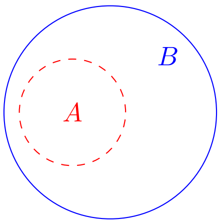
Advanced Algorithms 6 November 2018

Lecture 8: Streaming and Sketching Algorithms II

Lecturer: Mohsen Ghaffari *Scribe: Davin Choo*

Recall that the k^{th} moment of a stream S is defined as $\sum_{j=1}^{n}(f_j)^k$. In this lecture, we will continue the analysis for estimating the zeroth moment of a stream, and show an algorithm that estimates the k^{th} moment of a stream, due to [AMS96]. We will see how Tricks 1 and 2 from the previous lecture can be used to improve the estimation precision and amplify the success probabilities in our analysis.

Remark In this lecture, we will often upper-bound probabilities using the following fact: If event A implies event B, then $\Pr[A] \le \Pr[B]$. One can visualize the probability space as follows:



1 Estimating the zeroth moment of a stream (Continued)

Recall the definition of pairwise independent hash functions and the algorithm presented at the end of the last lecture (Algorithm 1 due to [FM85]). Let D be the number of distinct elements in the stream S.

Definition 1 (Family of pairwise independent hash functions)**.** *$\mathcal{H}_{n,m}$ is a family of pairwise independent hash functions if*

- *(Hash definition):* $\forall h \in \mathcal{H}_{n,m}$, $h : \{1, \dots, n\} \to \{1, \dots, m\}$
- *(Uniform hashing):* $\forall x \in \{1, \dots, n\}$, $\Pr_{h \in \mathcal{H}_{n,m}}[h(x) = i] = \frac{1}{m}$
- *(Pairwise independent)* $\forall x, y \in \{1, \dots, n\}$, $x \neq y$, $\Pr_{h \in \mathcal{H}_{n,m}}[h(x) = i \wedge h(y) = j] = \frac{1}{m^2}$

Algorithm 1 FM($S = \{a_1, \dots, a_m\}$)

```
h ← Random hash from H_{n,n}
Z ← 0
for a_i ∈ S do                                    ▷ Items arrive in streaming fashion
    Z = max{Z, ZEROS(h(a_i))}    ▷ ZEROS(h(a_i)) = # trailing zeroes in binary representation of h(a_i)
end for
return 2^Z · √2                                   ▷ Estimate of D
```

Since the hash h is deterministic after picking a random hash from $\mathcal{H}_{n,n}$, $h(a_i) = h(a_j), \forall a_i = a_j \in [n]$.

Lemma 2. *If $X_1, \dots, X_n$ are pairwise independent indicator random variables and $X = \sum_{i=1}^{n} X_i$, then $Var(X) \le \mathbb{E}[X]$.*

Proof.

$$\begin{aligned}
\mathrm{Var}(X) &= \textstyle\sum_{i=1}^{n} \mathrm{Var}(X_i) && \text{The } X_i\text{'s are pairwise independent} \\
&= \textstyle\sum_{i=1}^{n} (\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2) && \text{Definition of variance} \\
&\le \textstyle\sum_{i=1}^{n} \mathbb{E}[X_i^2] && \text{Ignore negative part} \\
&= \textstyle\sum_{i=1}^{n} \mathbb{E}[X_i] && X_i^2 = X_i \text{ since } X_i\text{'s are indicator random variables} \\
&= \mathbb{E}[\textstyle\sum_{i=1}^{n} X_i] && \text{Linearity of expectation} \\
&= \mathbb{E}[X] && \text{Definition of expectation}
\end{aligned}$$

□

Theorem 3. *There exists a constant $C > 0$ such that* $\Pr[\frac{D}{3} \leq 2^Z \cdot \sqrt{2} \leq 3D] > C$.

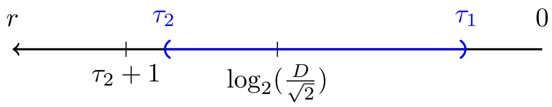
Proof. We will prove $\Pr[(\frac{D}{3} > 2^Z \cdot \sqrt{2}) \text{ or } (2^Z \cdot \sqrt{2} > 3D)] \leq 1 - C$ by separately analyzing $\Pr[\frac{D}{3} \geq 2^Z \cdot \sqrt{2}]$ and $\Pr[2^Z \cdot \sqrt{2} \geq 3D]$, then applying union bound. Define indicator variables

$$X_{i,r} = \begin{cases} 1 & \text{if } \textsc{zeros}(h(a_i)) \geq r \\ 0 & \text{otherwise} \end{cases}$$

