Estimating hybrid frequency moments of data streams

Sumit Ganguly, Mohit Bansal⋆, and Shruti Dube ⋆⋆

Indian Institute of Technology, Kanpur

Abstract. We consider the problem of estimating hybrid frequency moments of two dimensional data streams. In this model, data is viewed to be organized in a matrix form $(A_{i,j})_{1\le i,j,\le n}$. The entries $A_{i,j}$ are updated coordinate-wise, in arbitrary order and possibly multiple times. The updates include both increments and decrements to the current value of $A_{i,j}$. The hybrid frequency moment $F_{p,q}(A)$ is defined as $\sum_{j=1}^n \left(\sum_{i=1}^n |A_{i,j}|^p\right)^q$ and is a generalization of the frequency moment of one-dimensional data streams.

We present an $\tilde{O}(1)$ space[1] algorithm for the problem of estimating $F_{p,q}$ for $p \in [0,2]$ and $q \in [0,1]$. We also present a $\tilde{O}(n^{1-1/q})$ space algorithm for estimating $F_{p,q}$ for $p \in [0,2]$ and $q \in (1,2]$.

1 Introduction

The data stream model of computation is an abstraction for a variety of practical applications arising in network monitoring, sensor networks, RF-id processing, database systems, online web-mining, etc.. A problem of basic utility and relevance in this setting is the following *hybrid frequency moments estimation* problem. Consider a networking application where a stream of packets with schema (*src-addr*, *dest-addr*, *nbytes*, *time*) arrives at a router. The problem is to warn against the following scenario arising out of a distributed denial of service attack, where, a few destination addresses receive messages from an unusually large number of distinct source addresses. This can be quantified as follows: let A be an $n \times n$ matrix where $A_{i,j}$ is the count of the number of messages from node i to node j. Then $A^0_{i,j}$ is 1 if i sends a message to j and is 0 otherwise. Thus, $\sum_{i=1}^n A^0_{i,j}$ counts the number of distinct sources that send at least one message to j. Define the hybrid moment $F_{0,2}(A) = \sum_{j=1}^n(\sum_{i=1}^n A^0_{i,j})^2$. In an attack scenario, $F_{0,2}(A)$ becomes large compared to its average value. Since n can be very large (e.g., in the millions), it is not feasible to store and update the traffic matrix A at network line speeds. We propose instead to use the data streaming approach to this problem, namely, to design a sub-linear space data structure that, (a) processes updates to the entries of A, and, (b) provides a randomized algorithm for approximating the value of $F_{0,2}(A)$.

Quantities such as $F_{0,2}(A)$ are known as the hybrid moment of a matrix A. They are more generally defined [19] as follows. Given an $n \times n$ integer matrix A with columns $A_1, A_2, \ldots, A_n$, the hybrid frequency moment $F_{p,q}(A)$ is the qth moment of the n-dimensional vector $[F_p(A_1), F_p(A_2), \ldots, F_p(A_n)]$. That is,

$$F_{p,q}(A) = \sum_{j=1}^n \left(\sum_{i=1}^n A^p_{i,j}\right)^q = \sum_{j=1}^n (F_p(A_j))^q \ .$$

⋆ Current Affiliation: University of California, Berkeley.

⋆⋆ Current Affiliation: McKinsey & Company, New Delhi.

[1] The $\tilde{O}$ notation suppresses factors of the form $(\log^{O(1)} n) \cdot (\log^{O(1)} F_{1,1}) \cdot \epsilon^{-\Omega(1)}$.

Data Stream Model. We will be interested in algorithms in the data stream model, that is, the input is abstracted as a potentially infinite sequence σ of records of the form (pos, i, j, Δ), where, $i, j \in \{1, 2, \ldots, n\}$ and $\Delta \in \mathbb{Z}$ is the change to the value of $A_{i,j}$. The pos attribute is simply the sequence number of the record. Each input record (pos, i, j, Δ) changes $A_{i,j}$ to $A_{i,j} + \Delta$. In other words, the $A_{i,j}$ is the sum of the changes made to the (i, j)th entry since the inception of the stream:

$$A_{i,j} = \sum_{(pos,i,j,\Delta)\in\sigma} \Delta, \quad 1 \le i, j \le n \ .$$

In this paper, we consider the problems of estimating $F_{p,q}$ and allow general matrix streams, that is, matrix entries may be positive, zero or negative.

Prior work. Hybrid frequency moments $F_{p,q}(A)$ are a generalization of the frequency moment $F_p(a)$ of an n-dimensional vector a, defined as $F_p(a) = \sum_{j=1}^{n} |a_i|^p$. The problem of estimating $F_p(a)$ has been studied in the data stream model where the input is a stream of updates to the components of a. This problem has been influential in the development of algorithms for data streams. We will say that a randomized algorithm computes an ϵ-approximation to a real valued quantity L, provided, it returns $\hat{L}$ such that $|\hat{L} - L| < \epsilon L$, with probability $\ge \frac{3}{4}$.

Alon, Matias and Szegedy [1] present a seminal randomized sketch technique for ϵ-approximation of $F_2(a)$ in the data streaming model using space $O(\epsilon^{-2} \log F_1(a))$ bits. Using the techniques of [1], it is easily shown that deterministically estimating $F_p(a)$ for any real $p \ge 0$ requires $\Omega(n)$ space [11]. Hence, work in the area of sub-linear space estimation of moments has considered only randomized algorithms. Estimation of $F_0(a)$ was first considered by Flajolet and Martin in [9]; the work in [1] presents a modern version of this technique for estimating $F_0(a)$ to within a constant multiplicative factor and using space $O(\log n)$. Gibbons and Tirthapura [13] present an ϵ-approximation algorithm using space $O(\epsilon^{-2} \log F_1(a))$; this is further improved in [3]. The use of p-stable sketches was proposed by Indyk [14] for estimating $F_p(a)$, for $0 < p \le 2$, using space $\tilde{O}(1)$. Indyk and Woodruff [15] present a near optimal space algorithm for estimating F_p, for $p > 2$. Woodruff [21] presents an $\Omega(\epsilon^{-2})$ space lower bound for the problem of estimating F_p, for all $p \ge 0$, implying that the stable sketches technique is space optimal up to poly-logarithmic factors. A space lower bound of $\Omega(n^{1-2/p})$ was shown for the problem of estimating F_p for $p \ge 2$ in a series of developments [1, 2, 5]. Cormode and Muthukrishnan [8] present an algorithm for obtaining an ϵ-approximation for $F_{0,2}(A)$ using space $\tilde{O}(\sqrt{n})$. This is the only prior work on estimating hybrid moments of a matrix in the data stream model.

Contributions. We present randomized algorithms for the problem of estimating hybrid moments $F_{p,q}(A)$ of a matrix A in the data stream model. We consider the range $p \in [0, 2]$ and $q \in [0, 2]$. We present a novel variation of the stable sketches technique to obtain a $\tilde{O}(1)$ space algorithm for estimating $F_{p,q}$ in the range $p \in [0, 2]$ and $q \in [0, 1]$. For $p \in [0, 2]$ and $q \in (1, 2]$, we present an algorithm for estimating $F_{p,q}$ that uses $\tilde{O}(n^{1-1/q}/\epsilon^3)$ space.

