COMS W4995-3: Advanced Algorithms Jan 18, 2017

Lecture 1 – Counting, Morris' Algorithm, Probability

Instructor: *Alex Andoni* Scribes: *Noah Gallant*

1 The Counting Problem: count up to n

1.0.1 Normal counting and space

Let n be the number of events or ticks we would like to keep track of, for example the number of suspicious requests a router receives. We would like to keep track of this number, what is the space required (in bits)? For an exact count it is necessarily $\log(n)$ bits. Can we do any better getting the exact count? Nope!

1.0.2 Approximate Count

To use less space we can try to compute **approximate count**. Where if a represents the actual count, we define the approximate count, $\hat{a}$ as follows:

Definition 1:

$$a \leq \hat{a} \leq u \cdot a$$

Where u is called the approximation factor. Oftentimes, we will think of approximation being $u = 1 + \epsilon$, where ϵ is the "error" (e.g., $\epsilon = 0.1$ means that the algorithm can overestimate the count, by at most 10%).

Definition 2:

$$a \;/\; l \leq \hat{a} \leq u \cdot a$$
$$u, l \geq 1$$

Where $u \cdot l$ is our approximation factor. This second definition can be translated to the first:

$$\hat{a}' = l \cdot a \rightarrow$$
$$a \leq \hat{a}' \leq (u \cdot l) \cdot a$$

However, even using approximation, the optimal space is still $\Omega(\log(n))$.

Hence we can consider a further relaxation to our counting problem: **randomized approximate counting** wherein we only require that:

$$Pr[a \leq \hat{a} \leq u * a] \geq 90\%$$

1.0.3 Algorithm for Randomized Approximate Counting

It turns out we can solve the randomized approximate counting with much less space ($O(\log \log n)$ bits only). In particular, we will see the **Morris' Randomized Approximate Counting**, conceived in 1978, whose underlying idea is as follows. We use a counter $X \in \mathbb{Z}$.

1. initialize $X = 0$
2. at each event $X := X + 1$ with probability $\frac{1}{2^X}$, and leave unchanged with probability $1 - \frac{1}{2^X}$.
3. output (the estimate) is $\hat{a} = 2^X - 1$.

2 Probability

Let X be a random variable.

Definition 1 (Expectation)**.** *For a discrete random variable X, the expectation of X, $\mathbb{E}[X]$ is*

$$\mathbb{E}[X] = \sum_a a \Pr[X = a]$$

For a continuous random variable X, the expectation of X, $\mathbb{E}[X]$, is

$$\mathbb{E}[X] = \int a\phi(a)da$$

where ϕ is the probability density function of X.

Lemma 2 (Linearity of Expectation)**.** *Let X and Y be two random variables.* $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Lemma 3 (Markov's inequality)**.** *Let X be a non-negative random variable. For all $\lambda > 0$,*

$$\Pr[X > \lambda] \leq \frac{\mathbb{E}[X]}{\lambda}$$

Definition 4 (Variance)**.** *The variance of a random variable X, denoted var$[X]$, is*

$$Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Lemma 5 (Chebyshev's Inequality)**.** *For all $\lambda > 0$,*

$$\Pr[|X - \mathbb{E}[X]| > \lambda] \leq \frac{Var[X]}{\lambda^2}$$

3 Analysis of Morris' Algorithm

Claim 6. *Define X_n = value of X after n events. Then $\mathbb{E}[2^{X_n} - 1] = n$.*

Proof. The proof is by induction on n. The basis of the induction: for $n = 0$, $X = 0$ and therefore $2^{X_0} = 1 = n + 1$. So assume for inductive hypothesis that $\mathbb{E}[2^{X_{n-1}}] = (n - 1) + 1$. We now argue the inductive step. Note that

$$\mathbb{E}[2^{X_n}] = \sum_i 2^i \cdot \Pr[X_n = i].$$

Additionally,

$$\begin{aligned}
\Pr[X_n = i] &= \Pr[X \text{ is incremented} \wedge X_{n-1} = i - 1] + \Pr[X \text{ is not incremented} \wedge X_{n-1} = i] \\
&= \frac{1}{2^{i-1}} \cdot \Pr[X_{n-1} = i - 1] + \left(1 - \frac{1}{2^i}\right) \cdot \Pr[X_{n-1} = i].
\end{aligned}$$

Using the above two facts, we may write:

$$\begin{aligned}
\mathbb{E}[2^{X_n}] &= \sum_i 2^i \cdot \left(\frac{1}{2^{i-1}} \cdot \Pr[X_{n-1} = i - 1] + \left(1 - \frac{1}{2^i}\right) \cdot \Pr[X_{n-1} = i]\right) \\
&= \sum_i 2 \cdot \Pr[X_{n-1} = i - 1] + \sum_i 2^i \cdot \Pr[X_{n-1} = i] - \sum_i \Pr[X_{n-1} = i] \\
&= \mathbb{E}[2^{X_{n-1}}] + 1.
\end{aligned}$$

We now apply the inductive hypothesis to conclude:

$$\mathbb{E}[2^{X_n}] = ((n - 1) + 1) + 1 = n + 1.$$

□