-(1-p)^k)+1\cdot(1-p)^k\}.$$

If $\mathbf{X}$ is the random variable "the number of tests required under this strategy", this amounts to saying

$$E(\mathbf{X})=\frac{N}{k}\{(k+1)\cdot(1-(1-p)^k)+1\cdot(1-p)^k\}.$$

(c) Which k will minimize the expected number of tests necessary under plan (ii)? You may not be able to evaluate this explicitly, so you may have to just discuss this point, *i.e.*, what does it depend on and how?

***** Simplify the above to

$$E(\mathbf{X})=N[\frac{k}{k+1}-q^k],$$

where $q = 1 - p$. Then, to minimize the expected number of trials, take $\frac{dE}{dk} = E'$ and set it equal to 0. This gives

$$E'(k) = N[-\frac{1}{k^2} + \log(\frac{1}{q})q^k].$$

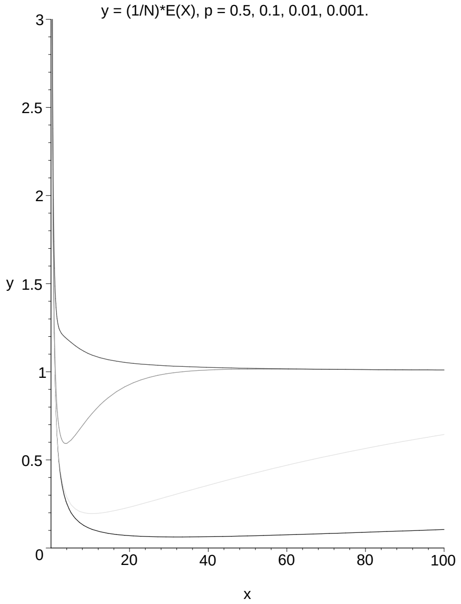
("log" $=$ natural log.) Graphically, this means we want the value of k, say k_o, where two curves cross,

$$y = \frac{1}{\log(\frac{1}{q})} \cdot q^{-k},$$

and

$$y = k^2.$$

These may or may not cross, dpending on the size of p (see the case $p = 0.5$ below, the top curve). Furthermore, even if they do cross, it may be for a value of k greater than N (see $p = 0.1$, 0.01, 0.001, graphs 2 - 4 below). If we let $q = 1 - p$ get very close to 1 (i.e., the disease is rarer and rarer), then k_0, the best value of k, moves to the right, regardless of the size of N. The crossing may not come before $k = N$, however, in which case the right endpoint of the interval, that is, $k = N$ would be optimal. This says that if a disease is rare enough, or perhaps, the test is so rarely positive, you don't really expect too many positives, and you just lump everybody together. The q's for which this is true depend on the N. On the other hand, if the disease is common enough (i.e., p is close to 1, admittedly not very realistic, then the best choice of k is $k = 1$, since the exponential function will already be bigger than the quadratic when $k = 1$. This phenomenon does NOT depend on N. Also note that I have been treating k as a continnuous variable, which it is not. You would have to choose the nearest whole number to the minimum found continuously.



Possible values of p: 0.5, 0.1, 0.01, 0.001,
from top to bottom.

Problem 2. A Statistical Test.

A U of M professor got a ticket twelve times for illegal overnight parking. All twelve tickets were given either Tuesdays or Thursdays. Find the probability of this event. Was his renting a garage only for Tuesdays and Thursdays justified?

There are a couple of possible answers to this, depending on how you model the situation. The language of the problem is meant to be vague since that is how probability situations are often presented to you in the real world. Be sure to discuss the set-up of this problem carefully. Bonus to your group if you find several ways to model the situation.

(i) "Balls in urns" model: are eachof the seven days equally likely? Suppose there seven urns (the days of the week), and we want to distribute 12 tickets ("balls") in the urns. The probability that all 12 tickets get put into two of the urns ("Tuesday" and "Thursday") is just

$$\frac{\binom{12=2-1}{2-1}}{\binom{12+7-1}{7-1}} = \frac{13}{18564} = \frac{1}{1428}.$$

(ii) Suppose you model this as choosing twelve separate times (tickets indistinguishable) which day of the week to give a ticket. Then you get $P = (\frac{2}{7})^{12}$.

(iii) Here is a what I consider the most realistic model: this is a Bernoulli trials model where the police are to decide every day with equal probability, whether or not to drive down the street where the professor parks (assuming he will get a ticket every time they drive by his illegally parked car at night). A longish version of this possible solution to this problem is on the web at `http://www.math.lsa.umich.edu/ dburns/tix03.pdf`. Please refer to that for the details of this possible solution.

Problem 3. A problem in genetics.