2 Review: Hss algorithm

In this section, we review the *Hierarchical Sampling over Sketches* (HSS) proposed in [4] for estimating a class of metrics over data-streams of the following form

$$\Psi(\mathcal{S}) = \sum_{i: f_i \neq 0} \psi(|f_i|) \ . \tag{1}$$

Sampling sub-streams. The HSS algorithm uses a sampling scheme as follows. From the input stream $\mathcal{S}$, sub-streams $\mathcal{S}_0, \ldots, \mathcal{S}_L$ are created such that $\mathcal{S}_0 = \mathcal{S}$ and for $1 \leq l \leq L$, $\mathcal{S}_l$ is obtained from $\mathcal{S}_{l-1}$ by sub-sampling each distinct item appearing in $\mathcal{S}_{l-1}$ independently with probability $\frac{1}{2}$. At level 0, $\mathcal{S}_0 = \mathcal{S}$. $\mathcal{S}_l$ is a randomly sampled sub-stream of $\mathcal{S}_{l-1}$ with probability 1/2, for $l \geq 1$, based on the identity of the items. The sub-sampling scheme is implemented as follows. We assume that n is a power of 2. Let $h : [n] \rightarrow [0, \max(n^2, W)]$ be a random hash function drawn from a pair-wise independent hash family and $W \geq 2F_1$. Let $L_{\max} = \lceil \log(\max(n^2, W)) \rceil$. Define the random function level : $[n] \rightarrow [1, L_{\max}]$ as follows.

$$\text{level}(i) = \begin{cases} 1 & \text{if } h(i) = 0 \\ \text{lsb}(h(i)) & 2 \leq \text{level}(i) \leq L_{\max} \end{cases} .$$

where, $lsb(x)$ is the position of the least significant "1" in the binary representation of x. The probability distribution of the random level function is as follows.

$$\Pr\{\text{level}(i) = l\} = \begin{cases} \frac{1}{2} + \frac{1}{n} & \text{if } l = 1 \\ \frac{1}{2^l} & \text{otherwise.} \end{cases}$$

At each level $l \in \{0, 1, \ldots, L_{\max}\}$, the HSS algorithm keeps a frequency estimation data-structure denoted by DS_l, that takes as input the sub-stream $\mathcal{S}_l$, and returns an approximation to the frequencies of items that map to $\mathcal{S}_l$. The DS_l structure can be any standard data structure such as the COUNT-MIN sketch sketch structure [7] or the COUNTSKETCH structure [6], or any other data structure. Each stream update (pos, i, v) belonging to $\mathcal{S}_l$ is propagated to the frequent items data structures DS_l for $0 \leq l \leq \text{level}(i)$. Let $k(l)$ denote a space parameter for the data structure DS_l, for example, $k(l)$ is the size of the hash tables in the COUNT-MIN sketch or COUNTSKETCH structures. The values of $k(l)$ are the same for levels $l = 1, 2, \ldots, L$ and is four times the value for $k(0)$, that is, $k(1) = \ldots = k(L) = 4k(0)$. This non-uniformity is a technicality required by Lemma 1. We refer to $k = k(0)$ as the space parameter of the HSS structure.

Approximating f_i. Let $\Delta_l(k)$ denote the additive error of the frequency estimation by the data structure DS_l) at level l and using space parameter k. That is, we will assume that

$$|\hat{f}_{i,l} - f_i| \leq \Delta_l(k) \text{ with probability } 1 - 2^{-t}$$

where, t is a parameter and $\hat{f}_{i,l}$ is the estimate for the frequency of f_i obtained using the frequent items structure $DS_l(k)$.

Given a data stream, rank(r) is an item with the r^{th} largest absolute value of the frequency, where, ties are broken arbitrarily. We say that an item i has rank r if rank$(r) = i$. For a given value of k, $1 \le k \le n$, the set $Top(k)$ is the set of items with rank $\le k$. The residual second moment [6] of a data stream, denoted by $F_2^{res}(k)$, is defined as the second moment of the frequency of the data stream after the top-k frequencies have been removed, that is, $F_2^{res}(k) = \sum_{r>k} f^2_{\mathrm{rank}(r)}$. The residual first moment [7] of a data stream, denoted by F_1^{res}, is analogously defined as the first frequency moment of the data stream after the top-k frequencies have been removed, that is, $F_1^{res} = \sum_{r>k} |f_{rank(r)}|$.

Let $F_1^{res}(k,l)$ and $F_2^{res}(k,l)$ respectively denote $F_1^{res}(k)$ and $F_2^{res}(k)$ of the sub-stream $\mathcal{S}_l$. Lemma 1 relates the random values $F_1^{res}(k,l)$ and $F_2^{res}(k,l)$ to their corresponding non-random values $F_1^{res}(k)$ and $F_2^{res}(k)$, respectively.

Convention. For the sake of simplicity in notation, in this section, we will use f_i to denote $|f_i|$.

Lemma 1. *[10]*

1. *For $l \ge 1$ and $k \ge 2$,* $\Pr\left\{F_1^{res}(k,l) \le \frac{F_1^{res}(2^{l-2}k)}{2^{l-1}}\right\} \ge 1 - 2e^{-k/6}$.
2. *For $l \ge 1$,* $\Pr\left\{F_2^{res}(k,l) \le \frac{F_2^{res}(2^{l-2}k)}{2^{l-1}}\right\} \ge 1 - 2e^{-k/6}$.

Group definitions. At each level l, the sampled stream $\mathcal{S}_l$ is provided as input to a data structure DS_l, that when queried, returns an estimate $\hat{f}_{i,l}$ for any $i \in [n]$ satisfying

$$|\hat{f}_{i,l} - f_i| \le \Delta_l, \quad \text{with prob. } 1 - 2^{-t} \ .$$

Here, t is a parameter that will be fixed in the analysis and the additive error Δ_l is a function of the algorithm used by DS_l. Fix a parameter $\bar{\epsilon}$ which will be closely related to the given accuracy parameter ϵ, and is chosen depending on the problem. For example, in order to estimate F_p, $\bar{\epsilon}$ is set to $\frac{\epsilon}{4p}$. Therefore,

$$\hat{f}_{i,l} \in (1 \pm \bar{\epsilon}) f_i, \quad \text{provided, } f_i > \frac{\Delta_l}{\bar{\epsilon}}, \quad \text{and } i \in \mathcal{S}_l, \text{ with prob. } 1 - 2^{-t} \ .$$

Define the following event

$$\textsc{GoodEst} \equiv |\hat{f}_{i,l} - f_i| < \Delta_l, \text{ for each } i \in \mathcal{S}_l \text{ and } l \in \{0, 1, \ldots, L\} \ .$$

By union bound,

$$\Pr\{\textsc{GoodEst}\} \ge 1 - n(L+1)2^{-t} \ . \tag{2}$$

The analysis is conditioned on the event GoodEst.
Define a sequence of geometrically decreasing thresholds $T_0, T_1, \ldots, T_L$ as follows.

$$T_l = \frac{T_0}{2^l}, \quad l = 1, 2, \ldots, L \text{ and } \frac{1}{2} < T_L \le 1 \ . \tag{3}$$

Consequently, $L = \lceil \log T_0 \rceil$. Note that L and $L_{\max}$ are distinct parameters. The threshold values T_l's are used to partition the elements of the stream into groups $G_0, \ldots, G_L$ as follows.

$$G_0 = \{i \in \mathcal{S} : \ |f_i| \ge T_0\} \quad \text{and} \quad G_l = \{i \in \mathcal{S} : \ T_l < |f_i| \le T_{l-1}\}, \quad l = 1, 2, \ldots, L \ .$$

An item i is said to be *discovered as frequent* at level l, provided, i maps to $\mathcal{S}_l$ and $\hat{f}_{i,l} \geq Q_l$, where, Q_l, $l = 0, 1, 2 \ldots, L$, is a parameter family. The values of Q_l are chosen as follows.

$$Q_l = T_l(1 - \bar{\epsilon}) \tag{4}$$

The space parameter $k(l)$ is chosen at level l as follows.

$$\Delta_0 = \Delta_0(k) \leq \bar{\epsilon} Q_0, \qquad \Delta_l = \Delta_l(4k) \leq \bar{\epsilon} Q_l, l = 1, 2, \ldots, L \ . \tag{5}$$

The value of T_0 is a critical parameter for the Hss parameter and its precise choice depends on the problem that is being solved. For example, for estimating F_p, T_0 is chosen as $\frac{1}{\bar{\epsilon}(1-\bar{\epsilon})}\left(\frac{\hat{F}_2}{k}\right)^{1/2}$.

