**CS 761: Randomized Algorithms** Fall 2019

# Lecture 5 — October 4, 2019

*Prof. Gautam Kamath* *By: Xinyu Liu, Shiqi Xiao*
*Edited by Vedat Levi Alev*

**Disclaimer:** These notes have not been subject to the usual scrutiny reserved for formal publications.

Recall from last lecture:

Given $(id, count)$ pairs $(a_i, l_i)$ where $a_i \in [m]$ for $i = 1...n$ one at a time, we want to answer questions about stream, with limited space $\mathcal{O}(poly \log(m, n))$ ideally.

We have seen

- Heavy Hitters: Count min sketch;
- Distinct elements problem.

## 1 Streaming

Let $x_i = \sum_{j:i=a_j} l_j$ where $l_j \geq 0$ indicate the final count for symbol $i$ in a stream, we define the $p^{th}$ frequency moment be $F_p = \sum_{i\in[m]} x_i^p$.

There are some classic problems related to specific $p^{th}$ frequency moment, for example:

- $p = 0$: distinct element problem;
- $p = 1$: trivial (count elements problem);
- $p = 2$: measure of skewness.

If we want to get the explicit $F_p$, we will need at least linear space. In order to do it in log space, we have to sacrifice accuracy.

Let $\hat{F}_p$ be an estimate of $F_p$, we want to find an $\hat{F}_p$ which is a $(1 \pm \epsilon)$ approximation to $F_p$ with probability at least $1 - \delta$.

### 1.1 Second Frequency Moment $F_2$ Estimation

Here we introduce Alon-Matias-Szegedy algorithm for approximate $F_2$ in log space.

**Algorithm 1** (Alon-Matias-Szegedy)**.**

*Input: number of symbols $m$, length of stream $n$, and the stream.*

*Output: $\hat{F}_2$ (i.e. the estimate of $F_2$).*

1. *Choose Rademacher random variables $r_i$ i.i.d for each $i = 1, \ldots, m$, i.e.* $\Pr[r_i = \pm 1] = 1/2$*;*
2. *Initialize $Z = 0$;*
3. *While processing the stream, for each pair $(a_j, l_j)$, update $Z = Z + r_i l_j$ where $i = a_j$.*
4. *Output $\hat{F}_2 = Z^2$.*

**Remark 2.** *Using 4-wise independent random variables as $r_i$, we will need $\mathcal{O}(\log m)$ space.*

**Remark 3.** *Updating $Z$ during the process to maintain a partial sum of what we have seen so far, so we do not need to count every $x_i = \sum_{j:i=a_j} l_j$ and store all of $\{x_i | i = 1, ..., m\}$ in linear space.*

**Example 4.** *Assume $m = 3$, $r_1 = 1, r_2 = 1, r_3 = -1$, and stream is $(3,1), (1,1), (2,1), (1,1), (2,1), (1,1), (1,1)$. Calculate the explicit $F_2$ and estimate $\hat{F}_2$ using Alon-Matias-Szegedy algorithm.*

$F_2 = 4^2 + 2^2 + 1^2 = 21$.

$Z = \sum_{j=1}^{n} r_{a_j} l_j = 5$, *hence* $\hat{F}_2 = Z^2 = 25$.

Let us consider the expected value and variance of $\hat{F}_2$.

**Claim 5.** $E[\hat{F}_2] = F_2$.

*Proof.* Note that

$$E[r_i r_j] = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

then

$$\begin{aligned} E[\hat{F}_2] &= E[(\sum_{i,j:i=a_j} r_i l_j)^2] \\ &= E[(\sum_i r_i \sum_{j:i=a_j} l_j)^2] \\ &= E[(\sum_i r_i x_i)^2] \\ &= \sum_{1 \le i,j \le m} x_i x_j E[r_i r_j] \\ &= \sum_i x_i^2 \\ &= F_2. \end{aligned}$$

