15-859(M): Randomized Algorithms **Homework 3**
Date: Oct 18, 2004 **Due:** Nov 01, 2004

Ground Rules

The homework consists of a few exercises followed by some questions. The exercises will not be graded, and are given to help you better understand the course material. You are allowed to work in small groups, but must turn in solutions individually. Please let us know, for each question, if you worked in a group, or if you have seen the question before.

Exercises

1. **(Power of d choices.)** Show that if we were to randomly throw n balls into n bins by choosing d bins and placing the ball in the least loaded one, then the maximum load is

$$\frac{\ln \ln n}{\ln d} + O(1) \qquad \textbf{whp} \tag{1}$$

(The proof in Lecture 10 considered the case of $d = 2$.)

2. **(Reducing the error in BPP.)** Recall that the definition of a language L to be in BPP (from Lecture 2) required the probability of error of a randomized algorithm $\mathcal{A}$ for L be at most $1/4$. Show that repeating the algorithm t times and outputing the majority answer can reduce the error probability to at most $1/2^{O(t)}$. Can you prove this using Chernoff bounds? Without using Chernoff bounds?

3. **(Sample Complexity.)** Suppose the random variable X takes values in the interval $[a, b]$, and the mean $E[X] = \mu$. How many samples do you need to so that the mean lies in the range $[\mu - \epsilon, \mu + \epsilon]$ with probability at least $3/4$?

Questions

1. **(Dominating Sets.)** 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 δ 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

$$n \frac{1 + \ln(1 + \delta)}{(1 + \delta)} \tag{2}$$

2. **(Non-discrete Chernoff.)** Extend the Chernoff bound given in Lecture 9 to the case of arbitrary random variables $X_i \in [0, 1]$.

   (a) Show that $f(x) = e^{tx}$ is a convex function.

   (b) If C is a r.v. in $[0, 1]$, and B is a Bernoulli $\{0, 1\}$ r.v. 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 independent random variables over $[0, 1]$.

3. **(More Balls and Bins.)** Let us analyse another occupancy problem which involves multiple rounds, but here we remove balls instead of bins. (This can be thought of as analysing a system for contention resolution.)

We start off by throwing n balls into n bins in the first round. After the round $i \geq 1$, we remove every ball that occupied a bin by itself in round i, and in the following round $i+1$, we throw the remaining balls into the n bins. (One can imagine the lonely balls getting service, whereas none of the colliding balls receive service.) The process ends when there are no more balls left.

Show that the number of rounds for which this process runs is at most $c \log \log n$ **whp**.

4. **(Randomized Lower Bounds.)** Consider a set U: any red-blue coloring of U can be seen as an assignment $\chi : U \to \{1, -1\}$. The *discrepancy* of a set $S \subseteq U$ is $\mathsf{disc}_\chi(S) = |\sum_{x \in S} \chi(x)|$. The discrepancy of a set system $\mathcal{F}$ with n sets is at

$$\mathsf{disc}_\chi(\mathcal{F}) = \max_{S_i \in \mathcal{F}} \mathsf{disc}_\chi(S_i) = \max_{S_i \in \mathcal{F}} \big| \sum_{x \in S_i} \chi(x) \big| \tag{3}$$

In Lecture 9, we saw that a random ± 1 coloring for a family $\mathcal{F}$ of n subsets of U with $|U| = n$ gives a discrepancy of $f(n) = O(\sqrt{n \log n})$. We will now use a probabilistic argument to give a lower bound: there exists a family $\mathcal{F}$ of $n+1$ subsets of U for which $\mathsf{disc}(\mathcal{F})$ is at least $g(n) = \frac{\sqrt{n}}{c}$.

   (a) Fix some assignment $\chi : U \to \pm 1$. If we pick a random subset $A \subseteq U$ by including each element of U in A independently with prob $\frac{1}{2}$. Show that for some c and this fixed assignment χ:

$$\Pr_A[\mathsf{disc}_\chi(A) > \frac{\sqrt{n}}{c}] > 1/2. \tag{4}$$

   (b) Now prove that there *exists* a family $\mathcal{F}$ for which *all* assignments $\chi : U \to \pm 1$ have discrepancy $> \frac{\sqrt{n}}{c}$. (Hint: what if you pick n sets independently as above?)

5. **(Expanders.)** A (d, c, α)-expander is a graph $G = (V, E)$ where each node has degree at most d, and every subset $S \subseteq V$ with at most cn nodes has $|N(S)| \geq \alpha |S|$. Recall that $N(S)$ is the set of neighbors of S, and may include some nodes in S but not others.

   Starting with a set V of n nodes, add a random matching between the vertices thus: (a) choose a random permutation $v_1, v_2, \ldots, v_n$ of the nodes, and (b) add the edges (i, v_i) for all i. (We may have parallel edges and self loops – if $v_i = i$ – in this graph.) Perform this process $d = 600$ times.

   Prove that $G = (V, E)$ is a $(2d, \frac{7}{20}, \frac{3}{2})$-expander with probability at least $\frac{1}{2}$.