Assignment 2

1. Consider the definition of GapDist_k presented in Lecture 9. Let us modify the definition slightly as follows:

   Given: $\tilde{x} = x_1, x_2, \ldots, x_n$, where for all $1 \leq i \leq n$, $x_i \in [m]$, and $k \in \mathbb{N}$

   Output "Yes" if the number of distinct elements in $\tilde{x}$ is $\geq (1+\varepsilon)^{k+2}$,
   "No" if the number of distinct elements in $\tilde{x}$ is $\leq (1+\varepsilon)^{k-2}$

   (a) Modify the algorithm presneted in Section 9.1 appropriately to design a randomized algorithm for the above version of GapDIST_k which is correct with probability at least $1-\delta$ for any parameter $1 \geq \delta > 0$.

   (b) Let us call the alorithm you designed in the above step as $\mathcal{A}_\delta$. Design an algorithm, $\mathcal{A}_{\delta'}$, which for any given $\varepsilon > 0$, gives a $(1+\varepsilon)$-approximation algorithm for F_0 with probability at least $1-\delta'$ using $\mathcal{A}_\delta$. Analyze the space used by your algorithm.

2. Make approrpiate modifications to the algorithms presented in Section 11.1 and 11.2 in the notes so that they work in the turnstile model. Work out all the calculations for the modified algorithms. Read about the turnstile model at:
   `http://en.wikipedia.org/wiki/Streaming_algorithm#Models`

3. In the algorithm presented in Section 11.2, say the last line is changed from

   $$\text{"On query } a\text{, output Median}_{1\leq i\leq t}\{g_i(a)C[i][h_i(a)]\}\text{"}$$

   to

   $$\text{"On query } a\text{, output } \frac{\sum_{i=1}^{t}\{g_i(a)C[i][h_i(a)]\}}{t}$$

   The rest of the algorithm is kept as it is. Analyze the performance of this modified algorithm.

Assignment 3

1. Recall the definition of ε-NNS and ε-PLEB as described in class and as described in [IW]. Let $r_{\max}$ and $r_{\min}$ denote $\max_{i \neq j} d(p_i, p_j)$ and $\min_{i \neq j} d(p_i, p_j)$, respectively. Prove that ε-NNS can be solved by making $O\left(\log\log\left(\frac{r_{\max}}{r_{\min}}\right)\right)$ queries to ε-PLEB. Analyse the space complexity of the algorithm.

2. Give an example graph for which Feigenbaum et al. [FKMSZ]'s algorithm cannot achieve better than 6 competitive ratio.

Tutorial 1

Notation: Let $[m]$ denote the set $\{1, 2, \ldots, m\}$. Let $m > 10n$.

1. Recall Morris' algorithm and its analysis we saw in class. Let s_i denote the number of bits needed to store Y_i. Prove that $\mathbb{E}(s_i) = O(\log \log n)$. What is the $\mathbb{V}ar(s_i)$? From this conclude that $\Pr[s_i \geq 100 \log \log n] \leq 1/4$.

2. Suppose Alice decides to toss a coin till she sees a Head or has tossed the coin at most k times, where k is such that $k \geq 1$. Assume she is tossing a fair coin, that is the probability of head and tail is equal to $1/2$ each. What is expected number of Tails Alice sees?

3. Given a stream of $2n$ numbers from $[m]$ with the property that it contains $n + 10$ distinct elements and one other element which appears exactly $n - 10$ times, design a one-pass randomized algorithm that outputs the element that appears $n - 10$ times using space $O(\log m + \log n)$ with probability at least $9/10$.

4. Let $\{0, 1\}^n$ denote the set of all $0/1$ strings of length n. For example, $\{0, 1\}^2 = \{00, 01, 10, 11\}$. Given a string $w \in \{0, 1\}^{2n}$ design an algorithm which accepts w if it is a palindrome and rejects it otherwise. A string w is said to be a palindrome if $w = \mathrm{rev}(w)$, where $\mathrm{rev}(w)$ is simply the string w reversed. For example, 1011 is not a palindrome, whereas 11011011 is a palindrome.

Type of algo	space used	# passes	error	points you get
Deterministic	$O(n)$	1	0	2
Deterministic	$O(\log n)$	n	0	2
Randomized	$O(\log n)$	1	$\leq 1/4$	5

Tutorial 2

1. Read Section 7.1 of the uploaded notes. Compute $\mathbb{E}(1/Z)$, where the Z is a variable defined as in the algorithm in Section 7.1.

2. Using the construction of Exercise 8 in Lecture 0 uploaded on the course page, give a construction for 4-wise independent family of functions.

3. Consider the algorithm in Section 7.2 in the notes uploaded on the course page. Modify the algorithm appropriately so that it is an (ε, δ) approximation for the given ε and δ factors.

4. Let $X_1, X_2, \ldots, X_n$ be 0-1 pairwise independent random variables. Prove or disprove the following: If $X = \sum_{i=1}^{n} X_i$ then $\mathbb{V}ar(X) \leq \mathbb{E}(X)$.

5. **Hint:** The following theorem may be useful in coming up with a solution for Problem 4 of Tutorial 1:

   Theorem (Shwartz-Zippel Theorem). *Let $p(x_1, x_2, \ldots, x_n)$ be a non-zero polynomial of degree d with coefficients from $\mathbb{F}_q$, where q is a prime. Then*

$$\Pr_{a_1, a_2, \ldots, a_n \in \mathbb{F}_q} [p(a_1, a_2, \ldots, a_n) = 0] \leq \frac{d}{q}$$