Assignment 3 - Homework Exercises on Streaming Algorithms

Duy Pham - 0980384
Maciej Wikowski - 0927420
Pattarawat Chormai - 0978675

October 21, 2015

Str.I-1

II-1

(i)

We know that $\mathrm{E}[\widetilde{\Phi}(\sigma)] = \Phi(\sigma)$ and $\mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right] = (1/3) \cdot \Phi(\sigma)$, thus:

$$\Pr[|\widetilde{\Phi}(\sigma) - \Phi(\sigma)| \geq c \cdot \Phi(\sigma)] = \Pr[|\widetilde{\Phi}(\sigma) - \mathrm{E}[\widetilde{\Phi}(\sigma)] \geq \ 3c \cdot \mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right]]$$

According to Chebyshev Inequality, we have:

$$\begin{aligned}
&\Pr[|\widetilde{\Phi}(\sigma) - \mathrm{E}[\widetilde{\Phi}(\sigma)|] \geq 3c \cdot \mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right]] \\
&= \Pr[|\widetilde{\Phi}(\sigma) - \mathrm{E}[\widetilde{\Phi}(\sigma)|] \geq 3c \cdot \sqrt{(}\,\mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right]) \cdot \sqrt{(}\,\mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right])] \\
&\leq \frac{1}{9 \cdot c^2 \cdot \mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right]}
\end{aligned}$$

To let 1/6 be the upper bound of this probability, we have to choose a value of c such that:

$$\begin{aligned}
9 \cdot c^2 \cdot \mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right] &= 6 \\
c &= \sqrt{(}\frac{6}{9\,\mathrm{Var}\left[\widetilde{\Phi}(\sigma)\right]}) \\
&= \sqrt{(}\frac{2}{3\frac{\Phi(\sigma)}{3}})
\end{aligned}$$

Since we know that $\Phi(\sigma) > 3$, then:

$$c = \sqrt{(}\frac{2}{\Phi(\sigma)})) \quad < \sqrt{(}\frac{2}{3})$$

This value of c statisfies the condition that $0 < c < 1$.

(ii)

The idea of the new algorithm is to run ALG k times, then taking the median of the k return values $\widetilde{\Phi}(\sigma)$ (the Median trick). Before modifying the algorithm formally, we make some analysis as follows.

Let X_i be an indicator random variable, which is defined as:

$$X_i = \begin{cases} 1 \text{ if } \Pr[|\widetilde{\Phi}(\sigma) - \Phi(\sigma)| \geq c \cdot \Phi(\sigma)] \\ 0 \text{ otherwise} \end{cases}$$

Let $X = \sum_{i=1}^{k} X_i$ be the random variable representing the final outcome of the new algorithm. We have:

$$\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^{k} X_i] = \sum_{i=1}^{k} \mathrm{E}[X_i]$$

In this algorithm, we used the specified value of c in part (i), which confirms that:

$$\Pr[|\widetilde{\Phi}(\sigma) - \Phi(\sigma)| \geq c \cdot \Phi(\sigma)] \leq 1/6$$

which implies that
$\Pr[X_i = 1] \leq 1/6$

Because X_i is an indicator random variable, then:
$\mathrm{E}[X_i] = 1/6$

Thus, we have:

$$\mathrm{E}[X] = \sum_{i=1}^{k} \mathrm{E}[X_i] = k/6$$

When $X_i = 1$, it implies that we do not have the event described in the problem statement. The probability that we do not have it after performing the Median trick is:

$$\begin{aligned} \Pr[X > \frac{k}{2}] &= \Pr[X > 3\,\mathrm{E}[X]] = \Pr[X > (1+2)\,\mathrm{E}[X]] \\ &\leq (\frac{e^2}{3^3})^{\frac{k}{6}} \\ &= (\frac{e^2}{27})^{\frac{k}{6}} \end{aligned}$$

Thus, the probability that we get the event described is $1 - (\frac{e^2}{27})^{\frac{k}{6}}$. To get the desired value $(1 - \delta)$, we have to choose k such that:

$$\delta \geq (\frac{e^2}{27})^{\frac{k}{6}}$$
$$\iff \frac{1}{\delta} \leq (\frac{27}{e^2})^{\frac{k}{6}}$$
$$\iff \log\frac{1}{\delta} \leq \frac{k}{6}\log\frac{27}{e^2}$$
$$\iff k \geq \frac{6}{\log\frac{27}{e^2}} \cdot \log\frac{1}{\delta}$$

We can choose

$$k = \lceil 4\log\frac{1}{\delta} \rceil$$

In conclusion, we modify the algorithm as represented in algorithm 1.

Algorithm 1 Revised Algorithm Taking ALG as Sub-Routine

Require: A stream $\sigma =< a_1, \cdots, a_m >, \delta$
Ensure: Statistics $\widehat{\Phi}(\sigma)$
Operation:
  Choose $k := \lceil 4\log\frac{1}{\delta} \rceil$
  Init $J := \emptyset$
  for i from 1 to k **do**
    $\widetilde{\Phi}_i(\sigma) :=$ ALG ()
    $J = J \cup \widetilde{\Phi}_i(\sigma)$
  end for
    return The median of set J

The algorithm needs to store the set J of k elements, and each run of ALG stores $B(n, m)$ bits, so the total number of bits used by the algorithm is $O(kB(n, m))$.