Thus, $\hat{F}_2$ is unbiased. ∎

**Claim 6.** $Var[\hat{F}_2] \le 2(\sum_i x_i^2)^2 = 2F_2^2$.

*Proof.* Note that

$$E[\hat{F_2}^2] = \begin{cases} 0, & \text{if one symbol appears only once} \\ \sum_i x_i^4, & \text{if } i = j = k = l \\ 6\sum_{i \ne j} x_i^2 x_j^2, & \text{if } i = j, k = l \text{ (or some other 2 pairs match)} \end{cases}, \tag{2}$$

then

$$
\begin{aligned}
E[\hat{F_2}^2] &= E[Z^4] \\
&= \sum_{1 \le i,j,k,l \le m} x_i x_j x_k x_l E[r_i r_j r_k r_l] \\
&= \sum_i x_i^4 + 6 \sum_{i \ne j} x_i^2 x_j^2.
\end{aligned}
$$

So

$$
\begin{aligned}
Var[\hat{F_2}] &= E[\hat{F_2}^2] - E[\hat{F_2}]^2 \\
&= E[Z^4] - E[Z^2]^2 \\
&= (\sum_i x_i^4 + 6 \sum_{i \ne j} x_i^2 x_j^2) - (\sum_i x_i^2)^2 \\
&= \sum_i x_i^4 + 6 \sum_{i \ne j} x_i^2 x_j^2 - (\sum_i x_i^4 + 2 \sum_{i \ne j} x_i^2 x_j^2) \\
&= 4 \sum_{i \ne j} x_i^2 x_j^2 \\
&\le 2(\sum_i x_i^2)^2 \\
&= 2(E[\hat{F_2}])^2 \\
&= 2F_2^2.
\end{aligned}
$$

□

Hence, we have a bounded variance, which can be used in concentrating the mean by Chebyshev's inequality.

**Claim 7.** $\Pr[\hat{F_2} \in (1 \pm c\sqrt{2})F_2] \ge 1 - \frac{1}{c^2}$.

*Proof.* By Chebyshev's inequality,

$$
\begin{aligned}
Pr[|\hat{F_2} - E[\hat{F_2}]| \ge c\sqrt{Var[\hat{F_2}]}] &= Pr[|\hat{F_2} - F_2| \ge c\sqrt{2}F_2] \\
&\le \frac{1}{c^2}.
\end{aligned}
$$

□

**Remark 8.** *In order to further reduce the variance hence increase sufficiency, we could apply the idea of bootstrap aggregation to Algorithm 1:*

1. *Repeat the Alon-Matias-Szegedy algorithm $k$ times in parallel to obtain $\hat{F_2}^{(1)}, ..., \hat{F_2}^{(k)}$;*
2. *Output $\hat{F_2}' = \frac{1}{k} \sum_{i=[k]} \hat{F_2}^{(i)}$.*

Let us look at the expected value and variance of $\hat{F_2}'$ from the modified algorithm.

**Claim 9.** $E[\hat{F_2}'] = F_2$.

*Proof.*

$$\begin{aligned} E[\hat{F_2}'] &= \sum_{i=1}^{k} \frac{E[\hat{F_2}^{(i)}]}{k} \\ &= E[\hat{F_2}] \\ &= F_2. \end{aligned}$$

Thus, $\hat{F_2}'$ is unbiased. □

**Claim 10.** $Var[\hat{F_2}'] = \frac{1}{k} Var[\hat{F_2}]$.

*Proof.*

$$\begin{aligned} Var[\hat{F_2}'] &= \frac{1}{k^2} k Var[\hat{F_2}] \\ &= \frac{1}{k} Var[\hat{F_2}]. \end{aligned}$$

□

Here we can verify that $\hat{F_2}'$ is a $(1 \pm \epsilon)$ approximation to $F_2$ with probability at least $1 - \delta$ by Chernoff's inequality.

