CS5120: Probability & Computing **Jan-May 2020**

Lecture 13: Short Analysis of Morris' Algorithm

Lecturer: N.R.Aravind *Scribe: N.R.Aravind*

1 Morris' Algorithm

We recall the algorithm (due to Morris) for counting the length of a stream.

- Initialize $C = 0$.
- On seeing a new bit/item, increment C with probability $\frac{1}{2^C}$.
- When the stream ends, output 2^C as the approximation to the length of the stream.

Let $C(i)$ be the value of the counter after seeing i items. From the assignment problem, we get $E[2^{C(n)}] = n$ and $Var[2^{C(n)}] \leq n^2$. Thus, we can apply Proposition 1 and get a (ε, δ)-approximation of n, using $O(\frac{1}{\varepsilon^2}\log(\frac{1}{\delta})\log\log m)$ space.

2 Improving Error and Approximation Guarantee

We recall the following from the notes on mean estimation.

Proposition 1 *Let X be a random variables such that $\sigma[X] \leq cE[X]$. Let $r = \left\lceil 32\log\left(\frac{1}{\delta}\right)\right\rceil$, $s = \left\lceil \frac{16c}{\epsilon^2}\right\rceil$, and let $\{Y_{i,j} : 1 \leq i \leq r, 1 \leq j \leq s\}$ be independent samples of X. Let $Y_i = \frac{Y_{i,1} + \ldots + Y_{i,s}}{s}$ for $i = 1, \ldots, r$ and let Y be the median of $Y_1, \ldots, Y_r$. Then Y is a (ε, δ)-approximation of $E[X]$.*