6.895 Randomness and Computation	February 11, 2008

Homework 1

Lecturer: Ronitt Rubinfeld — *Due Date: February 25,2008*

Homework guidelines: You may work with other students, as long as (1) they have not yet solved the problem, (2) you write down the names of all other students with which you discussed the problem, and (3) you write up the solution on your own. No points will be deducted, no matter how many people you talk to, as long as you are honest. If you already knew the answer to one of the problems (call these "famous" problems), then let me know that in your solution writeup – it will not affect your score, but will help me in the future. It's ok to look up famous sums and inequalities that help you to solve the problem, but don't look up an entire solution.

The following problems are to help you understand the upcoming lectures. Please make sure you can do them, and if you have any questions, please talk to the course staff within a week. (Though, hopefully, some of them will be fun to think about for a while). Do not turn them in.

1. Let f be any linear function over a finite group G, that is, such that $\forall x, y \in G, f(x) + f(y) = f(x+y)$. You know nothing about f (except that it is linear). However, you are given a program P that computes f correctly on 7/8 of the inputs in G. (Again you know nothing about which inputs are correct or incorrect – just a bound on the total number of incorrect values). Assume that P runs in time T_G. Give a randomized algorithm which on input any $x \in G$ and security parameter β, outputs the correct value $f(x)$ with probability $1-\beta$. The algorithm should run in time $O(T_G \cdot \log 1/\beta)$.

2. (This is the "Von Neumann trick", which you don't really need for upcoming lectures, but it's cute). Given a coin with probability p of getting "heads", give a procedure for simulating one toss of a fair coin ($p = 1/2$). The procedure should run in expected time that is polynomial in $\frac{1}{p} + \frac{1}{1-p}$.

3. DO THIS BY WEDNESDAY 2/13. (This one isolates the essential reasoning that we will use in each of our Fourier analytic arguments). You are given $n \times n$ matrices A, B, C whose elements are from $\mathcal{Z}_2$ (integers mod 2). Show a (randomized) algorithm running in $O(n^2)$ time which verifies $A \cdot B = C$. The algorithm should always output "pass" if $A \cdot B = C$ and should output "fail" with probability at least 3/4 if $A \cdot B \neq C$. Assume the field operations $+, \times, -$ can be done in $O(1)$ steps.

The following problems are to be turned in. TURN YOUR SOLUTION IN TO EACH PROBLEM ON A SEPARATE PIECE OF PAPER WITH YOUR NAME ON EACH ONE.

1. You are given an approximation scheme $\mathcal{A}$ for f such that $Pr[\frac{f(x)}{1+\epsilon} \leq \mathcal{A}(x) \leq f(x)(1+\epsilon)] \geq 3/4$, and $\mathcal{A}$ runs in time polynomial in $1/\epsilon, |x|$. Construct an approximation scheme $\mathcal{B}$ for f such that $Pr[\frac{f(x)}{1+\epsilon} \leq \mathcal{B}(x) \leq f(x)(1+\epsilon)] \geq 1-\delta$, and $\mathcal{B}$ runs in time polynomial in $\frac{1}{\epsilon}, |x|, \log \frac{1}{\delta}$.

2. A 3-SAT formula takes the "and" of a set of clauses, where each clause takes the "or" of a set of literals (each literal is a variable, or the negation of a variable). Show that

for any 3-SAT formula in which every clause contains literals corresponding to 3 distinct variables, there is an assignment that satisfies at least 7/8 of the clauses.

3. Denote the complete graph on n nodes by K_n. Let $R(t)$ be the minimal n such that for any two-coloring of the edges of K_n, there is a subset of the vertices of K_n, of size t, such that all edges between vertices in this subset are the same color.

   Show that if $\binom{m}{t}2^{1-\binom{t}{2}} < 1$ then $R(t) > m$. (i.e., show that if $\binom{m}{t}2^{1-\binom{t}{2}} < 1$, then there is a coloring such that there is no subset of vertices of size t such that the edges joining these vertices are all one color).

4. To actually find a perfect matching, we can reuse the algorithm given in class as a subroutine by removing one edge at a time from the graph (together with both endpoints) and testing whether the resulting graph still has a perfect matching. If so, include the edge in the matching, otherwise, remove the edge (but not the endpoints) and try another edge. After at most a number of stages equal to the number of edges in the original graph, we have constructed a matching. However, this method is inherently sequential. We will now try to give a good parallel algorithm. Let's try to use the algorithm from class as a subroutine, but find a way to make all the subroutine calls at the same time.

   - First, let's prove a key lemma: Let $S_1, \ldots, S_k$ be subsets of a set S such that $|S| = m$. For all $i \neq j$, $S_i \neq S_j$. Let each element $x \in S$ have a weight w_x chosen independently and uniformly at random from the set $\{0, 1, \ldots, 2m-1\}$. Then show that

     $$Pr[\exists \text{ a unique set } S_i \text{ of minimum weight}] \geq \frac{1}{2}$$

     A hint is given on the course website.
   - Let B be the matrix formed by replacing x_{ij} with $2^{w_{ij}}$ in the Frobenius matrix. Use the lemma above to show that the following algorithm, which can be implemented to run in polylog time on polynomially many processors, outputs a perfect matching:
     - Calculate w, the largest power of 2 that divides $det(B)$
     - For each edge (i, j) in parallel do:
       * compute $t_{ij} = det(B_{ij})\frac{2^{w_{ij}}}{2^w}$ where B_{ij} is the (i, j) minor of B
       * place (i, j) in M iff t_{ij} is an odd integer
     - if M is a perfect matching then output M else fail.

5. Say that f_1, f_2, f_3, mapping from group G to H, are *linear consistent* if there exists a linear function $\phi : G \to H$ (that is $\forall x, y \in G, \phi(x) + \phi(y) = \phi(x + y)$) and $a_1, a_2, a_3 \in H$ such that $a_1 + a_2 = a_3$ and $f_i(x) = \phi(x) + a_i$ for all $x \in G$. A natural choice for a test of linear consistency is to verify that

   $$Pr_{x,y \in_r G}[f_1(x) + f_2(y) \neq f_3(x + y)] \leq \delta$$

   for some small enough choice of δ.

   - Assume G, H are Abelian. Show that f, g, h are linear-consistent iff for every $x, y \in G$ $f(x) + g(y) = h(x + y)$.

- Let $G = \{+1, -1\}^n$ and $H = \{+1, -1\}$. First note that since $a_i \in \{+1, -1\}$, then linear consistent f_i must be linear functions or "negations" of linear functions. We refer to the union of linear functions and the negations of linear functions as the *affine functions*. In class we expressed the minimum distance of f to a linear function. Express the minimum distance of a function f to an affine function.
- Show that if f_1, f_2, f_3 satisfy the above test, then for each $i \in \{1, 2, 3\}$, there is an affine function g_i such that $Pr_{x \in_r G}[f_i(x) \neq g_i(x)] \leq \delta$.
- (Extra credit) Show that there are linear consistent functions g_1, g_2, g_3 such that for $i \in \{1, 2, 3\}$, $Pr_{x \in_r G}[f_i(x) \neq g_i(x)] \leq \frac{1}{2} - \frac{2\gamma}{3}$ where $\gamma = \frac{1}{2} - \delta$.