**Claim 11.** $Pr[\hat{F_2}' \in (1 \pm \epsilon) F_2] \geq 1 - \delta$.

*Proof.* By Chernoff's inequality, we have

$$\begin{aligned} Pr[|\hat{F_2}' - F_2| \geq \epsilon F_2] &\leq \frac{Var[\hat{F_2}']}{\epsilon^2 F_2^2} \\ &= \frac{\frac{1}{k} Var[\hat{F_2}]}{\epsilon^2 F_2^2}. \end{aligned} \quad (3)$$

Let $k \in \mathcal{O}(\frac{1}{\delta \epsilon^2})$, then equation 3 becomes

$$\begin{aligned} Pr[|\hat{F_2}' - F_2| \geq \epsilon F_2] &\leq \frac{\frac{1}{k} 2 F_2^2}{\epsilon^2 F_2^2} \\ &= \frac{\delta \epsilon^2}{\epsilon^2} \\ &= \delta. \end{aligned}$$

Hence, $Pr[\hat{F_2}' \in (1 \pm \epsilon) F_2] \geq 1 - \delta$. □

**Claim 12.** *Remark 8 needs* $\mathcal{O}(\frac{1}{\delta \epsilon^2}(\log(m) + \log(n)))$ *space.*

**Observation 13.** *Another view of this problem:*

Let $S \in R^{k \times m}$ where $\Pr[S_{ij} = \pm 1] = \frac{1}{2}$, and let $x = \sum_{j \in [n]} l_j \cdot e_j$ where $l_j$ is the $j^{th}$ element and it is from pair $(a_j, l_j)$, then $\hat{F}_2 = Sx = [z_1, \ldots, z_k]^\top$.

We have

$$\begin{aligned} \hat{F_2}' &= \frac{\sum_{j \in [k]} \hat{F_2}^{(j)}}{k} \\ &= \frac{\sum_{j \in [k]} z_j^2}{k} \\ &= \frac{\|Sx\|_2^2}{k} \\ &= (1 \pm \epsilon)\|x\|_2^2. \end{aligned}$$

Hence, $\hat{F}_2$ could be approximated by calculating the L2-norm of $x$.

## 1.2 Dimension Reduction

Given $t$ points $x_1, ..., x_t \in R^m$, we want to reduce dimension to $k$ where $k \ll m$ while keeping the pairwise distance.

**Observation 14.** *When $t = 2$, we have points $x_1, x_2$ with dimension $m$, and $Sx_1, Sx_2$ with dimension $k = \mathcal{O}(\frac{1}{\delta \epsilon^2})$.*

*Then with probability at least $1 - \delta$, we have*

$$\begin{aligned} \frac{\|Sx_1 - Sx_2\|_2^2}{k} &= \frac{\|S(x_1 - x_2)\|_2^2}{k} \\ &= (1 \pm \epsilon)\|x_1 - x_2\|_2^2. \end{aligned}$$

**Observation 15.** *For a general $t$, we have points $x_1, ..., x_t$ with dimension $m$, and $Sx_1, ..., Sx_t$ with dimension $k = \mathcal{O}(\frac{1}{\delta \epsilon^2})$.*

*Then with probability at least $1 - \delta$, for any $Sx_i, Sx_j$ where $1 \le i, j \le t$, we have*

$$\begin{aligned} \|Sx_i, Sx_j\|^2 &= \frac{\|S(x_i - x_j)\|_2^2}{k} \\ &\in (1 \pm \epsilon)\|x_i - x_j\|_2^2 \end{aligned}$$

*where $k = \mathcal{O}(\frac{1}{\delta \epsilon^2})$, and $\delta \le \frac{\delta'}{\binom{t}{2}}$, so $k = \mathcal{O}(\frac{t^2}{\delta' \epsilon^2})$ .*

However, $k = t - 1$ is trivial and exact, so this $k = \mathcal{O}(\frac{t^2}{\delta' \epsilon^2})$ blows up dimension instead of reducing dimension. In fact, we are able to approximate $F_2$ in a much smaller dimension given the Johnson–Lindenstrauss Lemma.