and $X_r = \sum_{i=1}^{m} X_{i,r} = |\{a_i \in S : \textsc{zeros}(h(a_i)) \geq r\}|$. Notice that $X_n \leq X_{n-1} \leq \cdots \leq X_2 \leq X_1$ since $\textsc{zeros}(h(a_i)) \geq r + 1 \Rightarrow \textsc{zeros}(h(a_i)) \geq r$. Now,

$$\begin{aligned}
\mathbb{E}[X_r] &= \mathbb{E}[\textstyle\sum_{i=1}^{m} X_{i,r}] && \text{Since } X_r = \textstyle\sum_{i=1}^{m} X_{i,r} \\
&= \textstyle\sum_{i=1}^{m} \mathbb{E}[X_{i,r}] && \text{By linearity of expectation} \\
&= \textstyle\sum_{i=1}^{m} \Pr[X_{i,r} = 1] && \text{Since } X_{i,r} \text{ are indicator variables} \\
&= \textstyle\sum_{i=1}^{m} \frac{1}{2^r} && \text{Since } h \text{ is a uniform hash — } r \text{ zeros in coin flips} \\
&= \frac{D}{2^r} && \text{Since } h \text{ hashes same elements to the same value}
\end{aligned}$$

Denote τ_1 as the *smallest integer* such that $2^{\tau_1} \cdot \sqrt{2} > 3D$, and τ_2 as the *largest integer* such that $2^{\tau_2} \cdot \sqrt{2} < \frac{D}{3}$. We see that if $\tau_1 < Z < \tau_2$, then $2^Z \cdot \sqrt{2}$ is a 3-approximation of D.



- If $Z \geq \tau_1$, then $2^Z \cdot \sqrt{2} \geq 2^{\tau_1} \cdot \sqrt{2} > 3D$
- If $Z \leq \tau_2$, then $2^Z \cdot \sqrt{2} \leq 2^{\tau_2} \cdot \sqrt{2} < \frac{D}{3}$

$$\begin{aligned}
\Pr[Z \geq \tau_1] &\leq \Pr[X_{\tau_1} \geq 1] && \text{Since } Z \geq \tau_1 \Rightarrow X_{\tau_1} \geq 1 \\
&\leq \frac{\mathbb{E}[X_{\tau_1}]}{1} && \text{By Markov's inequality} \\
&= \frac{D}{2^{\tau_1}} && \text{Since } \mathbb{E}[X_r] = \frac{D}{2^r} \\
&\leq \frac{\sqrt{2}}{3} && \text{Since } 2^{\tau_1} \cdot \sqrt{2} > 3D
\end{aligned}$$

$$\begin{aligned}
\Pr[Z \leq \tau_2] &\leq \Pr[X_{\tau_2+1} = 0] && \text{Since } Z \leq \tau_2 \Rightarrow X_{\tau_2+1} = 0 \\
&\leq \Pr[\mathbb{E}[X_{\tau_2+1}] - X_{\tau_2+1} \geq \mathbb{E}[X_{\tau_2+1}]] && \text{Implied} \\
&\leq \Pr[|X_{\tau_2+1} - \mathbb{E}[X_{\tau_2+1}]| \geq \mathbb{E}[X_{\tau_2+1}]] && \text{Adding absolute sign} \\
&\leq \frac{\mathrm{Var}[X_{\tau_2+1}]}{(\mathbb{E}[X_{\tau_2+1}])^2} && \text{By Chebyshev's inequality} \\
&\leq \frac{\mathbb{E}[X_{\tau_2+1}]}{(\mathbb{E}[X_{\tau_2+1}])^2} && \text{By Lemma 2} \\
&\leq \frac{2^{\tau_2+1}}{D} && \text{Since } \mathbb{E}[X_r] = \frac{D}{2^r} \\
&\leq \frac{\sqrt{2}}{3} && \text{Since } 2^{\tau_2} \cdot \sqrt{2} < \frac{D}{3}
\end{aligned}$$

Putting together,

$$\begin{aligned}
\Pr[(\tfrac{D}{3} > 2^Z \cdot \sqrt{2}) \text{ or } (2^Z \cdot \sqrt{2} > 3D)] &\leq \Pr[\tfrac{D}{3} \geq 2^Z \cdot \sqrt{2}] + \Pr[2^Z \cdot \sqrt{2} \geq 3D] && \text{By union bound} \\
&\leq \frac{2\sqrt{2}}{3} && \text{From above} \\
&= 1 - C && \text{For } C = 1 - \frac{2\sqrt{2}}{3} > 0
\end{aligned}$$

