(ϵ, δ)-Estimators:
How to count very efficiently up to n

MSC - CR01 - Randomness in Science

Lecture 3 by Nicolas Schabanel

Jan 9, 2014

1 LogLog-counter

Theorem 1. *There is no deterministic counter that can count up to n with less than $\lceil \log_2 n \rceil$ bits of memory.*

Proof. A machine with K bits of memory can be in at most 2^K differents states. If $K < \lceil \log_2 n \rceil$, then $2^K < n$ and there are two distinct integers $i \neq j$ for which the counter is in the same state and the counter may thus not distinguish between them. □

Consider the following randomized counter:

VARIABLES:	A variable i
INIT:	$i := 0$
INC:	$i := \begin{cases} i+1 & \text{with probability } 1/2^i \\ i & \text{otherwise} \end{cases}$
VAL:	return $2^i - 1$

INIT is called to reset the counter. INC is called every time the counter needs to be incremented. And VAL is called every time we want to know the value of the counter.

Let I_n and X_n be the two random variables for the value of i and the value returned by VAL after n calls to INC ($X_n = 2^{I_n} - 1$).

Note that this algorithm requires only $O(\log i) = O(\log \log n)$ bits of memory to store i and to increment i with probability $1/2^i$ since it is enough to flip i times an unbaised coins which requires only to count up to i, i.e. $O(\log i)$ bits of memory.

Theorem 2. *For all $n \geqslant 0$, $\mathbb{E}[X_n] = n$.*

Proof. We proceed by recurrence. For $n = 0$, $X_0 = 2^0 - 1 = 0$. Let us now assume that $\mathbb{E}[X_n] = n$. Then,

$$\begin{aligned}
\mathbb{E}[X_{n+1} | X_n = x = 2^i - 1] &= \frac{1}{2^i}(2^{i+1} - 1) + (1 - \frac{1}{2^i})(2^i - 1) \\
&= 2 - \frac{1}{2^i} + 2^i - 1 - 1 + \frac{1}{2^i} \\
&= 2^i = X_n + 1
\end{aligned}$$

It follows that $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n] + 1 = n + 1$ by recurrence hypothesis. □

In order to get a precise evaluation of the quality of the approximation, let us evaluate the variance for X_n.

Theorem 3. *For all* $n \geqslant 0, \mathbb{E}[X_n^2] \leqslant \frac{3}{2}n^2$.

Proof. Let us proceed by recurrence to show that $\mathbb{E}[X_n^2] \leqslant (an+b)^2$ for some constants a and b to be chosen later. For $n = 0$, $X_0^2 = 0 \leqslant (a \cdot 0 + b)^2$ for all a and b. Let us now assume that $\mathbb{E}[X_n^2] \leqslant (an+b)^2$. Then,

$$\begin{aligned}
\mathbb{E}[X_{n+1}^2 | X_n = x = 2^i - 1] &= \frac{1}{2^i}(2^{i+1}-1)^2 + (1-\frac{1}{2^i})(2^i-1)^2 \\
&= \frac{1}{2^i}(4 \cdot 2^{2i} - 4 \cdot 2^i + 1) + (1-\frac{1}{2^i})(2^{2i} - 2 \cdot 2^i + 1) \\
&= 4 \cdot 2^i - 4 + \frac{1}{2^i} + 2^{2i} - 2 \cdot 2^i + 1 - 2^i + 2 - \frac{1}{2^i} \\
&= 2^{2i} + 2^i - 1 \\
&= (2^i - 1)^2 + 3 \cdot (2^i - 1) + 1 \\
&= (X_n)^2 + 3X_n + 1
\end{aligned}$$

It follows that:

$$\begin{aligned}
\mathbb{E}[X_{n+1}] &= \mathbb{E}[X_n^2] + 3\,\mathbb{E}[X_n] + 1 \\
&\leqslant (an+b)^2 + 3n + 4 = a^2n^2 + (2ab+3)n + 1 \quad \text{(by recurrence hypothesis)} \\
&\leqslant (a(n+1)+b)^2 = a^2n^2 + 2a(a+b)n + (a+b)^2,
\end{aligned}$$

for a and b chosen such that $2a^2 \geqslant 3$ and $(a+b)^2 \geqslant 1$, for instance take $a^2 = \frac{3}{2}$ and $b = 0$ which yields the claimed result. □

It follows that $\mathbb{V}\text{ar}(X_n) = \mathbb{E}[X_n^2] - \mathbb{E}[X_n]^2 \leqslant n^2/2 = O(\mathbb{E}[X_n]^2)$. As we will see in the next section, one can sample very efficiently and with an arbitrarily small error, an arbitrarily close approximation of the expected value of a random variable whose variance is of the same order as the square of its expectation.

2 (ϵ, δ)-estimators

Definition 1. *An* (ϵ, δ)-estimator Z *for a value* μ *is a random algorithm computing a random value* Z *such that:*

$$\Pr\{|Z - \mu| \geqslant \epsilon\mu\} \leqslant \delta$$

where ϵ *and* δ *are two values in* $[0, 1)$.