**Lemma 16** (Johnson–Lindenstrauss Lemma)**.** *Given $0 < \epsilon < 1$, a point $x \in R^m$, and a number $k = \mathcal{O}(\frac{\log(\frac{1}{\delta})}{\epsilon^2})$. Let $S \in R^{k \times m}$ where $S_{ij} \sim N(0, 1)$, then we have $\frac{\|Sx\|_2^2}{k} \approx (1 \pm \epsilon)\|x\|_2^2$.*

*Proof.* This proof is heavily related to Gaussian distribution, let us review some useful properties first.

**Fact 17** (Gaussian distribution $N(\mu, \sigma^2)$)**.** *The probability distribution function is* $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$.

*Properties:*

1. *If* $G_1 \sim N(\mu_1, \sigma_1^2), G_2 \sim N(\mu_2, \sigma_2^2)$*, then* $G_1 + G_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
2. *If* $G \sim N(0, 1)$*, then* $\sigma G \sim N(0, \sigma^2)$.

Define $Y' = \frac{\|Sx\|_2^2}{k}$. We want to obtain an expression of Gaussian distribution for it.

Let us first focus on each element of vector $Sx$, denote as $z_j$, we have

$$\begin{aligned} z_j &= [Sx]_j \\ &= \sum_i x_i S_{ij} \\ &= N(0, \sum_i x_i^2) \\ &= \|x\|_2^2 N(0, 1). \end{aligned}$$

Then we can rewrite $Y'$ as

$$\begin{aligned} Y' &= \frac{\|Sx\|_2^2}{k} \\ &= \frac{\sum z_j^2}{k} \\ &= \frac{\sum(\|x\|_2 G_j)^2}{k} \\ &= \|x\|_2^2 \frac{\sum G_j^2}{k} \end{aligned} \tag{4}$$

where $G_1, ..., G_m \sim N(0, 1)$ i.i.d..

By definition of Gaussian distribution, we have $E[G_j^2] = Var[G_i] + E[G_j]^2 = 1$, thus the expectation of $Y'$ is

$$\begin{aligned} E[Y'] &= \|x\|_2^2 E\left[\frac{\sum_j G_j^2}{k}\right] \\ &= \|x\|_2^2. \end{aligned}$$

Hence, the expectation of $Y'$ is as desired. Next, let us prove $Y'$ belongs to the $\epsilon$-bounds of $\|x\|_2^2$.

Define $Y = \frac{\sum G_j^2}{k}$, we want to prove $Y \in (1 \pm \epsilon)$ with probability $1 - \delta$.

Using Chernoff bound,

$$\begin{aligned} \Pr[Y \geq 1 + \epsilon] &= \Pr[e^{tkY} \geq e^{tk(1+\epsilon)}] \\ &\leq \frac{E[e^{tkY}]}{e^{tk(1+\epsilon)}} \\ &= \prod_{i \in [k]} \left( \frac{E[e^{tG_j^2}]}{e^{t(1+\epsilon)}} \right). \end{aligned} \tag{5}$$

Calculate the expected value of $e^{tG_j^2}$ as

$$\begin{aligned} E[e^{tG_j^2}] &= \frac{1}{\sqrt{2\pi}} \int e^{tu^2} e^{-\frac{u^2}{2}} du \\ &= \frac{1}{\sqrt{1-2t}} \end{aligned} \tag{6}$$

for $t < \frac{1}{2}$.