□

Although the analysis tells us that there is a small success probability ($C = 1 - \frac{2\sqrt{2}}{3} \approx 0.0572$), one can use t independent hashes and output the mean $\frac{1}{k}\sum_{i=1}^{k}(2^{Z_i} \cdot \sqrt{2})$ (Recall Trick 1). With t hashes, the variance drops by a factor of $\frac{1}{t}$, improving the analysis for $\Pr[Z \leq \tau_2]$. When the success probability $C > 0.5$, one can then call the routine k times independently and return the median (Recall Trick 2).

While Tricks 1 and 2 allows us to strength the success probability C, more work needs to be done to improve the approximation factor from 3 to $(1 + \epsilon)$. To do this, we look at a slight modification of Algorithm 1, due to [BYJK+02].

Algorithm 2 FM+($S = \{a_1, \ldots, a_m\}, \epsilon$)

$N \leftarrow n^3$
$t \leftarrow \frac{c}{\epsilon^2} \in \mathcal{O}(\frac{1}{\epsilon^2})$ ▷ For some constant $c \geq 28$
$h \leftarrow$ Random hash from $\mathcal{H}_{n,N}$ ▷ Hash to a larger space
$T \leftarrow \emptyset$ ▷ Maintain t smallest $h(a_i)$'s
for $a_i \in S$ **do** ▷ Items arrive in streaming fashion
  $T \leftarrow t$ smallest values from $T \cup \{h(a_i)\}$ ▷ If $|T \cup \{h(a_i)\}| \leq t$, take everything
end for
$Z = \max_{t \in T} T$
return $\frac{tN}{Z}$ ▷ Estimate of D

Remark For a cleaner analysis, we treat the *integer* interval $[N]$ as a *continuous* interval in Theorem 4. Note that there may be a rounding error of $\frac{1}{N}$ but this is relatively small and a suitable c can be chosen to make the analysis still work.

Theorem 4. *In* FM+*, for any given* $0 < \epsilon < \frac{1}{2}$, $\Pr[|\frac{tN}{Z} - D| \leq \epsilon D] > \frac{3}{4}$.

Proof. We first analyze $\Pr[\frac{tN}{Z} > (1+\epsilon)D]$ and $\Pr[\frac{tN}{Z} < (1-\epsilon)D]$ separately. Then, taking union bounds and negating yields the theorem's statement.

If $\frac{tN}{Z} > (1+\epsilon)D$, then $\frac{tN}{(1+\epsilon)D} > Z = t^{th}$ smallest hash value, implying that there are $\geq t$ hashes *smaller* than $\frac{tN}{(1+\epsilon)D}$. Since the hash uniformly distributes $[n]$ over $[N]$, for each element a_i,

$$\Pr[h(a_i) \leq \frac{tN}{(1+\epsilon)D}] = \frac{\frac{tN}{(1+\epsilon)D}}{N} = \frac{t}{(1+\epsilon)D}$$

Let $d_1, \ldots, d_D$ be the D distinct elements in the stream. Define indicator variables

$$X_i = \begin{cases} 1 & \text{if } h(d_i) \leq \frac{tN}{(1+\epsilon)D} \\ 0 & \text{otherwise} \end{cases}$$

and $X = \sum_{i=1}^{D} X_i$ is the number of hashes that are *smaller* than $\frac{tN}{(1+\epsilon)D}$. From above, $\Pr[X_i = 1] = \frac{t}{(1+\epsilon)D}$. By linearity of expectation, $\mathbb{E}[X] = \frac{t}{(1+\epsilon)}$. Then, by Lemma 2, $\text{Var}(X) \leq \mathbb{E}[X]$. Now,

