The AMS Estimator of F_k

So far, we have seen estimators for F_0, F_2. Today, we explore a general estimator for F_k where $k \geq 2$.

The motivation is we pick an element at random from the stream. We count the number of instances of the picked element from that point onwards. The basic estimator is $m(r^k - (r-1)^k)$ where r is the number of instances and m the length of the stream.

However, as we do not know m beforehand, we have to pick one on the fly as we saw in the week 1 problem set.

AMS Estimator on stream $\sigma = (s_1, \dots, s_m)$ from $\{1, \dots, n\}$.

- Initialize: Set $r = 0$, $a = 0$.
- Process s_j:
  - With prob. $\frac{1}{j}$, set $a \leftarrow s_j$, $r \leftarrow 0$
  - If $s_j = a$, $r \leftarrow r+1$
- Output: $m\left(r^k - (r-1)^k\right)$.

Analysis

Let A, R be the r.v.'s for the values of a, r at the end of alg., respectively.

Let X denote the output.

We begin by considering probabilities conditioned on $A = i$.

If $A = i$, then R is equally likely to be any of the values $\{1, \dots, f_i\}$ depending on which of the f_i occurances of i was picked by the alg.

$$\Pr\left(R = l \mid A = i\right) = \begin{cases} \frac{1}{f_i} & 1 \le l \le f_i \\ 0 & l > f_i \end{cases} .$$

$$\Rightarrow E\left(X \mid A = i\right) = E\left(m\left(R^k - (R-1)^k\right) \,\middle|\, A = i\right)$$

$$= \sum_{l=1}^{f_i} \frac{1}{f_i} m\left(l^k - (l-1)^k\right)$$

$$= \frac{m}{f_i}\left(f_i^k - 0^k\right) \qquad \text{(by telescoping sums)}$$

$$= m f_i^{k-1}$$

Therefore, the general expectation is

$$E(X) = \sum_{i=1}^{n} \Pr(A = i)\, E\left(X \mid A = i\right)$$

$$= \sum_{i=1}^{n} \frac{f_i}{m} m f_i^{k-1} = \sum_i f_i^k = F_k . \qquad \checkmark$$

So, the X produced is indeed an unbiased estimator for F_k.

To bound the variance,

$$Var(X) = E(X^2) - E(X)^2 \le E(X^2) = \sum_{i=1}^{n} \overset{\frac{f_i}{m}}{\bullet} \sum_{l=1}^{f_i} \frac{1}{f_i} m^2 \left(l^k - (l-1)^k\right)^2$$

$$= m \sum_{i=1}^{n} \sum_{l=1}^{f_i} \left(l^k - (l-1)^k\right)^2$$

Recall the mean value thm: $\exists\, y \in [x-1, x]$ s.t.

$$f(x) - f(x-1) = f'(y)$$

for continuous differentiable fns f. Apply this with $f(x) = x^k$

$$\Rightarrow \quad x^k - (x-1)^k = k\,y^{k-1} \le k\,x^{k-1} \qquad (\text{for } x \ge 1).$$

Applying this to the previous bnd.,

$$\begin{aligned}
\mathrm{Var}(X) &\le m \sum_{i=1}^{n} \sum_{l=1}^{f_i} k\, l^{k-1} \left(l^k - (l-1)^k\right) \\
&\le m \sum_{i=1}^{n} k\, f_i^{k-1} \underbrace{\sum_{l=1}^{f_i} \left(l^k - (l-1)^k\right)}_{\text{Again by telescoping sums}} \\
&= mk \sum_{i=1}^{n} f_i^{k-1} \cdot f_i^{k} \\
&= k\, F_1\, F_{2k-1}.
\end{aligned}$$

If we want to apply the median-of-means technique from Lec 2, it would be useful if $\mathrm{Var}(X)$ can be expressed as a multiple of $E(X)^2$ (or at least a bound of $\mathrm{Var}(X)$ can be expressed as such).