Substitute equation 6 into equation 5:

$$\begin{aligned} \Pr[Y \geq 1+\epsilon] &\leq \prod_{i \in [k]} \left( \frac{E[e^{tG_j^2}]}{e^{t(1+\epsilon)}} \right) \\ &= \left( \frac{\frac{1}{\sqrt{1-2t}}}{e^t e^{t\epsilon}} \right)^k \\ &\leq \left( \frac{1}{e^t \sqrt{1-2t}} \right)^k \frac{1}{e^{t\epsilon k}}. \end{aligned} \tag{7}$$

We can rewrite $\frac{1}{e^t\sqrt{1-2t}}$ as an exponential expression

$$\begin{aligned} \frac{1}{e^t \sqrt{1-2t}} &= e^{-t-\frac{1}{2}\log(1-2t)} \\ &= \exp(t^2 + \mathcal{O}(t^3)). \end{aligned} \tag{8}$$

Let $t = \Theta(\epsilon)$, and substitute equation 8 into equation 7:

$$\begin{aligned} \Pr[Y \geq 1+\epsilon] &\leq \left( \frac{1}{e^t \sqrt{1-2t}} \right)^k \frac{1}{e^{t\epsilon k}} \\ &\leq \exp(kt^2 + \mathcal{O}(t^3 k) - t\epsilon k) \\ &\leq \exp(-\Theta(k\epsilon^2)). \end{aligned} \tag{9}$$

By symmetry, we know $\Pr[Y \notin 1 \pm \epsilon] \leq 2\exp(-\Theta(k\epsilon^2))$.

Hence, substituting (4) gives

$$\begin{aligned} \Pr[Y' \in (1\pm\epsilon)\|x\|_2^2] &= \Pr[\|x\|_2^2 \frac{\sum G_j^2}{k} \in (1\pm\epsilon)\|x\|_2^2] \\ &= \Pr\left[ \frac{\sum G_j^2}{k} \in (1\pm\epsilon) \right] \\ &\geq 1 - 2\exp(-\Theta(k\epsilon^2)). \end{aligned}$$

Let $k = \mathcal{O}(\frac{\log(\frac{1}{\delta})}{\epsilon^2})$, we have $\Pr[Y' \in (1\pm\epsilon)\|x\|_2^2] \geq 1-\delta$ as desired. □

**Corollary 18.** *Let $S \in R^{k \times m}$ be a matrix, where each element $S_{ij} \sim N(0,1)$, then*

$$\text{for any } \vec{x}_1, \ldots, \vec{x}_t \in \mathbb{R}^m, \quad \Pr\left[\left\|\frac{S\vec{x}_i - S\vec{x}_j}{\sqrt{k}}\right\|_2^2 = (1 \pm \epsilon) \cdot \|\vec{x}_i - \vec{x}_j\|_2^2\right] = 1 - \delta$$

*for all points simultaneously if $k \geq \mathcal{O}(\frac{\log(t/\delta)}{\epsilon^2})$.*

There is another way to view the Johnson–Lindenstrauss Lemma, we could consider number-medians and distribution-medians instead.

**Claim 19.** *Given $S_{ij} \sim N(0,1), Sx = [z_1, \ldots, z_k]^\top$ where $k$ is a constant, and $z_j \sim N(0, \|x\|_2^2)$, then $\hat{F_2}' = \frac{z_1^2 + \ldots + z_k^2}{k} \to \|x\|_2^2$.*

*Proof.* In this proof, we need to use the relationships between the cumulative distribution function (CDF) and the median of a distribution. First, we introduce Lemma 20 and 21 to help complete this proof.

**Lemma 20.** *Let $u_1, ..., u_k$ be i.i.d. random variables with cumulative distribution function $F$ and median $m$. Let $u = median(u_i, ..., u_k)$, then*

$$\Pr\left[F(u) = \left(\frac{1}{2} \pm \epsilon\right)\right] \geq 1 - \exp(-\Theta(\epsilon^2 k)).$$

*Proof.* Let event $E_i$ represent $F(u_i) < \frac{1}{2} - \epsilon$, then $\Pr[E_i] = \frac{1}{2} - \epsilon$. We have $F(u) < \frac{1}{2} - \epsilon$ if and only if more than $\frac{k}{2}$ of $E_i$'s hold.