$$\begin{aligned}
\Pr[\tfrac{tN}{Z} > (1+\epsilon)D] &\leq \Pr[X \geq t] && \text{Since the former implies the latter} \\
&= \Pr[X - \mathbb{E}[X] \geq t - \mathbb{E}[X]] && \text{Subtracting } \mathbb{E}[X] \text{ from both sides} \\
&\leq \Pr[X - \mathbb{E}[X] \geq \tfrac{\epsilon}{2}t] && \text{Since } \mathbb{E}[X] = \tfrac{t}{(1+\epsilon)} \leq (1-\tfrac{\epsilon}{2})t \\
&\leq \Pr[|X - \mathbb{E}[X]| \geq \tfrac{\epsilon}{2}t] && \text{Adding absolute sign} \\
&\leq \tfrac{\text{Var}(X)}{(\epsilon t/2)^2} && \text{By Chebyshev's inequality} \\
&\leq \tfrac{\mathbb{E}[X]}{(\epsilon t/2)^2} && \text{Since Var}(X) \leq \mathbb{E}[X] \\
&\leq \tfrac{4(1-\epsilon/2)t}{\epsilon^2 t^2} && \text{Since } \mathbb{E}[X] = \tfrac{t}{(1+\epsilon)} \leq (1-\tfrac{\epsilon}{2})t \\
&\leq \tfrac{4}{c} && \text{Simplifying with } t = \tfrac{c}{\epsilon^2} \text{ and } (1-\tfrac{\epsilon}{2}) < 1
\end{aligned}$$

Similarly, if $\frac{tN}{Z} < (1-\epsilon)D$, then $\frac{tN}{(1-\epsilon)D} < Z = t^{th}$ smallest hash value, implying that there are $< t$ hashes *smaller* than $\frac{tN}{(1-\epsilon)D}$. Since the hash uniformly distributes $[n]$ over $[N]$, for each element a_i,

$$\Pr[h(a_i) \leq \frac{tN}{(1-\epsilon)D}] = \frac{\frac{tN}{(1-\epsilon)D}}{N} = \frac{t}{(1-\epsilon)D}$$

Let $d_1, \ldots, d_D$ be the D distinct elements in the stream. Define indicator variables

$$Y_i = \begin{cases} 1 & \text{if } h(d_i) \leq \frac{tN}{(1-\epsilon)D} \\ 0 & \text{otherwise} \end{cases}$$

and $Y = \sum_{i=1}^{D} Y_i$ is the number of hashes that are *smaller* than $\frac{tN}{(1-\epsilon)D}$. From above, $\Pr[Y_i = 1] = \frac{t}{(1-\epsilon)D}$. By linearity of expectation, $\mathbb{E}[Y] = \frac{t}{(1-\epsilon)}$. Then, by Lemma 2, $\text{Var}(Y) \leq \mathbb{E}[Y]$. Now,

$$
\begin{array}{rcll}
\Pr[\frac{tN}{Z} < (1-\epsilon)D] & \leq & \Pr[Y \leq t] & \text{Since the former implies the latter} \\
& = & \Pr[Y - \mathbb{E}[Y] \leq t - \mathbb{E}[Y]] & \text{Subtracting } \mathbb{E}[Y] \text{ from both sides} \\
& \leq & \Pr[Y - \mathbb{E}[Y] \leq -\epsilon t] & \text{Since } \mathbb{E}[Y] = \frac{t}{(1-\epsilon)} \geq (1+\epsilon)t \\
& \leq & \Pr[-(Y - \mathbb{E}[Y]) \geq \epsilon t] & \text{Swap sides} \\
& \leq & \Pr[|Y - \mathbb{E}[Y]| \geq \epsilon t] & \text{Adding absolute sign} \\
& \leq & \frac{\text{Var}(Y)}{(\epsilon t)^2} & \text{By Chebyshev's inequality} \\
& \leq & \frac{\mathbb{E}[Y]}{(\epsilon t)^2} & \text{Since } \text{Var}(Y) \leq \mathbb{E}[Y] \\
& \leq & \frac{(1+2\epsilon)t}{\epsilon^2 t^2} & \text{Since } \mathbb{E}[Y] = \frac{t}{(1-\epsilon)} \leq (1+2\epsilon)t \\
& \leq & \frac{3}{c} & \text{Simplifying with } t = \frac{c}{\epsilon^2} \text{ and } (1+2\epsilon) < 3
\end{array}
$$

Putting together,

$$
\begin{array}{rcll}
\Pr[|\frac{tN}{Z} - D| > \epsilon D]] & \leq & \Pr[\frac{tN}{Z} > (1+\epsilon)D]] + \Pr[\frac{tN}{Z} < (1-\epsilon)D]] & \text{By union bound} \\
& \leq & 4/c + 3/c & \text{From above} \\
& \leq & 7/c & \text{Simplifying} \\
& \leq & 1/4 & \text{For } c \geq 28
\end{array}
$$

□

2 Estimating the k^{th} moment of a stream

In this section, we describe algorithms from [AMS96] that estimates the k^{th} moment of a stream, first for $k = 2$, then for general k. Recall that the k^{th} moment of a stream S is defined as $F_k = \sum_{j=1}^{n} (f_j)^k$.