Hierarchical samples. Items are sampled and placed into sampled groups $\bar{G}_0, \bar{G}_1, \ldots, \bar{G}_L$ as follows. The estimated frequency of an item i is defined as

$$\hat{f}_i = \hat{f}_{i,r}, \text{ where, } r \text{ is the lowest level such that } \hat{f}_{i,r} > Q_r \ .$$

The sampled groups are defined as follows.

$$\bar{G}_0 = \{i : |\hat{f}_i| \geq T_0\} \text{ and } \bar{G}_l = \{i : T_l < |\hat{f}_i| \leq T_{l-1} \text{ and } i \in \mathcal{S}_l\}, 1 \leq l \leq L \ .$$

The choices of the parameter settings satisfy the following properties. We use the following standard notation. For $a, b \in \mathbb{R}$ and $a < b$, (a, b) denotes the open interval defined by the set of points between a and b (end points not included), $[a, b]$ represents the closed interval of points between a and b (both included) and finally $[a, b)$ and $(a, b]$ respectively, represent the two half-open intervals. Partition a frequency group G_l, for $1 \leq l \leq L - 1$, into three adjacent sub-regions:

$$\begin{aligned}
\textit{lmargin}(G_l) &= [T_l, T_l + \bar{\epsilon} Q_l], \quad l = 0, 1, \ldots, L-1 \text{ and is undefined for } l = L. \\
\textit{rmargin}(G_l) &= [Q_{l-1} - \bar{\epsilon} Q_{l-1}, T_{l-1}), \quad l = 1, 2, \ldots, L \text{ and is undefined for } l = 0. \\
\mathrm{mid}(G_l) &= (T_l + \bar{\epsilon} Q_l, Q_{l-1} - \bar{\epsilon} Q_l), \quad 1 \leq l \leq L - 1
\end{aligned}$$

These regions respectively denote the *lmargin* (left-margin), *rmargin* (right-margin) and *middle-region* of the group G_l. An item i is said to belong to one of these regions if its true frequency lies in that region. The middle-region of groups G_0 and G_L are each extended to include the right and left margins, respectively. That is,

$$\begin{aligned}
\textit{lmargin}(G_0) &= [T_0, T_0 + \bar{\epsilon} Q_0) \text{ and } \mathrm{mid}(G_0) = [T_0 + \bar{\epsilon} Q_0, F_1] \\
\textit{rmargin}(G_L) &= [Q_{L-1} - \bar{\epsilon} Q_{L-1}, T_{L-1}) \text{ and } \mathrm{mid}(G_0) = [0, Q_{L-1} - \bar{\epsilon} Q_{L-1}) \ .
\end{aligned}$$

Estimator. The sample is used to compute the estimate $\hat{\Psi}$. We also define an idealized estimator $\bar{\Psi}$ that assumes that the frequent items structure is an oracle that does not make errors.

$$\hat{\Psi} = \sum_{l=0}^{L} \sum_{i \in \bar{G}_l} \psi(\hat{f}_i) \cdot 2^l \qquad \bar{\Psi} = \sum_{l=0}^{L} \sum_{i \in \bar{G}_l} \psi(f_i) \cdot 2^l \tag{6}$$

Lemma 2 shows that the expected value of $\bar{\Psi}$ is Ψ, assuming the event GoodEst holds.

Lemma 2. *[10]* $\mathsf{E}\big[\bar{\Psi} \mid \textsc{GoodEst}\big] = \Psi$.

Notation. Let $l(i)$ denote the index of the group G_l such that $i \in G_l$.

Lemma 3. *[10]*
$$\mathsf{Var}\big[\bar{\Psi} \mid \textsc{GoodEst}\big] \leq \sum_{\substack{i\in[n]\\ i\notin(G_0 - lmargin(G_0))}} \psi^2(f_i)\cdot 2^{l(i)+1} \enspace .$$

The error incurred by the estimate $\hat{\Psi}$ is $|\hat{\Psi} - \Psi|$, and can be bounded as the sum of two error components.
$$|\hat{\Psi} - \Psi| \leq |\bar{\Psi} - \Psi| + |\hat{\Psi} - \bar{\Psi}| = \mathcal{E}_1 + \mathcal{E}_2$$
Here, $\mathcal{E}_1 = |\Psi - \bar{\Psi}|$ is the error due to sampling and $\mathcal{E}_2 = |\hat{\Psi} - \bar{\Psi}|$ is the error due to the estimation of the frequencies. By Chebychev's inequality
$$\mathsf{Pr}\Big\{\mathcal{E}_1 \leq 3(\mathsf{Var}\big[\bar{\Psi}\big])^{1/2} \mid \textsc{GoodEst}\Big\} \geq \frac{8}{9} \enspace .$$
Notation. Define a real valued function $\pi : [n] \to \mathbb{R}$ as follows.
$$\pi_i = \begin{cases} \Delta_{l(i)}\cdot|\psi'(\xi_i(f_i,\Delta_l))| & \text{if } i\in G_0 - \mathit{lmargin}(G_0) \text{ or } i\in \mathrm{mid}(G_l)\\ \Delta_{l(i)}\cdot|\psi'(\xi_i(f_i,\Delta_l))| & \text{if } i\in \mathit{lmargin}(G_l),\ \text{for some } l>1\\ \Delta_{l(i)-1}\cdot|\psi'(\xi_i(f_i,\Delta_{l-1}))| & \text{if } i\in \mathit{rmargin}(G_l)\end{cases}$$
where, the notation $\xi_i(f_i, \Delta_l)$ returns the value of t that maximizes $|\psi'(t)|$ in the interval $[f_i - \Delta_l, f_i + \Delta_l]$.
$$\Pi_1 = \sum_{i\in[n]} \pi_i, \tag{7}$$
$$\Pi_2 = 3\Bigg(\sum_{i\in[n] i\notin G_0 - lmargin(G_0)} \pi_i^2\cdot 2^{l(i)+1}\Bigg)^{1/2} \tag{8}$$
$$\Lambda = 3\Bigg(\sum_{l=1}^{L}\psi(T_{l-1})\psi(G_l)2^{l+1} + \psi(T_0+\Delta_0)\psi(lmargin(G_0))\Bigg)^{1/2} \tag{9}$$
Here, the notation $\psi(G_l)$ denotes $\sum_{i\in G_l}\psi(f_i)$ and likewise $\psi(lmargin(G_0)) = \sum_{i\in lmargin(G_0)}\psi(f_i)$. It can be shown that
$$\Lambda \geq 3(\mathsf{Var}\big[\bar{\Psi}\big])^{1/2} \geq \mathcal{E}_1, \qquad \text{assuming } \textsc{GoodEst} \enspace .$$

Lemma 4. *[10]*
$$\mathsf{E}\big[\mathcal{E}_2 \mid \textsc{GoodEst}\big] \leq \Pi_1,\ \textit{and}\ \mathsf{Var}\big[\mathcal{E}_2 \mid \textsc{GoodEst}\big] \leq \frac{\Pi_2^2}{9} \enspace .$$
Therefore, $\mathsf{Pr}\{\mathcal{E}_2 \leq \Pi_1 + \Pi_2 \mid \textsc{GoodEst}\} \geq \frac{8}{9}$.