<u>Lemma</u> For $x_1, \dots, x_n \ge 0$ reals and $k \ge 1$ real. Then

$$\left(\sum_i x_i\right)\left(\sum_i x_i^{2k-1}\right) \le n^{1-1/k} \left(\sum x_i^k\right)^2.$$

<u>Cor</u> $\quad F_1 \cdot F_{2k-1} \le n^{1-1/k}\, F_k^{\,2}.$

Pf of Lemma

Let $M = \max_i x_i$. Then

$$M^{k-1} = \left(M^k\right)^{k-1/k} \le \left(\sum_i x_i^k\right)^{(k-1)/k}$$

As the fn $x \longmapsto x^k$ is convex,

$$\left(\frac{1}{n}\sum_i x_i\right)^k \le \frac{1}{n}\sum_i x_i^k$$

$$\frac{1}{n}\sum x_i^k \le \left(\frac{1}{n}\sum_i x_i^k\right)^{1/k}$$

$$\begin{aligned}
\Rightarrow \quad \left(\sum_i x_i\right)\left(\sum_i x_i^{2k-1}\right) &\le \left(\sum_i x_i\right)\left(M^{k-1}\sum x_i^k\right) \\
&\le \left(\sum_i x_i\right)\left(\sum_i x_i^k\right)^{(k-1)/k}\left(\sum x_i^k\right) \\
&\le n\left(\frac{1}{n}\sum_i x_i^k\right)^{1/k}\left(\sum_i x_i\right)^{2-1/k} \\
&= n^{1-1/k}\left(\sum_i x_i^k\right)^2 .
\end{aligned}$$

$$\Rightarrow \quad E(X) = F_k \quad \text{and} \quad Var(X) \le k n^{1-1/k} F_k^2 .$$

We can apply median-of-means to get a (ε, δ)-approximation using

$$O\left(\frac{1}{\varepsilon^2}\log\left(\frac{1}{\delta}\right)\frac{Var(X)}{E(X)^2}\left(\log m + \log n\right)\right)$$

$$= O\left(\frac{1}{\varepsilon^2}\log\left(\frac{1}{\delta}\right) k\, n^{1-1/k}\left(\log m + \log n\right)\right)$$

Note, as $k \to \infty$, this isn't that great cause each f_i can be calculated precisely with $O(n \log m)$ space. The alg. is also not efficient for calculating F_2; rather its just a good exercise in streaming algorithm analysis.

<u>Using the Johnson-Lindenstrauss Lemma to calculate F_2.</u>

The goal of calculating F_2 can be thought of as calculating $\|f\|$ where $f = (f_1, \dots, f_n)$. The premise of the predicament is that storing all of f is tedious $(O(n))$ and we want to store something smaller.

The idea is to store instead $y = Lf$ where L is a $k \times n$ matrix so $y \in \mathbb{R}^k$ and then by the JL Lemma for $k = \Omega\left(\frac{\ln(1/\delta)}{\epsilon^2 - \epsilon^3}\right)$

$$\Pr\left((1-\varepsilon)\|f\|^2 \le \|y\|^2 \le (1+\varepsilon)\|f\|^2 \right) > 1-\delta$$

$$= \mathbb{P}\left(\left| \|y\|^2 - F_2 \right| \le \varepsilon F_2 \right) > 1-\delta.$$

To generate a streaming algorithm out of this, notice

$$f = \sum_{j=1}^{m} x_j \quad \text{where} \quad x_j = (0, \dots, 0, \underbrace{1}_{s_j^{\text{th}} \text{ position}}, 0, \dots 0)$$

By linearity $y = Lf = \sum_{j=1}^{m} Lx_j = L_{s_j}$ where L_k is the k^{th} row of L.

So the algorithm here is generate $L = \left[\text{Gaussian}(0, \frac{1}{n})\right]^{k \times n}$ matrix and $y = 0$.

On stream element s_j, ~~adds~~ $y \leftarrow y + L_{s_j}$.

Return $\|y\|$.

