Math 425, Section 7, Winter 2003: Introduction to Probability.

Second Group Problems: Solutions

1. The St. Petersburg Paradox.

A person tosses a fair coin until a tail appears for the first time. If the tail appears on the nth flip, the person wins 2^n dollars. Let $\mathbf{X}$ denote the random variable "the player's winnings". Show that $E(\mathbf{X}) = +\infty$. This problem is known as the St. Petersburg paradox.
(a) Would you be willing to pay $1 million to play this game once?
(b) Would you be willing to pay $1 million for each game if you could play for as long as you liked and only had to settle up when you stopped playing?

Solution: This is a loose question, and some of the answers depend on your outlook on things. Nevertheless, there are some probability observations which should help you decide, and the question is asking you to isolate those. So, to begin with, we should check that $E(\mathbf{X}) = +\infty$. $\mathbf{X}$ is a function of the geometric random variable, "the arrival of the first tail". The probability that the first tail arrives at the n-th flip is $2^{-n}, n = 1, 2, ...$ Since the value of $\mathbf{X}$ if the first tail is at flip n is 2^n, we find

$$E(\mathbb{X}) = \sum_{n=1}^{n=+\infty} 2^n \cdot 2^{-n} = \sum_{n=1}^{n=+\infty} 1 \cdot 1 = +\infty.$$

(a.) Most people said "no" here, arguing that the probability of recovering your $1,000,000 were very small, approximately 1 in a million, which is correct and reasonable.

(b.) A lot of people said "no" here, also. However, since the expectation is infinite, one should look closer. In fact, if you are allowed to pay up at the end and can quit whenever you wish, then this is virtually giving you a fortune. The only problem is that in order to realize this gain, you will have to wait a long time. The longer you play, the more favorable the game is (that is when you choose beforehand how long you will play it). People have worked what a fair entry price is, and it actually grows with the number of plays. That is, if you play for n games, you shouldn't pay just $\$1,000,000 \times n$, but rather $\$1,000,000 \times n \log_2(n)$.

This question only carried 4 points out of 20: I just wanted you to think about the problem.

2. Reliability and Lifetimes.

A certain manufacturer wishes to study the reliability of the light bulbs it makes. Experience suggests that, in everyday use, light bulbs fail when they are switched on and not while they are functioning. Hence the manufacturer decides to study reliability by counting the number of times a light bulb can be turned on.

Let X be the number of times a bulb can be switched on until it fails for the first time. (For example, $X = 10$ means that the bulb *failed* on the *tenth* try to switch it on.) Suppose we make the following assumptions:

(i) Each time the light bulb is switched on, it either works or fails, independently of whether it worked on previous trials.

(ii) On any trial, the probability that the light bulb fails when it is switched on is p.

(a) Construct a probability model for X (*i.e.*, determine a probability mass function and specify the possible values of X). Verify that your mass function satisfies the condition

$$\Sigma_x p_X(x) = 1.$$

(*Hint:* Recall that the sum of the geometric series $\Sigma_{k=0}^{\infty} r^k = 1/(1-r)$, for $r \in (0,1)$.)

(b) Without referring to the formula for p_X derived in part (a), describe the qualitative behavior which you think the graph of p_X should have. Then, using your formula for p_X, construct graphs for $p = \frac{1}{4}$ and $p = \frac{1}{2}$, and try to show that p_X behaves in the way you predicted. Which value of p do you think the manufacturer would prefer?

(c) Determine the probability that a given light bulb functions at least 100 times. For what values of p is this at least 95% of the time?

(d) Intuitively, what do you think the mean value of X should be? Prove your conjecture using the mass function p_X constructed in part (a).

(e) Specify the location of the mean on each of the graphs constructed in (b). What is the relationship of the mean to the median in these cases? Can you show this for all values of p?

3. Another problem in genetics: Sex Related Characters.

This is a continuation of the problem in the first group problem set. We said earlier that both genes and chromosomes appear together in pairs. There is one important exception to this. Recall that the *sex chromosomes*, denoted X and Y, are a pair of coupled chromosomes and they determine an individual's sex. All females are of genotype XX and all males are of genotype XY. Now many genes situated on the X-chromosome have no corresponding gene on the Y-chromosome which is simply significantly shorter than the X-chromosome. In particular, according to the elementary model in the earlier problem, the offspring of any mating pair should be 50% male and 50% female, the difference being whether the individual receives the Y- or X-chromosome from its father. Thus, in males, genes which are on the X-chromosome only appear singly. Typical examples of this are two sex-linked genes which cause colorblindedness and hemophilia. With respect to any such genes, females can have one of three different genotypes if there are two alleles at this locus, namely AA, Aa, and aa, while males can have only two, namely, A and a. Let's assume again *random mating* with genotype frequencies in the female population of $u, 2v, w$ for AA, Aa, aa, respectively. As before, set $p = u+v, q = v+w$. The frequencies of the two male genotypes A and a we will call p', q', respectively. The probability of a (first generation) female descendant to be of genotype AA, Aa or aa will be denoted $u_1, 2v_1, w_1$, while we denote the genotype frequencies of the (first generation) male descendants with ${p'}_1, {q'}_1$. Now a male descendant must receive his X chromosome from his mother, so

$${p'}_1 = p, \; {q'}_1 = q.$$

For the three female genotypes, we find frequencies, as earlier,

$$u_1 = pp', \; 2v_1 = pq' + qp', \; w_1 = qq'.$$

Hence,

$$p_1 = u_1 + v_1 = \frac{1}{2}(p + p'), \text{ and } q_1 = v_1 + w_1 = \frac{1}{2}(q + q').$$

(a) What do you discern about the difference between the frequencies of the A and a genes in male and female [populations? More specifically, compute the difference between the male and female frequencies of the A and a genes in the first filial generation compared to the parental generation.

(b) Does Hardy's law from the previous problem set hold here? We are ignoring effects of mutation and other forms of "genetic drift" on the population.

(c) Let p_n, q_n be the frequencies for A, a in the female population at the n-th filial generation, and p'_n, q'_n the corresponding frequencies for the male population in that same generation. Calculate inductive equations for these quantities.

(d) Let

$$\alpha = \frac{1}{3}(2p + p'),\ \beta = \frac{1}{3}(2q + q').$$

Show that, as $n \to \infty$,

$$p_n \to \alpha,\ p'_n \to \alpha,\ q_n \to \beta,\ q'_n \to \beta.$$

Does this convergence occur rapidly or slowly?

(e) For a recessive gene which is sex-linked (such as that for colorblindness), the population tends to have a proportion of the male population with the recessive gene, say a above, given by β. What is the corresponding frequency of the female genotype aa In particular, if one in 100 males is colorblind, what is the probability that a female will be colorblind?