Lemma 5 presents the overall expression of error and its probability.

Lemma 5. *[10] Let* $\bar{\epsilon} \leq \frac{1}{3}$. *Then,*
$$\mathsf{Pr}\Big\{|\hat{\Psi} - \Psi| \leq \Lambda + \Pi_1 + \Pi_2\Big\} > \frac{7}{9}(1 - (n(L+1))2^{-t}) \enspace .$$

3 Preliminaries

In this section, we review salient properties of stable distributions and briefly review Indyk's [14] and Li's [18] techniques for estimating moments of one-dimensional vectors in the data streaming model. We use the notation $y \sim D$ to denote that a given random variable y follows a probability distribution D.

Indyk's estimator. The use of p-stable sketches was pioneered by Indyk [14] for estimating F_p, for $0 < p \leq 2$. A stable sketch is a linear combination

$$X = \sum_{i=1}^{n} a_i s_i$$

where $s_i \sim S(p, 1)$, $i \in [n]$ and *i.i.d.*. The first parameter in $S(p, 1)$ is the stability parameter and the second parameter is the scale factor (set to 1). By property of stable distributions,

$$X \sim S\left(p, (F_p(a))^{1/q}\right) \quad .$$

For estimating F_1, Indyk keeps $t = O(\frac{1}{\epsilon^2})$ independent 1-stable (Cauchy) sketches $X_1, X_2, \ldots, X_t$ and defines the estimator

$$\hat{F}_1 = (4/\pi) \cdot \text{median}_{r=1}^{t} |X_r|^q .$$

This estimator is shown to satisfy $\hat{F}_1 \in (1 \pm \epsilon)F_1$ with probability 15/16.

Further, Indyk shows that for stable distributions it suffices to, (a) truncate the support of the distribution $S(p, 1)$ beyond $(nmM)^{O(1)}$, and, (b) consider the approximation to the continuous $S(p, 1)$ distribution by discretizing it by a grid with interval size $(nmM/\epsilon)^{O(1)}$.

Indyk's application of Nisan's PRG. One final difficulty remains, namely, that the sketches $s_i \sim S(p, 1)$ were assumed to be independent. To simulate this would require $\Omega(n)$ random bits. Indyk proposes the following use of Nisan's pseudo-random generator (PRG) [20] for fooling space bounded computations. The total space S used by the randomized machine, not counting the random bits used, is $O(\epsilon^{-2} \log(\epsilon^{-1} nmM))$. First envision that the input stream is reordered so that all updates to a given item i arrive consecutively. Since sketches are linear, the value of the sketches are independent of the order. For each element i, the stable random variables $s_i(u)$ for $u = 1, 2, \ldots, t$ are computed from the ith chunk of S random bits obtained from Nisan's generator that stretches a seed of length $S \log n$ to nS bits, where, $S = O(\epsilon^{-2} \log(nmM\epsilon^{-1}))$. By Nisan's PRG, this fools any space S algorithm. The random seed size becomes $S \log n = O(\epsilon^{-2} \log(nmM\epsilon^{-1}) \log(n))$ and this dominates the space requirement of the F_1 estimation algorithm. The time taken to obtain the ith random bit chunk is $O(\epsilon^{-2} \log(\epsilon^{-1})(\log n))$ simple field operations on a field of size $O(nmM\epsilon^{-1})$. Indyk outlines an argument to extend the analysis of the estimator for F_1 to general F_p for $p \in (0, 2)$, by replacing 1-stable sketches by p-stable sketches. However, the space requirement as a function of p was not explicitly determined, which was subsequently resolved by Li using the geometric means estimator.

Li's estimator. Li [18] proposes several new estimators for the estimation of F_p for $p \in (0, 2)$. These estimators are defined on p-stable sketches $X_u = \sum_{i \in [n]} f_i s_i(u)$, $u = 1, 2, \ldots, t$. The geometric means estimator is defined as

$$\hat{Y}_{p,t} = C(p, p/t)^{-t} \prod_{i=1}^{t} |X_i|^{p/t} .$$

where,

$$C(p, q) = \frac{2}{\pi} \Gamma\big(1 - \frac{q}{\alpha}\big) \Gamma(q) \sin\big(\frac{\pi}{2}(q)\big), -1 < q < p \ .$$

This estimator is unbiased, that is, $\mathsf{E}\big[Y_{p,t}\big] = F_p$. Li [18] proves the following tail-bound[2]:

$$|\hat{Y}_{p,t} - F_p| < \epsilon F_p \text{ with prob. } 1/8 \text{ provided, } t \geq \frac{96(p^2 + 2)}{12\pi^2 \epsilon^2} .$$

For reference, we define the constant

$$K_L(p) = \frac{96(p^2 + 2)}{12\pi^2 \epsilon^2} = O(\epsilon^{-2}) \ . \tag{10}$$

$K_L(p)$ is not principally dependent on p, since, $p \in (0, 2]$.

Li uses Indyk's idea of applying Nisan's PRG to reduce the number of random bits. The space requirement is $O(\epsilon^{-2} \log(\epsilon^{-1} nmM)(\log n))$ and update time requirement remains $O(\epsilon^{-2}(\log \epsilon^{-1}) \log(n))$ operations on $\log(nmM)$ bit numbers. An interesting contribution of Li's work is to show that F_p can be estimated using space $\tilde{O}(\epsilon^{-2})$, independent of the value of p.

Kane, Nelson, Woodruff's (KNW) estimator for F_p. Kane, Nelson and Woodruff [17] present two estimators for estimating F_p for $p \in (0, 2)$ that we denote by KNW-I and KNW-II. Both these estimators use space that is tight with respect to the lower bounds, which was also improved in the same paper [17]. The estimators view the computation of the p-stable sketches as the multiplication of the $t \times n$ random matrix A with the n-dimensional frequency vector f. Each $A_{i,j} \sim \mathcal{D}_p$, where, $\mathcal{D}_p$ is the discretized and truncated version of $\mathrm{St}(p, 1)$. However, unlike Indyk and Li's proposal to use fully independent $A_{i,j}$'s, the KNW-I estimator requires just the following limited independence. (i) For each row value i, the column entries (i.e., $A_{i,j}$'s) are $O(\epsilon^{-p} \log^{3p}(1/\epsilon))$-wise independent, and, (ii) the rows of A are pair-wise independent. This can be achieved using a random seed of size $O(t \log(nmM)) = O\big(\epsilon^{-p} \log^{3p}(1/\epsilon) \log(nmM)\big)$. The update processing time requirement is $O(\epsilon^{-2-p} \log^{3p}(1/\epsilon))$. The KNW-II estimator further reduces the independence requirement among the variates in a single row of A to $\log(\epsilon^{-1}) / \log\log(\epsilon^{-1})$. This reduces the estimation time to $O\big(\epsilon^{-2}(\log \epsilon^{-1})^2 / (\log\log \epsilon^{-1})\big)$ simple operations on fields of size $(nmM)^{O(1)}$.

