Discrete Structures	**Problem Set 9**
CS 280 Spring 2005	**Due Wednesday, April 20, 2005**

Reading: Rosen edition 5: Sections 5.3 (or edition 4: Sections 4.5), and Section 13.6 of the Kleinberg-Tardos handout.

You may also want to review number theory and modular arithmetic from Chapter 2.

For this problem set we use the notation "(p mod q)" to denote that the remainder r that results from dividing p by q. We will assume for this Thus, for a natural number n, the quantity "(n mod 2)" is 0 if n is even, and 1 if n is odd."

(1) Suppose we are given a set of n variables $x_1, x_2, \ldots, x_n$, each of which can take one of the values in the set $\{0, 1\}$. We are also given a set of k equations; the r^{th} equation has the form

$$((x_i + x_j) \bmod 2) = b_r$$

for some choice of two distinct variables x_i, x_j, and for some value b_r that is either 0 or 1. Thus, each equation specifies whether the sum of two variables is even or odd.

Consider the problem of finding an assignment of values to variables that maximizes the number of equations that are satisfied (i.e. in which equality actually holds).

For example, suppose we are given the equations

$$((x_1 + x_2) \bmod 2) = 0$$

$$((x_1 + x_3) \bmod 2) = 0$$

$$((x_2 + x_4) \bmod 2) = 1$$

$$((x_3 + x_4) \bmod 2) = 0$$

over the four variables $x_1, \ldots, x_4$. Then it's possible to show that no assignment of values to variables will satisfy all equations simultaneously, but setting all variables equal to 0 satisfies three of the four equations.

(a) Use the probabilistic method to prove that for any set of n equations, there is a way to assign values to the variables that satisfies at least half of them. Recall that the method suggests that you set the variables independently at random, and prove that the probability that at least half of the n equations are satisfied is not 0.

(b) Suppose we drop the condition that each equation must have exactly two variables; in other words, now each equation simply specifies that the sum of an arbitrary subset of the variables, mod 2, is equal to a particular value b_r. We assume that each equation involves at least 2 variables. Does the claim in part (a) remain true? Prove it or provide a counter example.

(2) We considered two class of hash function. Both hash to a table of size p for some prime p. In the case of the simple hashing we assumed that the universe U is a set of large

numbers, and for $x \in U$ we define $f(x) = (x \bmod p)$. This is only one particular function, and hence we cannot expect to make any probabilistic statements about it. A related class of function is $\mathcal{H}_{simple} = \{f_\alpha | 0 \leq \alpha < p\}$, where $f_\alpha(x) = (\alpha x \bmod p)$.

In class we have defined a universal class of hash-functions that assume that the universe U' is of the form $(x_1, x_2, \ldots, x_r)$ for some integer r where $0 \leq x_i < p$ for all i. We defined a has function $h_a(x) = ((\sum_i x_i a_i \bmod))p$ where $a = (a_1, a_2, \ldots, a_r)$ is a vector of integers, with $0 \leq a_i < p$ for all i, and we have shown that for two vectors $x, x' \in U'$ the probability $Pr(h_a(x) = h_a(x')) = 1/p$ if the coordinates of a are chosen independently, uniformly at random.

(a) Is it true that for every integer $0 \leq k < p$ and every integer x, the probability $P(f_\alpha(x) = k) = 1/p$ if α is randomly chosen $0 \leq \alpha < p$ with each value equally likely? Prove or provide a counter example with a proof.

(b) Is it true that for every integer $0 \leq k < p$ and every two integers x and y, the probability $P(f_\alpha(x) = f_\alpha(y)) = 1/p$ if α is chosen randomly with all values $0 \leq \alpha < p$ equally likely? Prove or provide a counter example with a proof.

(c) Is it true that for every integer $0 \leq k < p$ and every vector $x \in U'$, the probability $P(h_a(x) = k) = 1/p$ if a is a vector whose coordinates are chosen randomly and independently with all values $0 \leq a_i < p$ equally likely? Prove or provide a counter example with a proof.

(d) Recall that we have shown in class (and is also in the notes) that for every two vectors $x, x' \in U'$ the probability $Pr(h_a(x) = h_a(x')) = 1/p$ if the coordinates of a are chosen independently, uniformly at random. Let $A(x, k)$ be the random event that $h_a(x) = k$. Consider two different elements $x, x' \in U$ is $A(x, k)$ and $A(x', k)$ independent? Prove or provide a counter example with a proof. Note that for both $A(x, k)$ and $A(x', k)$ the choices of x, x' and k are fixed, and the randomness is over the different choices of the coordinates a_i.