2.1 $k = 2$

For each element $i \in [n]$, we associate a random variable $r_i \in_{u.a.r.} \{-1, +1\}$.

Algorithm 3 AMS-2($S = \{a_1, \ldots, a_m\}$)

```
For each i ∈ [n], assign r_i ∈_{u.a.r.} {−1, +1}      ▷ For now, this takes O(n) space
Z ← 0
for a_i ∈ S do                                        ▷ Items arrive in streaming fashion
    Z ← Z + r_i                                        ▷ At the end, Z = Σ_{i=1}^n r_i f_i
end for
return Z^2                                            ▷ Estimate of F_2 = Σ_{i=1}^n (f_i)^2
```

Lemma 5. *In* AMS-2*, if random variables* $\{r_i\}_{i\in[n]}$ *are pairwise independent, then* $\mathbb{E}[Z^2] = \sum_{i=1}^{n} f_i^2 = F_2$*. That is,* AMS-2 *is an unbiased estimator for the* 2^{nd} *moment.*

Proof.

$$
\begin{array}{rcll}
\mathbb{E}[Z^2] & = & \mathbb{E}[(\sum_{i=1}^{n} r_i f_i)^2] & \text{Since } Z = \sum_{i=1}^{n} r_i f_i \text{ at the end} \\
& = & \mathbb{E}[\sum_{i=1}^{n} r_i^2 f_i^2 + 2\sum_{1\leq i<j\leq n} r_i r_j f_i f_j] & \text{Expanding } (\sum_{i=1}^{n} r_i f_i)^2 \\
& = & \sum_{i=1}^{n} \mathbb{E}[r_i^2 f_i^2] + 2\sum_{1\leq i<j\leq n} \mathbb{E}[r_i r_j f_i f_j] & \text{Linearity of expectation} \\
& = & \sum_{i=1}^{n} \mathbb{E}[r_i^2] f_i^2 + 2\sum_{1\leq i<j\leq n} \mathbb{E}[r_i r_j] f_i f_j & f_i\text{'s are (unknown) constants} \\
& = & \sum_{i=1}^{n} f_i^2 + 2\sum_{1\leq i<j\leq n} \mathbb{E}[r_i r_j] f_i f_j & \text{Since } (r_i)^2 = 1, \forall i \in [n] \\
& = & \sum_{i=1}^{n} f_i^2 + 2\sum_{1\leq i<j\leq n} \mathbb{E}[r_i]\mathbb{E}[r_j] f_i f_j & \text{Since } \{r_i\}_{i\in[n]} \text{ are pairwise independent} \\
& = & \sum_{i=1}^{n} f_i^2 + 2\sum_{1\leq i<j\leq n} 0 \cdot f_i f_j & \text{Since } \mathbb{E}[r_i] = 0, \forall i \in [n] \\
& = & \sum_{i=1}^{n} f_i^2 & \text{Simplifying} \\
& = & F_2 & \text{Since } F_2 = \sum_{i=1}^{n} (f_i)^2
\end{array}
$$

□

Lemma 6. *In* AMS-2*, if random variables* $\{r_i\}_{i\in[n]}$ *are 4-wise independent, then* $Var[Z^2] \leq 2(\mathbb{E}[Z^2])^2$.

Proof. As before, $\mathbb{E}[r_i] = 0$ and $\mathbb{E}[r_i^2] = 1$ for all $i \in [n]$. By 4-wise independence, the expectation of any product of ≤ 4 different r_i's is the product of their expectation, which is zero. For instance, $\mathbb{E}[r_i r_j r_k r_l] = \mathbb{E}[r_i]\mathbb{E}[r_j]\mathbb{E}[r_k]\mathbb{E}[r_l] = 0$. Note that $r_i^2 = r_i^4 = 1$ and $r_i = r_i^3$.

$$\begin{aligned}
\mathbb{E}[Z^4] &= \mathbb{E}[(\textstyle\sum_{i=1}^n r_i f_i)^4] && \text{Since } Z = \textstyle\sum_{i=1}^n r_i f_i \text{ at the end} \\
&= \textstyle\sum_{i=1}^n \mathbb{E}[r_i^4] f_i^4 + 6\sum_{1\leq i<j\leq n} \mathbb{E}[r_i^2 r_j^2] f_i^2 f_j^2 && \text{Linearity of expectation and 4-wise independence} \\
&= \textstyle\sum_{i=1}^n f_i^4 + 6\sum_{1\leq i<j\leq n} f_i^2 f_j^2 && \text{Since } \mathbb{E}[r_i^4] = \mathbb{E}[r_i^2] = 1, \forall i \in [n]
\end{aligned}$$