HSS *estimator.* An estimator for F_p based on the HSS technique was presented in [12] for estimating F_p. Though it uses sub-optimal space $O(\epsilon^{-2-p}(\log(nmM)^2(\log n))$, it has the best update processing time so far, namely, $O(\log^2(nmM))$.

[2] Li proves a left and right tail bound separately; here we combine them into a single inequality

Estimating $F_{p,q}$: Simple cases. Estimation of hybrid moments generalizes the problem of estimating the regular moment $F_p(a)$ for an n-dimensional vector a. In particular, for any p, $F_{p,1}(A) = F_p(a)$ where a is the n^2-dimensional vector obtained by stringing out the matrix A row-wise (or column-wise). Therefore, $F_{p,1}(A)$ can be estimated using standard techniques for estimating F_p of one-dimensional vectors. This implies that for $0 \le p \le 2$, the space requirement for estimating $F_{p,1}$ is $\tilde{O}(\epsilon^{-2})$.

4 Bi-linear stable sketches for estimating $F_{p,q}$, $p \in [0,2]$, $q \in [0,1]$

In this section, we present a technique for estimating $F_{p,q}$ in the range $p \in [0,2]$ and $q \in [0,1]$ using bilinear stable sketches.

Consider two families of fully independent stable variables $\{x_{i,j} : 1 \le i \le j \le n\}$ and $\{\xi_j : 1 \le j \le n\}$, where, $x_{i,j} \sim S(p,1)$ and $\xi_j \sim S(q,1)$. A p,q bi-linear stable sketch is defined as

$$X = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{i,j} x_{i,j} \xi_j^{1/p} \ .$$

Corresponding to each stream update (pos, i, j, Δ), the bi-linear sketch is updated as follows:

$$X := X + \Delta \cdot x_{i,j} \cdot \xi_j^{1/p} \ .$$

A collection of $s_1 s_2$ bi-linear sketches $\{X_{u,v} \mid 1 \le u \le s_1, 1 \le v \le s_2\}$ is kept such that for each distinct value of v, the family of sketches $\{X_{u,v}\}_{u=1,2,\ldots,s_1}$ uses the independent family of stable variables $\{x_{i,j}(u,v)\}$ but uses the same family of stable variables $\{\xi_j(v)\}$. That is,

$$X(u,v) = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{i,j} x_{i,j}(u,v) (\xi_j(v))^{1/p}, \qquad u = 1,\ldots,s_1, v = 1,\ldots,s_2 \ . \tag{11}$$

We note that for $0 < q \le 1$, there exist stable distributions $S(q,1)$ with non-negative support. Thus, $\xi_j \sim S(q,1)$ is non-negative and $\xi_j^{1/p}$ is non-negative. The estimate $\hat{F}_{p,q}$ is obtained using the following steps.

Algorithm BILINSTABLE$(p, q, s_1, s_2, \{X(u,v)\}_{u \in [1,s_1], v \in [1,s_2]})$.

1. For $v = 1, 2, \ldots, s_2$, calculate $\hat{Y}(v)$ as follows.

$$\hat{Y}(v) = \text{StableEst}^{(p)}(\{X(u,v)\}_{u=1,\ldots,s_1}) \ .$$

2. Return the estimate $\hat{F}_{p,q}$ as follows.

$$\hat{F}_{p,q} = \text{StableEst}^{(q)}(\{\hat{Y}(v) \mid v = 1,\ldots,s_2\})$$

Fig. 1. Algorithm BILINSTABLE for estimating $F_{p,q}$

4.1 Analysis

In this section, we present an analysis of the bi-linear stable sketch algorithm. The cases, $p = 0$ and $q = 0$ are considered separately.

Lemma 6. *For each* $0 < p \leq 2$, $0 < q < 1$ *and* $\epsilon < 1/8$, *the estimator* $\text{BILINSTABLE}(p, q, s_1, s_2, \{X(u,v)\}_{u \in [1,s_1], v \in [1,s_2]})$ *with parameters* $s_2 = \frac{K_L(q)}{\epsilon^2}$ *and* $s_1 = \frac{K_L(p)}{\epsilon^2} \log \frac{1}{\epsilon}$ *satisfies* $|\hat{F}_{p,q} - F_{p,q}| \leq 3\epsilon F_{p,q}$ *with probability* $\frac{7}{8}$. *The constant* $K_L(p)$ *is the constant of Li's geometric means estimator for p-stable sketches, given by* (10).

Proof. Fix a value of v and for this value of v, let y be a value of the random vector $\xi(v)$ obtained by choosing $\xi_j(v)$ randomly from the stable distribution $S(q, 1)$, for each $j = 1, 2, \ldots, n$ and independently. Denote the random variable $X(u, v)$ conditional on the choice $\xi(v) = y$ as $X(u, v|\xi(v) = y)$. Therefore,

$$\begin{aligned} X(u, v|\xi(v) = y) &= \sum_{j=1}^{n} \sum_{i=1}^{n} A_{i,j} x_{i,j}(u, v) y_j^{1/p} \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} \left(A_{i,j} y_j^{1/p} \right) x_{i,j}(u, v) \end{aligned}$$

Moreover, it is important to note that the random variables $X(u, v|\xi(v) = y)$ are independent since the random variables $\{x_{i,j}(u, v)\}_{1 \leq i,j,u \leq n}$ are independent.

So we have by standard property of stable distributions that

$$X(u, v|\xi(v) = y) \sim S(p, b(y))$$

where,

$$b(y) = \left(\sum_{j=1}^{n} \sum_{i=1}^{n} |A_{i,j} y_j^{1/p}|^p \right)^{1/p} = \left(\sum_{j=1}^{n} y_j \sum_{i=1}^{n} |A_{i,j}|^p \right)^{1/p} = \left(\sum_{j=1}^{n} y_j (\|A_j\|_p)^p \right)^{1/p} .$$

The second equality (crucially) uses the fact that for $0 < q < 1$, the stable distribution $S(q, 1)$ has non-negative support implying that y_j is non-negative.

Let $\hat{Y}(v \mid \xi(v) = y)$ be the random variable obtained by applying StableEst to the values $X(1, v|\xi(v) = y), \ldots, X(s_1, v|\xi = y)$. We now choose Li's estimator and accordingly set $s_1 = K_L \epsilon^{-2} \log(1/\delta'))$, where, $K = K_L$ is the constant for Li's estimator. By properties of StableEst we have,

$$\hat{Y}(v \mid \xi(v) = y) = \lambda_v(y) \sum_{j=1}^{n} (\|A_j\|_p)^p y_j,$$

where, $\Pr\{1 - \epsilon \leq \lambda_v(y) \leq 1 + \epsilon\} \geq 1 - \delta'$. Therefore,

$$\hat{Y}(v) = \lambda_v(\xi(v)) \sum_{j=1}^{n} (\|A_j\|_p)^p \xi_j(v), \tag{12}$$

where, $\Pr\{1-\epsilon \le \lambda_v(\xi(v)) \le 1+\epsilon\} \ge 1-\delta'$.