* The downside to this method which has the same asymptotic complexity to the Tug-of-War Alg. seen in the first lectures is that we store $y \in \mathbb{R}^k$, so we would need to also pay attention to the decimal precision needed.

<u>Hoeffding's Inequality</u>

<u>Thm</u>. If $X_1, \ldots, X_n$ are independent r.v. such that $X_i \in [a_i, b_i]$ a.s. (i.e. $\Pr(a_i \le X_i \le b_i) = 1$) then for $S = \sum_i X_i$,

$$\Pr\left(S - E(S) \ge t\right) \le \exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right)$$

<u>Hoeffding's Lemma</u> Let X be a r.v. with $E(X) = 0$ and $a \le X \le b$ a.s. Then for all $\lambda \in \mathbb{R}$,

$$E\left(e^{\lambda X}\right) \le \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right).$$

Pf. of Lemma

As $x \mapsto e^{\lambda x}$ is convex, for $a \le x \le b$,

$$e^{\lambda x} \le \left(\frac{b-x}{b-a}\right) e^{\lambda a} + \left(\frac{x-a}{b-a}\right) e^{\lambda b}$$

$$\Rightarrow \quad E(e^{\lambda X}) \le \left(\frac{b-E(X)}{b-a}\right) e^{\lambda a} + \left(\frac{E(X)-a}{b-a}\right) e^{\lambda b}$$

$$= \left(\frac{b}{b-a}\right) e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}$$

Let $h = \lambda(b-a)$, $p = \frac{-a}{b-a}$, and $L(h) = -hp + \ln(1-p+pe^h)$

Then we can see

$E(e^{\lambda X}) \le e^{L(h)}$. Taylor expand $L(h)$ to get that $\left(\begin{array}{l} L(0)=0 \\ L'(0)=0 \\ L''(h) \le \frac{1}{4} \ \forall h \end{array}\right)$

$$L(h) \le \frac{1}{8} h^2 = \frac{1}{8}\lambda^2 (b-a)^2.$$

$$\Rightarrow \quad E(e^{\lambda X}) \le \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right). \qquad \square$$

Pf of Thm

For any r.v. X, $X - E(X)$ has mean 0. And if $a \le X \le b$ a.s.

$a - E(X) \le X - E(X) \le b - E(X)$ a.s. So applying the Lemma,

$$E\left(e^{\lambda(X-E(X))}\right) \le \exp\left(\frac{\lambda^2\left((b-E(X))-(a-E(X))\right)^2}{8}\right) = \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Assume $X_1, \dots, X_n$ indep. r.v. with $a_i \le X_i \le b_i$ a.s. and let

$S = \sum_i X_i$. Then for $s, t \ge 0$

$$\Rightarrow \Pr\left(S - E(S) \ge t\right) = \Pr\left(e^{s(S_n - E(S_n))} \ge e^{st}\right)$$

$$\le \frac{1}{e^{st}} E\left(e^{s(S_n - E(S_n))}\right) \qquad \text{(Markov's Ineq.)}$$

$$= \frac{1}{e^{st}} \prod_i E\left(e^{s(X_i - E(X_i))}\right)$$

$$\le \frac{1}{e^{st}} \prod_i \exp\left(\frac{s^2 (b_i - a_i)^2}{8}\right)$$

$$= \exp\left(-st + \frac{1}{8} s^2 \sum_{i=1}^{n} (b_i - a_i)^2\right)$$

Define $C := \sum_i (b_i - a_i)^2$. Then the minimizing s satisfies

$$0 = \frac{d}{ds}\left(-st + \frac{C}{8} s^2\right) = -t + \frac{C}{4} s \quad \Rightarrow \quad s = \frac{4t}{C}.$$

$$\Rightarrow \Pr\left(S - E(S) \ge t\right) \le \exp\left(\frac{-4t^2}{C} + \frac{C}{8} \frac{16t^2}{C^2}\right)$$

$$= \exp\left(\frac{-2t^2}{C}\right) = \exp\left(\frac{-2t^2}{\sum_i (b_i - a_i)^2}\right).$$