Randomized Algorithms (Blum/Gupta) **Homework 3**
Date: Feb 9, 2011 **Due:** Feb 23, 2011

Groundrules

- Homeworks will generally consist of *exercises*, easier problems designed to give you practice, and *problems*, that may be harder, trickier, and/or somewhat open-ended. You should do the exercises by yourself, but you may work with a friend on the harder problems if you want. One exception: no fair working with someone who has already figured out (or already knows) the answer. If you work with a friend, then write down who you are working with.
- If you've seen a problem before (sometimes we'll give problems that are "famous"), then say that in your solution (it won't affect your score, we just want to know). Also, if you use any sources other than the textbook, write that down too (it's fine to look up a complicated sum or inequality or whatever, but don't look up an entire solution).

Reading: Listed next to the lectures on the webpage.

Exercises

1. **(The Bias towards Satisfaction.)** In this problem we'll give yet another approximation algorithm for the MAX-SAT problem.
   (a) Suppose the MAX-SAT instance is such that all the length-1 clauses (also called "unit clauses") only contain variables that are un-negated. Now suppose we set each variable to 1 with probability $p \geq 1/2$, and 0 otherwise. Show that each clause is satisfied with probability at least $\min(p, 1-p^2)$. Hence show that for a suitable value of p, the expected number of clauses satisfied is at least $(\sqrt{5}-1)m/2 \approx 0.618m$, where m is the total number of clauses.
   (b) Now consider a formula without any restriction on the unit clauses. Clearly, we cannot hope to satisfy more than 50% of the clauses, since the formula might be $\phi = (x) \wedge (\neg x)$. However, show an algorithm that satisfies at least $c\,OPT$ of the clauses, where OPT is the maximum number of clauses that can be satisfied by any assignment, and $c = (\sqrt{5}-1)/2$.
2. **(How Many Unoccupied Bins?)** Suppose we throw n balls into n bins uniformly at random, and let X_i be the indicator r.v. for bin i having no balls in it (i.e., for it being unoccupied). If $E[X_i] = \mu$, and if $X = \sum_i X_i$ is the number of unoccupied bins, then $E[X] = n\mu$. Show that $E[X] \approx n/e$. Also that, for any constant $\epsilon > 0$, the tail probability $\Pr[X \geq (1+\epsilon)n/e] \leq 1/poly(n)$. Be careful: the X_i's are not independent! You are allowed to use results proven in previous assignments.
3. **(Lovasz Local Lemma and Colorings.)** Consider an undirected graph $G = (V, E)$, where each vertex v has a list $S(v)$ of allowed colors. A list-coloring χ of G assigns each vertex $v \in V$ a color from its list $S(v)$. A *proper* list-coloring is one that ensures that all edges are bichromatic.

   Suppose each vertex has a list of size $10k$. Moreover, for each $v \in V$ and $c \in S(v)$, there are at most k neighbors u of v that contain c in their color sets $S(u)$. (There is no bound on the degree of the underlying graph, though.) Show that there exists a proper list-coloring of G with these parameters.

   Hint: For each edge $e = \{u, v\}$ and color $c \in S(u) \cap S(v)$, let $\mathcal{B}_{e,c}$ to be the event that $\chi(u) = \chi(v) = c$.

Problems

1. **(Nearly Orthogonal Vectors.)** Call two vectors near-orthogonal if their inner product has small absolute value compared to the product of their lengths; in this problem we will show that while there are at most d orthogonal vectors in $\mathbb{R}^d$, there can be exponentially more near-orthogonal vectors.

Given a parameter $\epsilon > 0$, two vectors $\vec{x}, \vec{y} \in \mathbb{R}^d$ are called ϵ-orthogonal if

$$|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon \cdot \|\vec{u}\| \cdot \|\vec{v}\|.$$

Show that there exist $N = \exp(\Omega(\epsilon^2 d))$ unit vectors in $\{-1, 1\}^d$ that are mutually ϵ-orthogonal. (Hint: consider a random such set.)

Note: if you use a Chernoff bound other than the ones in the handout, you should prove it.

2. **(Coloring Sets.)** You are given a universe U of n elements and a collection of m sets over these elements: $S_1, S_2, \ldots, S_m$. Moreover, each of these sets contains *at least* k elements.

   Given a coloring of the elements in the universe with two colors red and blue, a set is "satisfied" if it is bichromatic—i.e., it contains both a red and a blue element. You are interested in colorings to maximize the number of satisfied sets.

   (a) Show that if $m < 2^{k-1}$, then there exists a coloring that satisfies all the sets. Give a (deterministic) polynomial time algorithm to find such a coloring.

   (b) Suppose you know that each set only intersects at most d many sets in this collection including itself. Prove that if $d < 2^{k-1}/e$, then there exists a coloring which satisfies all sets. Give an algorithm to find such a coloring if $d < 2^{k-4}$ that runs in expected polynomial time.

   (c) Consider the following algorithm:

   > *Tentatively color all elements red. Take a uniformly random ordering π of all the elements, and consider the elements in this order. If the current element being considered is the first element from some set S_i, recolor it blue.*

   Note that no set can be eventually be all red, so we need to bound the probability that some set has all its elements turned blue.

   i. Define the event $\mathcal{E}_{ij}$ ("S_i blames S_j") if the last (according to π) element of S_i is the first element of S_j (also according to π). Show that if S_i is all blue, then S_i blames someone. Hence, bound the expected number of all-blue sets, and the probability that the algorithm outputs a bad coloring, by $\sum_{i,j \in [m]} \Pr[\mathcal{E}_{ij}]$.

   ii. Calculate $\Pr[\mathcal{E}_{ij}]$.

   iii. Use this to show that if $m < ck^{1/4}2^k$ for a suitably small constant c, then the expected number of all-blue sets is strictly less than one. Hence infer the existence of a coloring that satisfies all the sets.

   Note: The best currently known argument shows that $m < c'\sqrt{k}\, 2^k$ also suffices, but requires more work. It is known that there are collections with $m > c''\, k^2 2^k$ for which there is no good two-coloring. Open question: what is the right answer?

3. **(Offline Cuckoos.)** In the two-choices lecture, we saw a random graph based proof that the better-of-two-choices algorithm for $r = 512n$ bins and n balls gives a load that is $O(\log \log n)$.

   (a) Show that if someone showed you the random choices of the n balls, you could find (in hindsight) find the allocation of each ball to one of its chosen bins, and which minimizes the max load in polynomial time.

   (In other words, you'd find which of their two choices the balls *could have* chosen to minimize the max bin load: of course, this uses all the choices and hence is not implementable as an online strategy.)

   (b) Show that this minimum max load is at most 3, with high probability.

   Just to put this in perspective: in the cuckoo hashing lecture, Rasmus showed that for $r \geq 4n$, the insertion procedure maintains a maximum load of one *throughout the run of the algorithm whp, by moving an expected constant number of keys at every step. So we're proving a much weaker guarantee—firstly, the load is a constant instead of* 1*; secondly, we're giving a guarantee only for the max-load at the end, and not a simple strategy that maintains a low max-load at all times.*