Advanced Algorithms 2021 10/29, 2021

Exercise 06

Lecturer: Mohsen Ghaffari *Teaching Assistant: Saeed Ilchi*

Note: In all the problems, suppose data stream elements are in $\{1, 2, \ldots, n\}$ *and let* m *be the length of the stream.*

1 Majority

In the class, we discussed the deterministic algorithm of Boyer and Moore which has space complexity $O(\log n + \log m)$ and outputs an element such that if the stream has a majority, the output is equal to this majority. Suppose $m \geq n$ and prove that any deterministic algorithm that is able to decide whether a stream has a majority element or not needs $\Omega(n)$ bits of space.

2 Deterministic Approximate Counting [Recommended]

Suppose $m \geq n$ and show that any deterministic algorithm that computes the exact number of distinct elements in a stream requires $\Omega(n)$ bits of space. Next, prove that even computing a 1.1-approximation of the number of distinct elements needs $\Omega(n)$ bits of space.

Hint:

For the second part, suppose a graph on subsets of $[n]$. We put an edge between two subsets if and only if the size of their symmetric difference is at most $\frac{n}{5}$. Find a large independent set in this graph.

3 Frequent Elements [Recommended]

Extend the majority algorithm of Boyer and Moore, which we discussed in the class, to an algorithm with space complexity of $O\left(k\left(\log n + \log m\right)\right)$ bits that outputs k elements that include those that appear in more than $1/k$ fraction of the stream.

Hint:

In the majority algorithm, we make bundles of size two. Here, think about bundles of size k.

4 Morris's Approximate Counting Algorithm

In Morris's approximate counting algorithm, which we discussed in the class, prove that

$$\mathbb{E}[2^{2X_m}] = \frac{3}{2}m^2 + \frac{3}{2}m + 1.$$

5 Pairwise Independent Hashing

Prove that given a prime number p, the random hash function $h_{a,b}(x) : \mathbb{F}_p \to \mathbb{F}_p$ defined as $h_{a,b}(x) = ax + b \ mod \ p$, where a and b are random numbers in $\{0, 1, \ldots, p-1\}$, is a pairwise independent hash function. That means, for every $x, y \in \{0, 1, \ldots, p-1\}$ where $x \neq y$, and every $i, j \in \{0, 1, \ldots, p-1\}$, we have $\Pr_{a,b}[h_{a,b}(x) = i \textit{ and } h_{a,b}(y) = j] = 1/p^2$.

Hint:

Fixing x, y, i, j gives us a system of linear equations on $\mathbb{F}_p$ with two equations and two variables (a and b). Prove that this system always has a unique solution.