The next step in the estimator of Figure 1 is to apply $\text{StableEst}^{(q)}$ to the set of random variables $\left\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\right\}$. To analyze this step, let us consider the StableEst estimators of Indyk and Li, denoted by $\text{StableEst}_\text{I}$ and $\text{StableEst}_\text{L}$ respectively. Using Indyk's median estimator,

$$\begin{aligned}\text{StableEst}_\text{I}^{(q)}\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\} &= C_I \text{median}_{v=1}^{s_2}\left\{|\hat{Y}(v)|^q\right\}\\ &= C_I \text{median}_{v=1}^{s_2}\left\{|\lambda_v(\xi(v))|^q \left|\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v)\right|^q\right\} .\end{aligned}$$

Since, $\lambda_v(\xi(v)) \in [1-\epsilon, 1+\epsilon]$ with prob. $1-\delta'$, we have

$$\text{StableEst}_\text{I}^{(q)}\left\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\right\} \in (1\pm\epsilon)^q \cdot C_I \text{median}_{v=1}^{s_2}\left\{\left|\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v)\right|^q\right\} \quad \text{with prob. } 1-s_2\delta' \ .$$

Since,

$$C_I \text{median}_{v=1}^{s_2}\left\{\left|\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v)\right|^q\right\} = \text{StableEst}_\text{I}^{(q)}\left\{\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v) \mid v = 1,2,\ldots,s_2\right\}$$

it follows that

$$\text{StableEst}_\text{I}^{(q)}\left\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\right\} \in (1\pm\epsilon)^q \text{StableEst}_\text{I}^{(q)}\left\{\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v) \mid v = 1,2,\ldots,s_2\right\} \quad \text{with prob. } 1-s_2\delta'. \quad (13)$$

A similar analysis can be done for Li's estimator.

$$\begin{aligned}\text{StableEst}_\text{L}^{(q)}\left\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\right\} &= C_L \prod_{v=1}^{s_2}\left|\hat{Y}(v)\right|^{q/s_2}\\ &= C_L \prod_{v=1}^{s_2}\left||\lambda_v(\xi(v))|^{q/s_2}\left|\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v)\right|^{q/s_2}\right.\end{aligned}$$

Since, $\lambda_v(\xi(v)) \in [1-\epsilon, 1+\epsilon]$ with prob. $1-\delta'$, we have

$$\text{StableEst}_\text{L}^{(q)}\left\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\right\} \in C_L \prod_{v=1}^{s_2}(1\pm\epsilon)^{q/s_2}\left|\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v)\right|^{q/s_2} \quad \text{with prob. } 1-s_2\delta'$$

Therefore,

$$\text{StableEst}_\text{L}^{(q)}\left\{\hat{Y}(v) \mid v = 1,2,\ldots,s_2\right\} \in (1\pm\epsilon)^q \text{StableEst}_\text{L}^{(q)}\left\{\sum_{j=1}^{n}(\|A_j\|_p)^p \xi_j(v) \mid v = 1,2,\ldots,s_2\right\} \quad \text{with prob. } 1-s_2\delta'. \quad (14)$$

The forms of equations (13) and (14) are similar and so we drop the subscript I or L.

Since $\xi_j(v) \sim S(q, 1)$ and independent, and $F_{p,q}(A) = \sum_{j=1}^{n} (\|A_j\|_p)^p$, it follows that

$$\sum_{j=1}^{n} (\|A_j\|_p)^p \xi_j(v) \sim S(q, (F_{p,q}(A))^{1/q}) \ .$$

We can now use one of the StableEst algorithms, namely, Indyk's estimator or Li's estimator. Let $s_2 = \frac{K}{\epsilon^2}$, where, $K = K_I$ if we use Indyk's stable estimator or $K = K_L$ for Li's estimator.

$$\text{StableEst}^{(q)} \left\{ \sum_{j=1}^{n} (\|A_j\|_p)^p \xi_j(v) \mid v = 1, 2, \ldots, s_2 \right\} \in (1 \pm \epsilon) \sum_{j=1}^{n} (\|A_j\|_p)^p$$

with probability 15/16. Combining with (13) or (14), we have,

$$\text{StableEst}^{(q)} \left\{ \hat{Y}(v) \mid v = 1, 2, \ldots, s_2 \right\} \in (1 \pm \epsilon)^{q+1} F_{p,q}(A) \text{ with prob. } 1 - s_2 \delta' - \frac{1}{16} \ . \quad (15)$$

Letting $\delta' < 1/(16 s_2)$, the success probability of the above equation becomes at least 14/16. Since, $\hat{F}_{p,q}(A)$ is defined as $\text{StableEst}^{(q)} \left\{ \hat{Y}(v) \mid v = 1, 2, \ldots, s_2 \right\}$, and $\epsilon \le 1/8$, we have,

$$\left| \hat{F}_{p,q}(A) - F_{p,q}(A) \right| \le 4\epsilon F_{p,q}(A) \qquad \text{with prob. } 7/8 \ .$$

□

4.2 Boundary cases

The above method does not work for estimating $F_{p,q}$ when, either $q = 1$ or when either p or q is 0. The first case, namely, $q = 1$ is not solved using the above method since, all families of stable distribution with stability parameter 1 (i.e., the Cauchy distributions) have negative support. That is, if $\xi_j \sim S(1, 1)$, then, ξ_j could be negative and so the bilinear summand $A_{i,j} x_{i,j} \xi_j^{1/p}$ may not be a real number. However, as was discussed in Section 3, the estimation for $F_{p,1}$ for the case $p \in [0, 2]$ can be performed nearly optimally in terms of space by viewing A as a single long vector of dimension n^2 and using the one-dimensional frequency moment estimation algorithm.

The second problem case arises when either p or q is 0, since, stable distributions are not known for these parameters. We address this case next. A solution to these issues is obtained by approximating $F_{p,q}$ by $F_{p',q'}$, where, p' and q' are chosen to be appropriately close to p to q respectively. Lemma 7 presents the statement of this claim.

Lemma 7. *For every* $\epsilon < 1/8$, $0 \le p \le 1$ *and* $0 \le q \le 1$

$$F_{p',q'} \ge F_{p,q} \ge (1 - 2\epsilon) F_{p',q'} \quad (16)$$

where, $p' = \max(p, t)$, $q' = \max(q, \epsilon)$ *and* $t \le \frac{\epsilon}{\log F_{1,1}}$.

Proof. By viewing the expression $F_{p,q}$ as a function of q and expanding $F_{p,q'}$ around $F_{p,q}$ for $q' > q$ using Taylor's series, we obtain

$$F_{p,q'} \le F_{p,q} + (q' - q) F_{p,q'} \ln F_{p,q'} \quad (17)$$

since,

$$\frac{d}{dx}F_{p,x} = \frac{d}{dx}\sum_{j=1}^{n}(F_p(A_j))^x = \sum_{j=1}^{n}(F_p(A_j))^x \ln F_p(A_j) \ .$$

For $0 \le p \le 1$ and $q' < 1$, we have $F_{p,q'} \le F_{1,1}$. Substituting in (17), we have

$$F_{p,q'} \ge F_{p,q} \ge F_{p,q'}\left(1 - \frac{q'-q}{\ln F_{1,1}}\right) \ . \tag{18}$$