The coefficient of $\sum_{1\leq i<j\leq n} \mathbb{E}[r_i^2 r_j^2] f_i^2 f_j^2$ is $\binom{4}{2}\binom{2}{2} = 6$. All other terms besides $\sum_{i=1}^n \mathbb{E}[r_i^4] f_i^4$ and $6\sum_{1\leq i<j\leq n} \mathbb{E}[r_i^2 r_j^2] f_i^2 f_j^2$ evaluate to 0 because of 4-wise independence.

$$\begin{aligned}
\text{Var}[Z^2] &= \mathbb{E}[(Z^2)^2] - (\mathbb{E}[Z^2])^2 && \text{Definition of variance} \\
&= \textstyle\sum_{i=1}^n f_i^4 + 6\sum_{1\leq i<j\leq n} f_i^2 f_j^2 - (\mathbb{E}[Z^2])^2 && \text{From above} \\
&= \textstyle\sum_{i=1}^n f_i^4 + 6\sum_{1\leq i<j\leq n} f_i^2 f_j^2 - (\sum_{i=1}^n f_i^2)^2 && \text{By Lemma 5 since 4-wise ind.} \Rightarrow \text{pairwise ind.} \\
&= 4\textstyle\sum_{1\leq i<j\leq n} f_i^2 f_j^2 && \text{Expand and simplify} \\
&\leq 2(\textstyle\sum_{i=1}^n f_i^2)^2 && \text{Upper bound} \\
&= 2(\mathbb{E}[Z^2])^2 && \text{By Lemma 5}
\end{aligned}$$

□

Theorem 7. *In* AMS-2, *if* $\{r_i\}_{i\in[n]}$ *are 4-wise independent,* $\Pr[|Z^2 - F_2| > \epsilon F_2] \leq \frac{2}{\epsilon^2}$ *for any* $\epsilon > 0$.

Proof.

$$\begin{aligned}
\Pr[|Z^2 - F_2| > \epsilon F_2] &= \Pr[|Z^2 - \mathbb{E}[Z^2]| > \epsilon \mathbb{E}[Z^2]] && \text{By Lemma 5} \\
&\leq \tfrac{\text{Var}(Z^2)}{(\epsilon \mathbb{E}[Z^2])^2} && \text{By Chebyshev's inequality} \\
&\leq \tfrac{2(\mathbb{E}[Z^2])^2}{(\epsilon \mathbb{E}[Z^2])^2} && \text{By Lemma 6} \\
&= \tfrac{2}{\epsilon^2}
\end{aligned}$$

□

Claim 8. $\mathcal{O}(k \log n)$ *bits of randomness suffices to obtain a set of* k*-wise independent random variables.*

Proof. Recall the definition of hash family $\mathcal{H}_{n,m}$. In a similar fashion[1], we consider hashes from the family (for prime p):

$$\begin{aligned}
\{h_{a_{k-1},a_{k-2},\dots,a_1,a_0} : h(x) &= \textstyle\sum_{i=1}^{k-1} a_i x^i \mod p \\
&= a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \cdots + a_1 x + a_0 \mod p, \\
&\quad \forall a_{k-1}, a_{k-2}, \dots, a_1, a_0 \in \mathbb{Z}_p\}
\end{aligned}$$

This requires k random coefficients, which can be stored with $\mathcal{O}(k \log n)$ bits. □

Observe that the above analysis only require $\{r_i\}_{i\in[n]}$ to be 4-wise independent. Claim 8 implies that AMS-2 only needs $\mathcal{O}(4 \log n)$ bits to represent $\{r_i\}_{i\in[n]}$.

Although the failure probability $\frac{2}{\epsilon^2}$ is large for small ϵ, one can repeat t times and output the mean (Recall Trick 1). With $t \in \mathcal{O}(\frac{1}{\epsilon^2})$ samples, the failure probability drops to $\frac{2}{t\epsilon^2} \in \mathcal{O}(1)$. When the failure probability is < 0.5, one can then call the routine k times independently, and return the median (Recall Trick 2). On the whole, for any given $\epsilon > 0$ and $\delta > 0$, $\mathcal{O}(\frac{\log(n)\log(1/\delta)}{\epsilon^2})$ space suffices to yield a $(1 \pm \epsilon)$-approximation algorithm that succeeds with probability $> 1 - \delta$.

