CS 787: Advanced Algorithms	Homework 2
Out: 03/08/13	**Due: 04/05/13**

Guidelines

Same as for HW1.

Exercises

1. Extend the Chernoff bound discussed in class to the case of arbitrary random variables $X_i \in [0,1]$ following the approach below.
   (a) Show that $f(x) = e^{tx}$ is a convex function.
   (b) Show that if C is a random variable taking values in $[0,1]$, and B is a $\{0,1\}$ random variable with $E[C] = E[B]$, then for any convex f, $E[f(C)] \leq E[f(B)]$.
   (c) Use these to reprove the Chernoff bounds for sums of independent random variables over $[0,1]$.

2. Recall that a vertex cover of a graph $G = (V, E)$ is a set of vertices $C \subset V$ such that each edge has at least one endpoint in C. Finding the vertex cover of the smallest cardinality is NP-complete.
   (a) Consider the following algorithm for Vertex Cover:
      i. Start with $C \leftarrow \emptyset$.
      ii. Pick an edge (u, v) such that $(u, v) \cap C = \emptyset$. Add an arbitrary endpoint to C.
      iii. If C is a vertex cover, halt, else go to Step (ii).

      Give an instance on which this algorithm may return a set which is $\Omega(n)$ times larger than the smallest vertex cover.
   (b) Now suppose we randomize the algorithm thus: when we pick an edge (u, v), we flip an unbiased coin to decide which endpoint to add to C. If k is the size of a smallest vertex cover, show that $E[|C|] \leq 2k$. (Hint: Consider any vertex v in the optimal vertex cover, and let $N(v)$ be the set of the neighbors of v as well as v itself. What is $E[|C \cap N(v)|]$?)
   (c) Suppose each vertex v had a weight $w(v)$, and the objective was to pick a set of smallest weight. Give an example to show that the above algorithms do not work for this problem. How would you alter the above randomized algorithm to still obtain a factor of 2 approximation? (Hint: vertices with high weight should be picked with lower probability and vertices with low weight with higher probability.)

3. Given a graph $G = (V, E)$, a dominating set $D \subseteq V$ is one where each vertex $v \in V$ is either in D or has a neighbor in D.

   Show that any graph with minimum degree d has a dominating set of size at most $O(\frac{n}{d} \log n)$. Can you prove the existence of a dominating set of size $n\frac{1+ln(1+d)}{(1+d)}$?

4. Problem 4.18 from the textbook.

Problems

1. Problem 4.25 from the textbook.

2. In the *Unique Set Cover* problem, we are given a set of n elements and a collection of m subsets of the elements. Our goal is to pick out a number of subsets so as to maximize the number of uniquely covered elements, those that are contained in exactly one of the picked subsets. (Note that the cost or number of subsets picked is not important).

   (a) Consider the following naïve algorithm—for some number $p < 1$, the algorithm picks each subset independently with probability p. Assuming that every element is contained in exactly F subsets, compute the expected number of elements uniquely covered. For what value of p is this expectation maximized?

   (b) Assuming that each element is contained in at most F subsets and at least $F/2$ subsets, give a constant factor approximation to unique set cover using the algorithm from part (a).

   (c) Extend the algorithm from part (b) to obtain an $O(\log m)$ approximation in general (without assumptions on the frequency of any element). (Hint: Try reducing this problem to the one in part (b) for an appropriate choice of F.)

   (d) Can you improve the approximation from part (c) to a factor of $O(\log n)$?

3. Consider coloring a set U of size n using two colors red and blue. The *discrepancy* of a set $S \subseteq U$ is the absolute difference of the number of red elements and the number of blue elements in S. Formally, we can describe the coloring of U as a function χ mapping elements in U to $\{1, -1\}$. The discrepancy of S can then be written as $\mathrm{disc}_\chi(S) = |\sum_{x \in S} \chi(x)|$.

   For a collection of sets $\mathcal{F}$ we define the discrepance of $\mathcal{F}$ as

   $$\mathrm{disc}_\chi(\mathcal{F}) = \max_{S \in \mathcal{F}} \mathrm{disc}_\chi(S) = |\sum_{x \in S} \chi(x)|.$$

   Given the family $\mathcal{F}$, our goal is to find a coloring with low discrepancy, that is, we want all of the sets in $\mathcal{F}$ to have a roughly balanced coloring. In this problem we will give bounds for how low the discrepancy can be.

   (a) Suppose that χ is a random coloring. That is, every element of U is mapped to 1 or -1 independently with equal probability. Prove that for any family $\mathcal{F}$ containing n sets, the discrepancy is $O(\sqrt{n \log n})$ with high probability (i.e. $1 - 1/n$).

   (b) We will now prove a lower bound on discrepancy. Fix an arbitrary assignment $\chi : U \to [1, -1]$. Note that this assignment is not random. Let A be a random subset of U, that is, each element of U is included in A independently with probability $1/2$. Show that for some constant c,

   $$\Pr\left[\mathrm{disc}_\chi(A) > \frac{\sqrt{n}}{c}\right] > \frac{1}{2}.$$

   (c) Now prove that there exists a family F of size n for which *all* assignments $\chi : U \to [1, -1]$ have discrepancy $> \frac{\sqrt{n}}{c}$. (Hint: what if you pick n sets independently as in part (b)?)