(e) Explain briefly how your answers to (a), (b), (c) and (d) effect which of the two classes of hash-functions may be better to use, if avoiding collisions is important.

(3) We have shown in class Markov's inequality that states that for every random non-negative variable X with expectation $E(X) = \mu$ and any $k > 1$ we have the following bound on the probability that X exceeds $k\mu$.

$$P(X \geq k\mu) \leq \frac{1}{k}.$$

(Note that in class we used strict inequalities in place of both inequalities. Markov inequality is valid in both forms.) Show that this bound is the strongest one can prove of this form. More formally, show that for any $\mu > 0$ and any $k > 1$, there is a nonnegative random variable X with $E(X) = \mu$, and $P(X \geq k\mu) = \frac{1}{k}$.

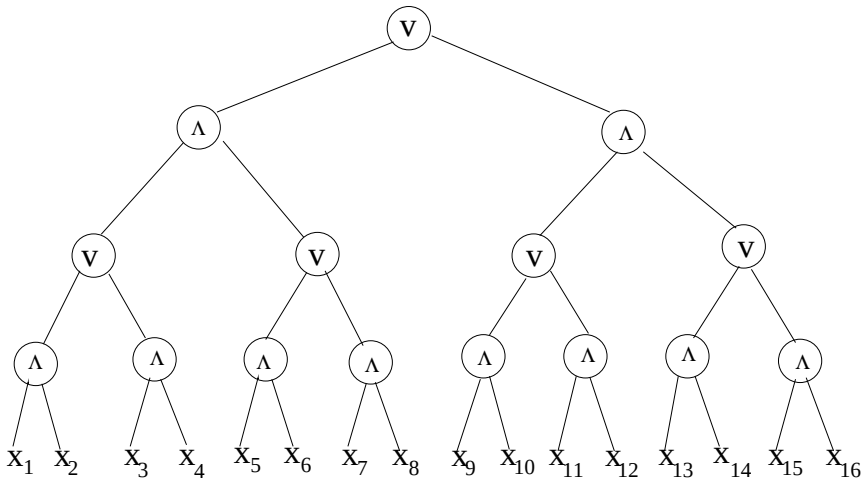

Figure 1: Circuit for Problem 6.

(4) Let X be a random variable, and Y and independent copy of the same random variable. (For example, if X is the outcome of a dice through, then Y is the outcome of an independent other dice though.) Let $\mu = E(X)$, and of course $\mu = E(Y)$ also. We know from linearity of expectation that if we define $Z = (X+Y)/2$ then $E(Z) = \mu$ also. What are possible relations between $\sigma(X)$ and $\sigma(Z)$ (can $\sigma(Z)$ smaller, bigger, or the same as $\sigma(X)$)? Of each option, either give an example, or show that it cannot happen. (You may want to use the fact that $Var(X+Y) = Var(X) + Var(Y)$ which we will prove on Friday.)

(5) We define co-variance of two random variables X and Y with expectations μ_X and μ_Y as $E((X-\mu_X)(Y-\mu_Y))$. Note that the co-variance of X with itself is exactly the variance of X. Show that the co-variance of two independent random variables X and Y is 0. You may want to wait till Friday's class before attempting this problem.

(6 optional) Consider a circuit built from $2n$ levels of alternating $\vee$ and $\wedge$ gates, and 4^n different variables. The odd levels have $\wedge$ gates and the even levels $\vee$ gates, with each gate taking the $\wedge$ or $\vee$ of two gates at the previous level. The circuit for $n = 2$ with $2n = 4$ levels, and $4^n = 16$ gates is shown in Figure 1.

To evaluate the circuit without randomization, we would need to look at all 4^n variables in the worst case. However, we claim that the following idea helps save on this in expectation. Start from the top of the formula. When you are at a gate, select one of the two sides at random, both sides equally likely, and evaluate that side first (using the same method recursively). Now evaluate the other side only of needed.

For example, if the current game is an $\vee$ gate, and the side evaluated comes out true, than we know that the "or" gate's value is true, and hence no need to evaluate the other

side. Unfortunately, with an $\vee$ gate, if the first evaluation comes out false, we still need to evaluate the other side.

If the current game is an $\wedge$ gate, and the side evaluated comes out true, than we still need to compute the other side. However, if the value is false, then we know that the "and" gate's value is false, and hence no need to evaluate the other side.

Show that the expected number of valiables that this method looks at with this method is at most 3^n.