Heritable characters depend on special carriers, called genes. As you know, genes appear in pairs on paired chromosomes. In the simplest situation, each of the two genes of a pair can assume one of two forms, called an *allele*, which we will denote typically A and a. Then three different pairs can be formed, AA, Aa, aa. These are called the three possible *genotypes* associated to the pair of genes. Each pair of genes determines one heritable factor, but each observable characteristic in an organism may depend on many of these factors. For some very specific characteristics, the influence of one pair of genes is so strong as to be regarded as exclusive (so-called *monogenic* characteristics). Reproductive cells (*gametes*) are formed by a splitting process and have only one of the two genes of the individual from any given gene pair. In mating the pairs are reconstituted, but with only one element of the gene pair from each of the parents. Thus if the mother is an AA genotype and the father is an aa genotype (these are called *homozygotes*, having the same gene for each of the pair), then the offspring will be necessarily of type Aa (*heterozygote*). The genotypes of offspring therefore depend on a chance process. At every occasion, there is an even chance that one or the other of the two genes in a gene pair will be passed on to the offspring by a parent. In particular, the genotypes of n offspring will correspond to the result of n independent trials, each of which is equivalent to the flipping of two fair coins, one for each parent. Thus, if an Aa and an Aa mate, any offspring should be of the genotypes AA, Aa, aa, with respective probabilities 1/4, 1/2, 1/4.

There is a second chance process in all this, and we standardize this by assuming *random mating.* We define this as follows. Each descendent of one generation is to be regarded as the product of a random selection of parents, and each selection is independent. This is obviously an idealization.

Suppose that the three genotypes are distributed among males and females of the population with frequencies $u : 2v : w$, where $u + 2v + w = 1$. Call $u, 2v, w$ the *genotype* *frequencies.* Put $p = u + v$, $q = v + w$. These are the *gene frequencies* for A and a, respectively. Thus, under random mating conditions, the genotype frequencies for the offspring generation will

$$u_1 = p^2, v_1 = 2pq, w_1 = q^2.$$

Verify this claim. Also, check what are the genotype frequencies of the second generation of descendents, i.e., that they are given by the same distributions as above. What about the third and later generations? This was first observed by Hardy and Weinberg.

We shall return to genetics for some future exercises.

*** **Solution:** The assumption is that the gene frequencies are the same for males and females (see GP #2, however). *Random mating* says that the genotype of the first generation of offspring is like two flips of a biased coin (A with probabillity p, a with probability $(1-p)$).

Table of Genotypes

Aa	aa
AA	aA

Table of Probabilities

pq	q^2
p^2	qp

Hence,

$$u_1 = p^2, 2v_1 = 2pq, w_1 = q^2,$$

and

$$p_1 = u_1 + v_1 = p;\ q_1 = v_1 + w_1 = q.$$

Notice that now the same thing happens in all succeeding generations, though the zero-th generation could have many solutions. (See GP #2, prob. 3.)

Problem 4. The prisoner's dilemma.

Three prisoners are informed by their jailer that one of them has been chosen at random to be executed, and the other two are to be set free. Prisoner A asks the jailer to tell him privately one of his fellow prisoners who will be set free, claiming there would be no harm in divulging this, since he already knows that at least one will go free. The jailer refuses to answer this question, pointing out that if A knew which of his fellow prisoners were to be set free, then his own probability of being executed would rise from 1/3 to 1/2, because he would then be one of only two prisoners. What do you think of the jailer's reasoning? This is problem 41 of Chapter 3 of Ross. This is a good problem for discussion.

*****Solution:** This can be solved readily with Bayes's Theorem, fro example. Let A be the event that Prisoner A is executed, and similarly for B, C. Let b be the event that the jailer says "Prisoner B will not be executed!" and similarly for c. We want to show that

$$P(A) = P(A \mid b) = \frac{1}{3}.$$

Note that we don't have complete information about the mechanism: if it is, in fact, Prisoner A who will be executed, then the jailer has a *choice* of whether to say "Prisoner B will not be executed!" or the same thing for Prisoner C. We ASSUME that he will choose to say either of these wiuth equal probability in this cae, that is, that

$$P(b \mid A) = P(c \mid A) = \frac{1}{2}.$$

But then, by Bayes,

$$P(A \mid b) = \frac{P(b \mid A)P(A)}{P(b \mid A)P(A) + P(b \mid A^c)P(A^c)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{1}{3}.$$

One may also solve this one by drawing conditional probability "trees", as we did a couple of times in class. You still must, of course, state explicitly that you will ASSUME that

$$P(b \mid A) = P(c \mid A) = \frac{1}{2}.$$