By viewing $F_{p,q'}$ as a function of p and using Taylor's series to expand $F_{p,q'}$ around p for $p' > p$, we have,

$$F_{p',q'} \le F_{p,q'} + (p'-p)q'F_{p',q'} \ln F_{1,1} \ .$$

Therefore,

$$F_{p',q'} \ge F_{p,q'} \ge F_{p',q'}(1 - q'(p'-p)\ln F_{1,1}) \ . \tag{19}$$

Substituting from (18), we have,

$$F_{p',q'} \ge F_{p,q} \ge F_{p',q'}\left(1 - \frac{q'-q}{\ln F_{1,1}}\right)\left(1 - q'(p'-p)\ln F_{1,1}\right) \ . \tag{20}$$

By choosing $q' = \max(q, \epsilon)$ yields $\frac{q'-q}{\ln F_{1,1}} \le \frac{\epsilon}{\ln F_{1,1}}$. Now suppose p' is chosen to be $\max(p,t)$ where, $t \le \frac{\epsilon}{\ln F_{1,1}}$. This implies

$$1 - q'(p'-p)\ln F_{1,1} \ge 1 - q'\epsilon \ge 1 - \epsilon$$

since $q' = \max(q,\epsilon) \le 1$. Substituting into (20) we obtain

$$F_{p',q'} \ge F_{p,q} \ge F_{p',q'}(1-\epsilon)^2 \ge (1-2\epsilon)F_{p',q'}$$

□

By Lemma 7, to obtain an ϵ-approximation to $F_{0,0}$, it suffices to obtain an $\epsilon/2$-approximation to $F_{\epsilon/\log F_{1,1}, \epsilon/\log F_{1,1}}$.

Following discussion in [14], q-stable sketches can be simulated using $O((1/q)\log n)$ bits of precision before and after the binary point. This follows from Levy's classical theorem on stable distribution: if $X \sim S(q,1)$ then $\Pr\left\{|X| < C_q n^{c/q}\right\} > 1 - 1/n^c$ for any $c > 0$, where, C_q is a constant dependent on q and is bounded above by an absolute constant. Thus, it is possible to approximate a single q-stable random variable using $O(c(1/q)\log n)$ random bits such that the resulting computation has error probability at most n^{1-c}.

Reducing random bits. There are $n^2 \cdot s_1 \cdot s_2$ p-stable random variables and $n \cdot s_2$ q-stable random variables. The random bits required under normal processing is $O(cs_2 \log n((1/p)s_1 n^2 + (1/q)s_2)n)$ that generates the necessary random variates with a distribution D such that the ℓ_1 difference of D from the corresponding true stable distribution is at most n^{-c}. For large enough constant c, the difference is negligible. We now use a technique of Indyk [14] to reduce the number of random bits. We briefly review Indyk's technique with regards to our problem.

First envision that the input stream is reordered so that all updates to a given matrix entry $A_{i,j}$ arrive consecutively. Then, for each element (i,j), the stable random variables $x_{i,j}(u,v)$ and $\xi_j(v)$ are computed from a set of independent random bits and the corresponding sketches are updated. The algorithm uses n^2 chunks of random bits, one chunk for each (i,j) and each chunk is of the size of $R = O(s_1 s_2 \log n(1/p) + s_2 \log n(1/q))$ bits. Denote the chunks as $\bar{X}_1, \ldots, \bar{X}_{n^2}$. The space requirement for storing the sketches is say S bits, where, $S = O(K_L(p)K_L(q)\epsilon^{-4}(\log \epsilon^{-1})(\log F_{1,1}))$. Now Nisan's pseudorandom generator (PRG) [20] for fooling space bounded Turing machines can be used to design a PRG G that expands $O(S \log R)$ bits to a sequence of n^2 chunks of size R bits each, denoted by $\widetilde{X}_1, \ldots, \widetilde{X}_{n^2}$. The construction of G guarantees that using $\widetilde{X}_j$ instead of $\bar{X}_j$ results in negligible error probability ($2^{-O(S)}$). Thus, in the ordered stream, the update corresponding to matrix entry (i,j) is updated using the random bits in $\widetilde{X}_{i,j}$. Since the difference is negligible, the pseudo-random sketches can be used to estimate the hybrid moment $F_{p,q}(A)$. Finally, Indyk observes that the sketches are updated using addition, which is a commutative and associative operation. Hence, G can be used just as well for the original stream that is arbitrarily ordered. We also note that the PRG G of Nisan is efficient in the sense that any S-length chunk $\widetilde{X}_j$ can be computed using $O(\log R)$ arithmetic operations over $O(S)$-bit words.

This gives us the following theorem. The constants in the space complexity expression are independent of p, q and n.

Theorem 1. *For every $p \in [0,2]$ and $q \in [0,1]$ and $\epsilon \leq 1/8$, there exists a randomized algorithm that returns $\hat{F}_{p,q}$ satisfying $|\hat{F}_{p,q} - F_{p,q}| < \epsilon F_{p,q}$ with probability $3/4$ using space $O(S \log(n^2))$, where, $S = O((\ln F_{1,1})\epsilon^{-4}) \log(\epsilon)^{-1})$.* □

5 Estimating hybrid moments: $F_{p,q}$ for $p \in [0,2]$, $q \in (1,2]$

In this section, we consider the problem of estimating the frequency moment $F_{p,q}(A)$, when $p \in [0,2]$ and $q \in (1,2]$.

We design a data structure ESTFREQ(p,k,δ) that processes the stream updates. Here $p \in [0,2]$, the matrix A is updated as a coordinate-wise stream, k is a space parameter k and δ is a confidence parameter. After the stream is processed, given any column index $j \in \{1,2,\ldots,n\}$ of the matrix A, the structure returns an estimate $\hat{F}_p(A_j)$ of $F_p(A_j)$ satisfying

$$|\hat{F}_p(A_j) - F_p(A_j)| \leq \frac{F_{p,1}(A)}{k}$$

with probability $1-\delta$. We first present the design of this structure.

5.1 The EstFreq data structure

The ESTFREQ(p,k,δ) data structure keeps a collection of $t = O(\log(1/\delta))$ hash tables $T_1, \ldots, T_t$, each consisting of $b = 8k$ buckets numbered $0, \ldots, b-1$. Associated with each hash table T_k is a hash function $h_k : \{1,\ldots,n\} \rightarrow \{0,\ldots,b-1\}$. The hash functions $\{h_k\}_{1 \leq k \leq t}$ are each drawn independently from a pair-wise independent family of hash functions. Associated with each hash table T_k we keep a family of p-stable random variables

$$\{x_{i,j,u,k} \mid 1 \leq i,j \leq n, 1 \leq u \leq U, 1 \leq k \leq t\}$$

where, $U = \Theta(1/\epsilon^2)$. We will assume that for any given i, j, u, k, a pseudo-random generator can be used to obtain the value of $x_{i,j,u,k}$ along the lines discussed by Indyk in [14]. Each bucket of a table T_k is an array of U p-stable sketches of the form

$$T_k[b, u] = \sum_{h(j)=b} \sum_{i=1}^{n} A_{i,j} x_{i,j,u,k}, \qquad u = 1, 2, \ldots, U \ .$$

Each stream update of the form (index, i, j, Δ) is processed as follows.