Using Chernoff bound,

$$\Pr[F(u) < \frac{1}{2} - \epsilon] \approx \exp(-\Theta(\epsilon^2 k)).$$

Symmetrically,

$$\Pr\left[F(u) = \left(\frac{1}{2} \pm \epsilon\right)\right] \geq 1 - \exp(-\Theta(\epsilon^2 k)).$$

□

**Lemma 21.** *Let $F$ be the CDF of $|G|$ where $G \sim N(0,1)$. If $F(z) = \frac{1}{2} \pm \epsilon$, then $z \in median(|G|) \pm \Theta(\epsilon)$.*

Then we start the formal proof. Define

$$\hat{F_2} = \frac{number - median(|z_1|, ..., |z_k|)}{distribution - median(|G|)}, G \sim N(0,1). \tag{10}$$

Since $|z_j| = \|x\|_2 |G_j|, G_j \sim N(0,1)$, equation 10 could be written as

$$\hat{F_2} = \|x\|_2 \frac{median(|G_1|, ..., |G_k|)}{median(|G|)}. \tag{11}$$

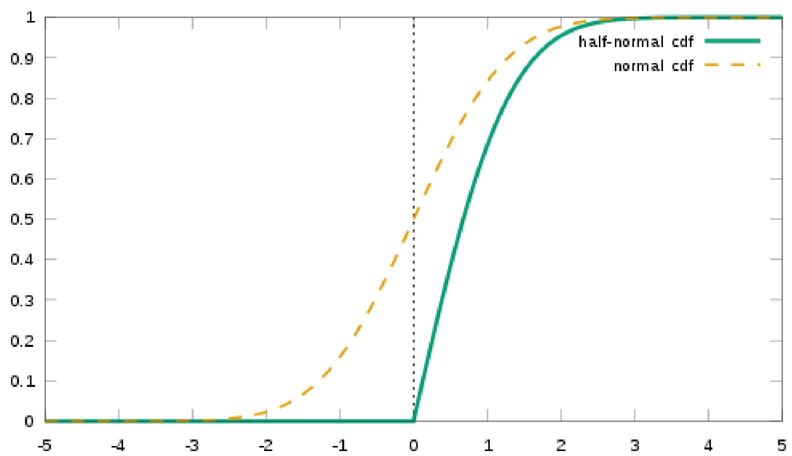


Figure 1: CDF of G

Let $u = median(|G_1|, ..., |G_k|)$, apply Lemma 20 to get $F(u) = \frac{1}{2} \pm \epsilon$ with high probability. Then from Lemma 21, we have $u \in median(|G|) \pm \Theta(\epsilon)$.

Substitute the result to equation 11:

$$\begin{aligned} Y &= \|x\|_2 \frac{median(|G_1|, ..., |G_k|)}{median(|G|)} \\ &= \|x\|_2 \frac{median(|G|) \pm \Theta(\epsilon))}{median(|G|)} \\ &= \|x\|_2 \left(1 \pm \Theta\left(\frac{\epsilon}{median(|G|)}\right)\right) \\ &= \|x\|_2 (1 \pm \epsilon). \end{aligned}$$

□

**Definition 22.** *Distance $D$ is p-stable if for $i = 1...k$, each $D_i \sim D$ satisfies $\sum_i x_i D_i \sim \|x\|_p D$ where $x = [x_1, \ldots, x_k]^\top$ and $x_1, \ldots, x_k \in R$. Note that this property only holds for $p \in (0, 2]$.*

**Remark 23.** *$\hat{F}_p \in (1 \pm \epsilon) F_p$ for $p \in (0, 2]$ with $\mathcal{O}(poly \log(m, n))$ space.*

## 1.3 Generalized Frequency Moment $F_p$ Estimation

In this section, we will talk about estimating the generalized frequency moment $F_p$.

First, let us look at an algorithm which will go through the entire streaming two times while estimating $F_p$.

