COS 521: Advanced Algorithm Design
Homework 5
Due: Tue, Jan 16

Collaboration Policy: You may collaborate with other students on these problems. Collaboration is limited to discussion of ideas only, and you should write up the solutions entirely *on your own* and *list your collaborators* as well as *cite any references* you may have used.

1. Consider the following variant of the parallel Maximal Independent Set algorithm we saw in class: Recall that the main idea was to select an independent set in each iteration so as to make significant progress. Suppose an independent set is picked in each iteration in the following way: Pick a random number integer $r(v) \in \{1, n^{10}\}$ for every vertex v. A vertex v is selected in the independent set if $r(v) < r(u)$ for all neighbors u of v. Prove that the expected number of iterations required for this new scheme is $O(\log m)$.

2. In class, we analyzed an estimator for the second frequency moment $F_2 = \sum_i m_i^2$ (here m_i is the number of copies of element i). Here the basic form of the estimator consisted of a counter $C = \sum_i x_i m_i$ ($x_i \in_R \{+1, -1\}$) and the value of the estimator was C^2. A number of independent copies of this scheme were required to obtain a good estimate with high probability. Consider the following variant to estimate $F_3 = \sum_i m_i^3$: Maintain a counter $C = \sum_i x_i m_i$, where each x_i is picked uniformly and at random from $\{1, \omega, \omega^2\}$ (here ω is a complex cube root of unity). For the purpose of this question, ignore the issues of independence in the choices of x_i and assume that all x_i are picked independently.

   (a) Show that $E[C^3] = F_3$.

   (b) Consider the estimator $\mathrm{Re}[C^3]$ = the real part of C^3. Give an upper bound on the number of independent copies of this estimator needed to obtain an $1+\epsilon$ approximation of F_3 with probability $1-\delta$.

3. Consider the family of bit vectors of size n. The purpose of this exercise is to design a compact representation for this family which allows us to determine whether the hamming distance of two bit vectors x and y is above or below a given distance threshold t.

   (a) For bit vector x, let x_i denote the ith coordinate of x. Suppose we pick a random bit vector r where $r_i = 1$ with probability p and each coordinate is picked independently of the others. Consider two bit vectors x and y such that $d(x, y) = r$. Compute $\Pr[x.r = y.r(\mod 2)]$.

   (b) Use your results in part (a) to design a compact sketch for the bit vector family to support the following (approximate) distance comparison: By examining the sketches

$s(x)$ and $s(y)$ we would like to distinguish between the cases $d(x, y) \leq t$ and $d(x, y) \geq (1 + \epsilon)t$ with probability $1 - \delta$. Give a bound on the size of the sketch as a function of ϵ and δ.