15-859(D) Randomized Algorithms

Homework # 3 **due: October 20, 1998**

(Do problem 5 before Thursday Oct 15th's class, since we'll be using the lemma you prove.)

Exercises:

1. The NP-hard *set cover problem* is: given subsets $S_1, S_2, \ldots, S_m$ of $V = \{1, 2, \ldots, n\}$, cover V with as few of the S_i's as possible.

   Suppose we are given an instance of set-cover whose optimal solution has size p (it uses p of the sets S_i). Show that the simple greedy algorithm ("pick the largest subset; remove the nodes it covers from V; repeat.") has the property that after finding p subsets, it has covered at least a $1 - 1/e$ fraction of V.

2. Deterministically amplifying the success probability of a randomized algorithm via hashing (Chor and Goldreich): Let H be a pairwise independent (strongly 2-universal) family of hash functions from $\{0,1\}^n$ to $\{0,1\}^n$, with the property that it takes only $O(n)$ random bits to select h from H at random. (For example, the "$(ax + b \bmod p) \bmod N$" generator will do). Suppose that we have a **BPP** algorithm that uses n random bits and has error probability $\leq 1/4$. To reduce our error, we can run the algorithm k times on the same input and take a majority vote. To reduce error without increasing the number of random bits used, instead of truly independent runs, we will use $h(i)$ as the "random input" for the ith trial. Show that with this approach, the probability the majority vote is wrong is $\leq 3/k$. Hint: use Chebyshev's inequality.

Problems:

3. Consider the setup in exercise 1 (an instance of set-cover whose optimal solution has size p). However, suppose now that V contains some unknown ("hidden") set H of "important points", and what we would like is to find p subsets that cover a *constant fraction of* H, rather than just cover a constant fraction of V. Describe a randomized algorithm to do this. It should find p subsets, and have the property that the expected number of points in H covered is at least $|H|(1 - 1/e)$.

4. Problem 6.11 in the book: show that if G is *regular* (all nodes have the same degree) then the cover time of G is bounded by $O(n^2 \log n)$. Thus, for regular graphs of degree greater than $\log n$, we get a better bound than our result for general graphs.

   Hint: Prove an upper bound on the diameter, as a function of the degree of the graph. Then use Theorem 6.9.

5. Hash mixing lemma (From Rudich). Let H be a pairwise independent hash family from $\{0,1\}^n \longrightarrow \{0,1\}^n$. Let $\epsilon = 2^{-n/3}$. Show that for all $A, B \subset \{0,1\}^n$, and for all but an ϵ fraction of $h \in H$,

   $$|Pr_{y \in \{0,1\}^n}[y \in A \wedge h(y) \in B] - Pr_{y,z \in \{0,1\}^n}[y \in A \wedge z \in B]| \leq \epsilon$$

   Hint: Define $Z_h^y = 1$ if $h(y) \in B$, 0 otherwise. Let $Z_h = \sum_{y \in A} Z_h^y$. h is "bad" if Z_h deviates too much from what it would be if h were a random function, namely, $|A| \cdot |B|/2^n$. Use Chebyshev to say that there are only a small fraction of "bad" h's (fewer than ϵ).