Exercise I, Sublinear Algorithms for Big Data Analysis 2024-2025

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students, and solve as many problems as you can. Problems marked (*) are more difficult, but also more rewarding. These problems have been taken from various sources on the Internet, too numerous to cite individually.

1 In class we saw a constant factor approximate randomized counting algorithm with space complexity $O(\log \log n)$, where n is the maximum value of the counter. Prove that any *deterministic* algorithm that provides a factor 2 approximation to the count must use $\Omega(\log n)$ space.

Solution. We can view the algorithm as a deterministic sequence of transitions between memory configurations. Consider a directed graph G where nodes are memory states of the algorithm, and directed edges (S_1, S_2) correspond to transition from state S_1 to S_2 when an event happens. Let S^* denote the initial state. Then the graph G is a directed path plus cycle: as events happen, we transition to new states, until a state is repeated. Let L denote the length of the 'stem' and C the length of the cycle. Then the algorithm outputs the same answer when L events happen and $L + C \cdot \lceil L/C \rceil$ events happen (as traversing the cycle any number of times brings the algorithm back to the same state). Since $L + C \cdot \lceil L/C \rceil \geq 2L$, this contradicts the assumption that the algorithm gives a factor two approximation, unless $L + C \cdot \lceil L/C \rceil > n$. Since

$$L + C \cdot \lceil L/C \rceil \leq L + C \cdot (L/C + 1) \leq L + C,$$

we thus get that $L + C > n$. Thus, the number of memory states that the algorithm can be in is at least $L + C > n$, and hence the space complexity is $\Omega(\log n)$.

□

2 Let X_n denote the random variable maintained by Morris' algorithm after n events have happened. Prove that $\mathbf{E}[2^{2X_n}] = O(n^2)$.

Solution. We prove by induction on n that $\mathbf{E}[2^{2X_n}] \leq An^2+Bn+C$ for some absolute constants A, B, C. The base is provided by $n = 0$ as long as $C \geq 1$.

We now prove the inductive step, $n \to n+1$. We have

$$\begin{aligned}
\mathbf{E}[2^{2X_{n+1}}] &= \sum_{j=0}^{\infty} \mathbf{Prob}[X_n = j] \cdot \mathbf{E}[2^{2X_{n+1}} | X_n = j] \\
&= \sum_{j=0}^{\infty} \mathbf{Prob}[X_n = j] \cdot (2^{-j} 2^{2(j+1)} + (1 - 2^{-j}) 2^{2j}) \\
&= \sum_{j=0}^{\infty} \mathbf{Prob}[X_n = j] \cdot (2^{2j} + 2^{j+2} - 2^{j}) \\
&= \sum_{j=0}^{\infty} \mathbf{Prob}[X_n = j] \cdot (2^{2j} + 3 \cdot 2^{j}) \\
&= \sum_{j=0}^{\infty} \mathbf{Prob}[X_n = j] \cdot 2^{2j} + 3 \sum_{j=0}^{\infty} \mathbf{Prob}[X_n = j] \cdot 2^{j} \\
&= \mathbf{E}[2^{2X_n}] + 3\mathbf{E}[2^{X_n}] \\
&= An^2 + Bn + C + 3(n+1) \quad \text{(by inductive hypothesis and result from class)}
\end{aligned}$$

We thus need to ensure that

$$\begin{aligned}
A(n+1)^2 + B(n+1) + C &= An^2 + (B + 2A)n + (A + B + C) \\
&\geq An^2 + (B + 3)n + (C + 3)
\end{aligned}$$

This means that we want $B + 3 \leq B + 2A$, i.e. $A \geq 3/2$ (set $A = 3/2$). We also want $A + B + C \geq C + 3$, which is satisfied when $B = 3/2$. Finally, we have $C = 0$ by the base case. We have shown that $\mathbf{E}[2^{2X_{n+1}}] = \frac{3}{2}n^2 + \frac{3}{2}n + 1$ for all $n \geq 0$.

□