CS/Math 240: Introduction to Discrete Mathematics	7/26/2007

Homework 6

Instructor: Jeff Kinne — TA: Mike Kowalczyk

This homework is due at the beginning of class on Thursday August 2, 2007. Mike will hold a review session at 11:45-12:45 on August 2 to discuss the solutions to these problems (that is after you have handed in the homework).

Problem 1

In this problem, we show that there exist strings that cannot be compressed by very much. Let x be some string of letters, and let C_x be a string which is source code for a program written in C. Now suppose that C_x is shorter than x, and that when C_x is executed it outputs x. Then C_x is said to be a *compression* of x, and we say that x can be *compressed* by $1 - \frac{|C_x|}{|x|}$. Note that in this model, we do not allow the program described by C_x to take any input.

We aim to show that there exist strings that cannot be compressed in this model by more than 1/2. We will use the probabilistic method to show that there must exist such strings (without having to actually demonstrate such a string).

Part a

Compute the total number of strings that are $2n$ letters long. For the sake of this problem, assume there are exactly 100 different letters that can be used (this includes, for example both capital and lower case, punctuation, white space, etc.).

Also give an upper bound for the number of different C programs of length at most n. For this, also assume that there are 100 different "letters" that can be used.

Part b

Use Part a) to show that there exists a string of length $2n$ that cannot be compressed by more than 1/2. What can you say about the fraction of strings of length $2n$ that can be compressed by at least 1/2?

Problem 2

In this problem, we consider the following experiment. We first roll 1 die. Let X be the value that comes up on that die. We then roll X more dice. Let $Y_1, Y_2, ..., Y_X$ be the values of these dice. In this problem, we look at the random variable $Y = Y_1 + Y_2 + ... + Y_X$ that is the sum of the second set of dice.

For example, suppose we roll a 1, meaning $X = 1$. We then roll 1 more die. If we roll a 6, then $Y_1 = 6$, and $Y = Y_1 = 6$.

Suppose we initially roll a 2, meaning $X = 2$. We then roll 2 more dice. If we roll a 3 and a 4, then $Y_1 = 3$, $Y_2 = 4$, and $Y = Y_1 + Y_2 = 3 + 4 = 7$.

Compute $E[Y]$. Make sure to justify (prove) your answer.
Hint: It may be helpful to first show that $E[Y] = \frac{1}{6} \cdot (E[Y_1] + E[Y_1 + Y_2] + ... + E[Y_1 + ... + Y_6])$. *Try using both of our "definitions" of expectation to do this.*

Problem 3

For this problem, we analyze the following randomized algorithm to solve the unordered search problem.
Rand-Search
Input: $\{x_1, ..., x_m\}, x$ with $m \geq 1$
Output: "yes" iff x is in the list

Set $L = \{1, 2, ..., m\}$.
for $i = 1$ **to** m
    Pick k uniformly at random from L.
    if $x_k = x$ **then** return "yes".
            else Remove k from L.
return "no"

First we point out that if x is not in the list, this algorithm considers the same number of positions in the list as linear search - they both will end up searching the entire list. So, for this problem, we assume that there is exactly 1 copy of x in the list. For this situation, we want to analyze the running time of *Rand-Search*.

Let Y be the random variable that corresponds to the number of positions that *Rand-Search* tries before finding x (including the successful attempt), in other words, Y is the largest value that i attains when we assume that x appears in the list exactly once.

Part a) For each value of $i = 1, 2, ..., m$, compute $\Pr[Y = i]$.

Part b) Compute $E[Y]$.

Part c) Compute $Var[Y]$.

Part d) Use Markov's inequality to bound the probability that Y is greater than $\frac{4m}{5}$.

Part e) Use Chebyshev's inequality to get a formula that bounds the probability that Y is greater than $\frac{4m}{5}$. What is the value of this formula for $m = 10$? For $m = 1000$? Is this estimate better or worse than the estimate computed in Part d)?

Problem 4

Recall that for independent random variables X_1 and X_2, $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$. Give an example of random variables Y_1 and Y_2 that are not independent and $Var(Y_1 + Y_2) \neq Var(Y_1) + Var(Y_2)$.