**Algorithm 24** (2 Pass)**.**

1. *In the first pass, pick $j \in [1, n]$ uniformly at random, wait until observe $i = a_j$;*
2. *In the second pass, compute $x_i$ = number of times item with label $i$ is seen;*

*3. Return* $\hat{F}_p = nx_i^{p-1}$.

Let us discuss the expected value and variance of $\hat{F}_p$ from the above algorithm.

**Claim 25.** $E[\hat{F}_p] = F_p$.

*Proof.*

$$\begin{aligned} E[\hat{F}_p] &= \sum_{i=1}^{m} \frac{x_i}{n}(nx_i^{p-1}) \\ &= \sum_{i=1}^{m} x_i^p \\ &= F_p. \end{aligned}$$

Thus, $\hat{F}_p$ is unbiased. □

**Claim 26.** $Var[\hat{F}_p] = nF_{2p-1}$.

*Proof.*

$$\begin{aligned} Var[\hat{F}_p] &\leq E[\hat{F}_p^{\,2}] \\ &= \sum_{i=1}^{m} \frac{x_i}{n}(nx_i^{p-1})^2 \\ &= nF_{2p-1}. \end{aligned}$$

Hence, we have bounded the variance, which can be used in concentrating the mean by Chebyshev's inequality. □

**Claim 27.** $nF_{2p-1} \leq m^{1-\frac{1}{p}}(F_p)^2$.

**Claim 28.** *This algorithm is* $\mathcal{O}(\frac{m^{1-\frac{1}{p}}}{\epsilon^2})$ *sufficient.*

*Proof.* Using Chebyshev's inequality,

$$\begin{aligned} \Pr\left[\frac{|\hat{F}_p - F_p|}{F_p} > \epsilon\right] &= \Pr[|\hat{F}_p - F_p| > \epsilon F_p] \\ &\leq \frac{Var[\hat{F}_p]}{\epsilon^2 F_p^2} \\ &\leq \frac{m^{1-\frac{1}{p}}}{\epsilon^2}. \end{aligned}$$

□

The 2 pass algorithm above can also be performed in the same pass, at the cost of accuracy.

**Algorithm 29** (1 Pass)**.**

1. *Pick* $j \in [1, n]$ *uniformly at random, wait until observe* $i = a_j$*;*
2. *In the rest of the pass, compute* $x'_i$ = *number of times label* $i$ *is seen after* $a_j$ *(inclusive);*
3. *Return* $\hat{F_p}' = n(x'^p_i - (x'_i - 1)^p)$*.*

Again, we will take a look at its expected value and variance.

**Claim 30.** $E[\hat{F_p}'] = \sum_{i=1}^m x_i^p$

*Proof.*

$$\begin{aligned}
E[\hat{F_p}'] &= nE[x'^p_i - (x'_i - 1)^p] \\
&= n\frac{1}{n}\sum_{i=1}^{m}\sum_{l=1}^{x_i}(l^p - (l-1)^p) \\
&= \sum_{i=1}^{m} x_i^p \\
&= F_p.
\end{aligned}$$

Thus, $\hat{F_2}'$ is unbiased. □

**Claim 31.** $Var[\hat{F_p}'] \leq p^2 Var[\hat{F_p}]$

*Proof.* Since

$$\begin{aligned}
\hat{F_p}' &= n(x'^p_i - (x'_i - 1)^p) \\
&\leq np(x'_i)^{p-1} \\
&\leq p(nx_i^{p-1}) \\
&= p\hat{F_p},
\end{aligned}$$

we have

$$\begin{aligned}
Var[\hat{F_p}'] &\leq p^2 Var[\hat{F_p}] \\
&\leq p^2 m^{1-\frac{1}{p}} F_p^2.
\end{aligned}$$

□

**Claim 32.** *This algorithm is* $\mathcal{O}(p^2 \frac{m^{1-\frac{1}{p}}}{\epsilon^2})$ *sufficient.*

*Proof.* See Proof for Claim 28. □