2.2 General k

The assumption of known m in AMS-k can be removed via reservoir sampling[2]. The idea is as follows: Initially, initialize stream length and J as both 0. When a_i arrives, choose to replace J with i with probability $\frac{1}{i}$. If J is replaced, reset r to 0 and start counting from this stream suffix onwards. It can be shown that the choice of J is uniform over current stream length.

Lemma 9. *In* AMS-k, $\mathbb{E}[Z] = \sum_{i=1}^n f_i^k = F_k$. *That is,* AMS-k *is an unbiased estimator for the* k^{th} *moment.*

[1] See https://en.wikipedia.org/wiki/K-independent_hashing
[2] See https://en.wikipedia.org/wiki/Reservoir_sampling

Algorithm 4 AMS-K($S = \{a_1, \ldots, a_m\}$)

```
m ← |S|                                   ▷ For now, assume we know m = |S|
J ∈_{u.a.r.} [m]                          ▷ Pick a random index
r ← 0
for a_i ∈ S do                            ▷ Items arrive in streaming fashion
    if i ≥ J and a_i = a_J then
        r ← r + 1     ▷ At the end, r = |{i ∈ [m] : i ≥ J and a_i = a_J}| = # a_J in suffix of stream
    end if
end for
Z ← m(r^k − (r − 1)^k)
return Z                                  ▷ Estimate of F_k = Σ_{i=1}^n (f_i)^k
```

Proof. When $J = i$, there are f_i choices for J. By telescoping sums, we have:

$$\begin{aligned}
\mathbb{E}[Z \mid J = i] &= \tfrac{1}{f_i}[m(f_i^k - (f_i-1)^k)] + \tfrac{1}{f_i}[m((f_i-1)^k - (f_i-2)^k)] + \cdots + \tfrac{1}{f_i}[m(1^k + 0^k)] \\
&= \tfrac{m}{f_i}[(f_i^k - (f_i-1)^k) + ((f_i-1)^k - (f_i-2)^k) + \cdots + (1^k + 0^k)] \\
&= \tfrac{m}{f_i} f_i^k
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Z] &= \textstyle\sum_{i=1}^n \mathbb{E}[Z \mid J = i] \cdot \Pr[J = i] && \text{Condition on the choice of } J \\
&= \textstyle\sum_{i=1}^n \mathbb{E}[Z \mid J = i] \cdot \frac{f_i}{m} && \text{Since choice of } J \text{ is uniform at random} \\
&= \textstyle\sum_{i=1}^n \frac{m}{f_i} f_i^k \cdot \frac{f_i}{m} && \text{From above} \\
&= \textstyle\sum_{i=1}^n f_i^k && \text{Simplifying} \\
&= F_k && \text{Since } F_k = \textstyle\sum_{i=1}^n f_i^k
\end{aligned}$$

□

Lemma 10. *For every n positive reals $f_1, f_2, \ldots, f_n$,*

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

Proof. Let $M = \max_{i\in[n]} f_i$, then $f_i \leq M$ for any $i \in [n]$ and $M^k \leq \sum_{i=1}^n f_i^k$. Hence,

$$\begin{aligned}
(\textstyle\sum_{i=1}^n f_i)(\sum_{i=1}^n f_i^{2k-1}) &\leq (\textstyle\sum_{i=1}^n f_i)(M^{k-1}\sum_{i=1}^n f_i^k) && \text{Pulling out a } M^{k-1} \text{ factor} \\
&\leq (\textstyle\sum_{i=1}^n f_i)(\sum_{i=1}^n f_i^k)^{(k-1)/k}(\sum_{i=1}^n f_i^k) && \text{Since } M^k \leq \textstyle\sum_{i=1}^n f_i^k \\
&= (\textstyle\sum_{i=1}^n f_i)(\sum_{i=1}^n f_i^k)^{(2k-1)/k} && \text{Merging the last two terms} \\
&\leq n^{1-1/k}(\textstyle\sum_{i=1}^n f_i^k)^{1/k}(\sum_{i=1}^n f_i^k)^{(2k-1)/k} && \text{Fact: } (\textstyle\sum_{i=1}^n f_i)/n \leq (\sum_{i=1}^n f_i^k/n)^{1/k} \\
&= n^{1-1/k}(\textstyle\sum_{i=1}^n f_i)^2 && \text{Merging the last two terms}
\end{aligned}$$

□

Remark $f_1 = n^{1/k}, f_2 = \cdots = f_n = 1$ is a tight example for Lemma 10, up to a constant factor.