Let us consider $X_n^1, \ldots, X_n^\ell$ ℓ random variables i.i.d as X_n (i.e. ℓ independent randomized counters). Let $\mu = \dfrac{X_n^1 + \cdots + X_n^\ell}{\ell}$.

Lemma 3.1. *For all* ϵ, ℓ *and* n, $\Pr\{|\mu - n| \geqslant \epsilon n\} \leqslant \dfrac{1}{2\ell\epsilon^2}$.

Note that the bound $1/2\ell\epsilon^2$ is *independent* of n!

Proof. Note that the expectation of each X_n^j is n, so is the expectation of their average μ. furthermore, since the X_n^j are independent, $\mathbb{V}\text{ar}(\mu) = \dfrac{1}{\ell}\mathbb{V}\text{ar}(X_n) \leqslant \dfrac{n^2}{2\ell}$. Thus, applying Chebychev inequality yields:

$$\Pr\{|\mu - n| \geqslant \epsilon n\} \leqslant \frac{\mathbb{V}\text{ar}(\mu)}{(\epsilon n)^2} \leqslant \frac{1}{\epsilon^2 n^2} \cdot \frac{n^2}{2\ell} = \frac{1}{2\ell\epsilon^2}.$$ □

Let us now consider k groups of ℓ random variables $X_n^{1,1}, \ldots, X_n^{\ell,k}$ i.i.d. as X_n. Let as before $\mu^j = \dfrac{X_n^{1,j} + \cdots + X_n^{\ell,j}}{\ell}$. And let now Z to be the median value of $\mu^1, \ldots, \mu^k$.

The key argument is that if the median is outside the interval $n \pm \epsilon n$, then *at least $k/2$ of the averages $\mu^1, \ldots, \mu^\ell$ are outside $n \pm \epsilon n$*, which is exponentially unlikely as soon as $\frac{1}{2\ell\epsilon^2} < \frac{1}{2}$ by Hoeffding's inequality.

Theorem 4. *For all $\epsilon, \delta \in (0,1)$ and all n, Z is a (ϵ, δ)-estimator for n when $\ell = 2/\epsilon^2$ and $k = 8\ln(1/\delta)$ (two constants!).*

Proof. Let us introduce the indicator random variable Y^j for the event "$\mu^j \notin (1 \pm \epsilon)n$", i.e. $Y^j = 1$ if $\mu^j \notin [n - \epsilon n, n + \epsilon n]$ and $Y^j = 0$ otherwise. According to the lemma above, $\mathbb{E}[Y^j] = \Pr\{Y^j = 1\} \leqslant 1/2\ell\epsilon^2 \leqslant 1/4$ for $\ell \geqslant 2/\epsilon^2$. The, from the observation above,

$$\begin{aligned}
\Pr\{|Z - n| \geqslant \epsilon n\} &\leqslant \Pr\{Y^1 + \cdots + Y^k \geqslant k/2\} \\
&\leqslant \Pr\{Y^1 + \cdots + Y^k - \mathbb{E}[Y^1 + \cdots + Y^k] \geqslant k/4\} \\
&\leqslant \exp\left(-2(k/4)^2 \Big/ \sum_{j=1}^{k} (1-0)^2\right) \quad \text{(by Hoeffding's inequality)} \\
&= \exp(-k/8) \leqslant \delta,
\end{aligned}$$

when $k \geqslant 8\ln(1/\delta)$. □

Note that the averaging of the counters allows to increase the precision to an arbitrary small ϵ whereas taking the median allows to reduce exponentially the error probability to an arbitrary small δ.

We conclude with the following LogLog-Counter.

VARIABLES:	$k\ell$ variables $i^{1,1}, \ldots, i^{\ell,k}$ with $k = \lceil 8\ln(1/\delta) \rceil$ and $\ell = \lceil 2/\epsilon^2 \rceil$
INIT:	Set $i^{p,q} := 0$ for all $p = 1..\ell$ and $q = 1..k$
INC:	Increment independently all counters: for all $p = 1..\ell$ and $q = 1..k$: set $i^{p,q} := \begin{cases} i^{p,q} + 1 & \text{with probability } 1/2^{i^{p,q}} \\ i^{p,q} & \text{otherwise} \end{cases}$
VAL:	for all $q = 1..k$: compute $\mu^q := \dfrac{2^{i^{1,q}} + \cdots + 2^{i^{\ell,q}}}{\ell} - 1$ return the median of the values $\mu^1, \ldots, \mu^k$

Theorem 5. *The algorithm above is a randomized counter using $O(\log(1/\delta)\log\log(n)/\epsilon^2) = O(\log\log n)$ bits of memory that returns a value within $n \pm \epsilon n$ with error probability at most δ for all $\epsilon, \delta \in (0,1)$, and all number of calls, n, to* INC.

Exercise 1. *Consider an arbitrary random variable with finite expectation $\mathbb{E}[X]$ and finite variance verifying $\mathbb{V}\text{ar}(X) \leqslant A \cdot \mathbb{E}[X]^2$ for some constant A. Design a (ϵ, δ)-estimator for computing $\mathbb{E}(X)$.*