CS5330: Assignment for Week 5

Due: Friday, 28th Feb 2020.

Here are solution sketches to the Week 5 problems. If anything is unclear, please talk to me or your TA.

1. Let V be the set of unit vectors in d dimensions whose each coordinate is either $1/\sqrt{d}$ or $-1/\sqrt{d}$. Call a pair of vectors ϵ-*orthogonal* if $|u^T v| \le \epsilon$. How large (in cardinality) a set of pairwise ϵ-orthogonal unit vectors from V can you construct?

   Sketch: Suppose we take a random set S of m vectors from V. For any two $u, v \in S$, $\Pr[|u^T v| > \epsilon] \le 2e^{-\varepsilon^2 d/2}$ by the Hoeffding bound. Therefore, $\Pr[\exists u, v \in S, |u^T v| > \varepsilon] \le \binom{m}{2} \cdot 2e^{-\varepsilon^2 d/2}$. Thus, for $m = e^{\epsilon^2 d/4}/2$, this probability is less than 1. Hence there exists a collection of $e^{\Omega(\varepsilon^2 d)}$ vectors which are pairwise ε-orthogonal.

2. ~~Suppose you can draw independent samples of a real random variable X that has expectation μ and standard deviation σ. Explain how to use only $O(\log n)$ samples from this source to generate a random variable Y with expectation μ such that $\Pr[|Y - \mu| > 2\sigma] < 1/n$.~~

3. Show that the randomized quicksort algorithm runs in time $O(n \log n)$ with high probability (i.e., with probability $1 - 1/n^c$ for some constant $c > 0$).

   Sketch: Let X_{ij} indicate if items i and j (indexed according to the sorted order) are compared by the algorithm. We are interested in $X = \sum_{i \in [n]} \sum_{j \in [n]: j > i} X_{ij}$.

   As discussed earlier, $X_{ij} = 1$ if and only if i or j are chosen as pivots before $\{i+1, \ldots, j-1\}$. Note that X_{ij} and $X_{i'j'}$ are independent if $i < i' < j' < j$. For $k = 0, 1, 2, \ldots, n-2$, let $Y_k = \sum_{i=1}^{i<(n-k+1)/2} X_{i,n-k+1-i}$. Check that each Y_k is the sum of independent random variables.

Now, $\mathbb{E}[Y_k] \le \sum_{i=1}^{i<(n-k+1)/2} \frac{2}{n-k-2i+2} \le 2H_{n-k}$. By the Chernoff Bound, $\Pr[Y_k > 12H_{n-k}] \le 2^{-12H_{n-k}} \le 1/(n-k)^4$.

Also, we have the trivial bound that $Y_k \le (n-k+1)/2 \le n-k$. Thus, we can write $\sum_{k=0}^{n-2} Y_k \le 10\sqrt{n} \cdot 10\sqrt{n} + \sum_{k:n-k>10\sqrt{n}} Y_k = 100n + \sum_{k:n-k>10\sqrt{n}} Y_k \le 100n + O(n \log n)$ with probability at least $1 - \sum_{i>10\sqrt{n}} 1/i^4 \ge 1 - 1/n$.

Other approaches are discussed in https://sarielhp.org/teach/13/fall_2013_cs_573/lec/slides/10_notes.pdf.