Theorem 11. *In* AMS-K, *$Var(Z) \leq kn^{1-\frac{1}{k}}(\mathbb{E}[Z])^2$*

Proof. Let us first analyze $\mathbb{E}[Z^2]$.

$$\begin{aligned}
\mathbb{E}[Z^2] &= \tfrac{m}{m}[(1^k - 0^k)^2 + (2^k - 1^k)^2 + \cdots + (f_1^k - (f_1-1)^k)^2 && \text{(A)} \\
&\quad + (1^k - 0^k)^2 + (2^k - 1^k)^2 + \cdots + (f_2^k - (f_2-1)^k)^2 \\
&\quad + \ \ldots \\
&\quad + (1^k - 0^k)^2 + (2^k - 1^k)^2 + \cdots + (f_n^k - (f_n-1)^k)^2] \\
&\leq m[k \cdot 1^{k-1}(1^k - 0^k) + k \cdot 2^{k-1} \cdot (2^k - 1^k) + \cdots + k \cdot f_1^{k-1} \cdot (f_1^k - (f_1-1)^k) && \text{(B)} \\
&\quad + k \cdot 1^{k-1}(1^k - 0^k) + k \cdot 2^{k-1} \cdot (2^k - 1^k) + \cdots + k \cdot f_2^{k-1} \cdot (f_2^k - (f_2-1)^k) \\
&\quad + \ \ldots \\
&\quad + k \cdot 1^{k-1}(1^k - 0^k) + k \cdot 2^{k-1} \cdot (2^k - 1^k) + \cdots + k \cdot f_n^{k-1} \cdot (f_n^k - (f_n-1)^k)] \\
&\leq m[k \cdot f_1^{2k-1} + k \cdot f_2^{2k-1} + \cdots + k \cdot f_n^{2k-1}] && \text{(C)} \\
&= k \cdot m \cdot F_{2k-1} && \text{(D)} \\
&= k \cdot F_1 \cdot F_{2k-1} && \text{(E)}
\end{aligned}$$

(A) By definition of $\mathbb{E}[Z^2]$ (condition on J and expand in the same style as the proof of Theorem 9).

(B) $\forall 0 < b < a$, $a^k - b^k = (a-b)(a^{k-1} + a^{k-2}b + \cdots + ab^{k-2} + b^{k-1}) \leq (a-b)ka^{k-1}$, with $a = b+1$

(C) Telescope each row, then ignore remaining negative terms

(D) $F_{2k-1} = \sum_{i=1}^{n} f_i^{2k-1}$

(E) $F_1 = \sum_{i=1}^{n} f_i = m$

Then,

$$\begin{aligned}
\text{Var}(Z) &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 && \text{Definition of variance} \\
&\leq \mathbb{E}[Z^2] && \text{Ignore negative part} \\
&\leq k \cdot F_1 \cdot F_{2k-1} && \text{From above} \\
&\leq kn^{1-1/k} F_k^2 && \text{By Lemma 10} \\
&= kn^{1-1/k} (\mathbb{E}[Z])^2 && \text{By Theorem 9}
\end{aligned}$$

□

Remark Proofs for Lemma 10 and Theorem 11 were omitted in class. The above proofs are presented in a style consistent with the rest of the scribe notes. Interested readers can refer to [AMS96] for details.

Remark One can apply an analysis similar to the case when $k = 2$, then use Tricks 1 and 2.

Claim 12. *For $k > 2$, a lower bound of $\widetilde{\Theta}(n^{1-\frac{2}{k}})$ is known.*

Proof. Theorem 3.1 in [BYJKS04] gives the lower bound. See [IW05] for algorithm that achieves it. □

References

[AMS96] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 20–29. ACM, 1996.

[BYJK+02] Ziv Bar-Yossef, TS Jayram, Ravi Kumar, D Sivakumar, and Luca Trevisan. Counting distinct elements in a data stream. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 1–10. Springer, 2002.

[BYJKS04] Ziv Bar-Yossef, Thathachar S Jayram, Ravi Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[FM85] Philippe Flajolet and G Nigel Martin. Probabilistic counting algorithms for data base applications. *Journal of computer and system sciences*, 31(2):182–209, 1985.

[IW05] Piotr Indyk and David Woodruff. Optimal approximations of the frequency moments of data streams. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 202–208. ACM, 2005.