Problems for Recitation 12 (5/1/98)

Problem 1. This is a real problem that came from a research project in computational biology last term. The CS PhD student hadn't taken 6.042 and so he needs our help. Can you solve the problem?

It is known that atoms occur in different isotopes. For example, Carbon atoms usually have 6 protons and 6 neutrons, but occasionally they have 7 neutrons. In fact 1.11% of all Carbon atoms have 7 neutrons and 98.89% have 6. From what we know, the appearance of the different isotopes seems to be a random event, unrelated to the functionality of the atom. We can tell which Carbon atoms have an extra neutron only by their atomic weight. The atomic weight of the generic Carbon atom is 12 and of its isotope is 13.

The research team is studying proteins composed of Carbon, Oxygen, Nitrogen and Hydrogen (the four building blocks of life). Their atomic weights and isotopes are listed below. The second column is the atomic weight and the third column is the frequency of the isotope.

Hydrogen	1	$P_{H,1}$
	2	$P_{H,2}$
Carbon	12	$P_{C,12}$
	13	$P_{C,13}$
Nitrogen	14	$P_{N,14}$
	15	$P_{N,15}$
Oxygen	15	$P_{O,15}$
	16	$P_{O,16}$
	17	$P_{O,17}$

The research team wants to write software that can predict the distribution of atomic weights for any protein. Describe a simple algorithm for computing the weight distribution of any protein that contains n_C carbon atoms, n_O oxygen atoms, n_N nitrogen atoms and n_H hydrogen atoms. [Hint: Use generating functions!]

Problem 2. Coupon Collector problem: We are trying to collect baseball cards. Assume there are n different cards. We can buy them one at a time. When we buy a card, it is equally likely that we get any one of the n cards, and purchases are mutually independent. We want to study how many cards we need to buy before we get one each of the n different kinds.

(a) Suppose we already have k different cards. What is the expected number of cards we need to buy to get a new card?

(b) Using the above, what is the expected number of cards we need to buy to get one of each kind?

(c) Suppose we buy $n \ln n + cn$ cards, where $c > 0$ is a real number. What is the probability that we don't have the first card?

(d) Suppose we buy $n \ln n + cn$ cards, what is an upper bound on the probability that we don't have all n cards?