Update(i, j, Δ)
for k **:=** 1 **to** t **do**
    b **:=** $h_k(j)$
    for u **:=** 1 **to** U **do**
        $T_k[b, u] := T_k[b, u] + \Delta \cdot x_{i,j,u,k}$
    endfor
endfor

The estimator for $F_p(A_j)$ is defined as follows. First, an estimate for $F_p(A_j)$ is obtained from each of the t tables and then the median of these estimates is returned. An estimate is obtained from each table T_k by first mapping j to its bucket $b = h_k(j)$ and then returning the StableEst of the p-stable sketches associated with this bucket as follows. Finally, the median of these estimates is returned. That is,

$$\hat{F}_p(A_j) = \text{median}_{k=1}^{t} \text{StableEst}^{(p)}(\{T_k[h_k(j), u]\}_{u=1,2,\ldots,U})$$

We will now analyze the data structure.

Lemma 8. *Let the number of buckets in each hash table of the* ESTFREQ(p, k, A) *structure be* $8k$ *and the number of hash tables be* $O(\log(1/\delta))$. *Also suppose that the number of stable sketches in each bucket of the hash tables is* $O(1/\epsilon^2)$. *Then,*

$$|\hat{F}_p(A_j) - F_p(A_j)| < \frac{\epsilon}{2} F_p(A_j) + \frac{(1+\epsilon/2)}{k} F_{p,1}(A)$$

with probability $1 - \delta$.

Proof. Fix a column A_j and fix a table T_k. Consider the bucket $b = h_k(j)$ to which A_j maps in this table. Let $X = X_{j,k}$ denote the following random variable.

$$X_{j,k} = \sum_{h_k(j')=h_k(j)} F_p(A_{j'}) \ .$$

It follows from the pair-wise independence of h_k that

$$\mathsf{E}\big[X - F_p(A_j)\big] = \frac{1}{8k}(F_{p,1}(A) - F_p(A_j)) \ .$$

By Markov's inequality,

$$\Pr\{X - F_p(A_j) > F_{p,1}(A)/k\} < 1/8 \ . \qquad (21)$$

Let $Y_k = \text{StableEst}^{(p)}(\{T_k[h_k(j), u]\}_{u=1,2,\ldots,U})$. Then, $|Y_k - X| \le \epsilon X$ with probability $1-1/16$ (say) since there are $O(1/\epsilon^2)$ p-stable sketches in each bucket. Conditional on the event $|Y_k - X| \le \epsilon X$, we have

$$\begin{aligned} |Y_k - F_p(A_j)| &\le \epsilon X + (X - F_p(A_j)) \\ &= (1+\epsilon)(X - F_p(A_j)) + \epsilon F_p(A_j) \\ &\le \frac{(1+\epsilon)F_{p,1}(A)}{k} + \epsilon F_p(A_j) \ . \end{aligned}$$

where the last inequality holds with probability $1 - 1/8 - 1/8 = 3/4$ by union bound. Unconditioning the dependence on the event $|Y_k - X| \le \epsilon X$ which holds with probability $1 - 1/16$ the success probability is at least $3/4 - 1/16 = 11/16$. By classical Chernoff's bounds, the probability of success can be boosted to $1 - \delta$ by returning the median of $O(\log(1/\delta))$ independent measurements.

Let ϵ be $\epsilon/2$ to obtain the statement of the lemma by increasing the number of stable sketches per bucket by a constant factor. □

5.2 Estimating $F_{p,q}$

In this section, we use the EstFreq structure in conjunction with the Hss technique to estimate $F_{p,q}$ for $p \in [0, 2]$ and $q \in (1, 2]$.

We will instantiate the Hss technique to use an EstFreq(p, k, δ) data structure at level $l = 0$ and an EstFreq$(p, 4k, \delta)$ structure as the frequent items structure at each level $l = 1, \ldots, L$. Set $\delta = 1/n^2$. Define the thresholds as follows. Let $\bar{\epsilon} = \epsilon/(4q)$.

$$T_0 = \frac{F_{p,1}}{k\bar{\epsilon}} \text{ and } T_l = \frac{T_0}{2^l} \ .$$

The groups are defined as follows.

$$G_0 = \{A_j \mid F_p(A_j) \ge T_0\} \text{ and } G_l = \{A_j \mid T_l < F_p(A_j) \le T_{l-1}\}$$

The function to be estimated is

$$\Psi(A) = \sum_{j=1}^{n} (F_p(A_j))^q \ .$$

We can now directly use the properties of the Hss technique to calculate the error.

Lemma 9.

$$\mathsf{Var}\big[\bar{\Psi} \mid \text{GoodEst}\big] \le \frac{4F_{p,1}F_{p,2q-1}}{\bar{\epsilon}k} \ .$$

Therefore, $\mathcal{E}_1 \le \epsilon F_{p,q}$ provided, $k \ge \frac{36 \cdot n^{1-1/q}}{q \cdot \epsilon^3}$.

Proof. By Lemma 3,

$$\begin{aligned} \mathsf{Var}\big[\bar{\Psi} \mid \text{GoodEst}\big] &\le \sum_{\substack{i \in [n] \\ i \notin (G_0 - lmargin(G_0))}} \psi^2(f_i) \cdot 2^{l(i)+1} \\ &= \sum_{A_j \in lmargin(G_0)} 2(F_p(A_j))^{2q} + \sum_{l=1}^{L} \sum_{A_j \in G_l} (F_p(A_j))^{2q} \cdot 2^{l+1} \end{aligned} \quad (22)$$

We first consider the second summation expression above.

$$\begin{aligned}\sum_{l=1}^{L}\sum_{A_j\in G_l}(F_p(A_j))^{2q}\cdot 2^{l+1} &\le \sum_{l=1}^{L}\sum_{A_j\in G_l}(T_{l-1})(F_p(A_j))^{2q-1}\cdot 2^{l+1}\\ &\le \sum_{l=1}^{L}\sum_{A_j\in G_l}\frac{T_0}{2^{l-1}}(F_p(A_j))^{2q-1}\cdot 2^{l+1}\\ &\le 4T_0\sum_{l=1}^{L}(F_p(A_j))^{2q-1} \quad . \end{aligned} \tag{23}$$

The first summand of (22) simplifies to

$$\begin{aligned}\sum_{A_j\in lmargin(G_0)} 2(F_p(A_j))^{2q} &\le 2T_0(1+\bar\epsilon)\sum_{A_j\in lmargin(G_0)}(F_p(A_j))^{2q-1}\\ &\le 4T_0\sum_{A_j\in lmargin(G_0)}(F_p(A_j))^{2q-1} \quad .\end{aligned}$$

Adding with the *RHS* of (23), we have

$$\begin{aligned}\mathsf{Var}\big[\bar\Psi \mid \textsc{GoodEst}\big] &\le 4T_0\sum_{A_j\in lmargin(G_0)}(F_p(A_j))^{2q-1} + 4T_0\sum_{l=1}^{L}\sum_{A_j\in G_l}(F_p(A_j))^{2q-1}\\ &\le 4T_0F_{p,2q-1}(A) = \frac{4F_{p,1}F_{p,2q-1}}{\bar\epsilon k} \quad . \end{aligned}\tag{24}$$

We can now obtain an upper bound on $\mathcal{E}_1$. Using the definition of $\mathcal{E}_1$ and (24), we obtain

$$\mathcal{E}_1 \le 3(\mathsf{Var}\big[\bar\Psi\big])^{1/2} \le 6\left(\frac{F_{p,1}F_{p,2q-1}}{\bar\epsilon k}\right)^{1/2} \quad .$$

Using standard identities, $F_{p,1} \le n^{1-1/q}F_{p,q}^{1/q}$. Further,

$$\begin{aligned}F_{p,2q-1} &= \sum_{j=1}^{n}(F_p(A_j))^{2q-1} \le \left(\max_{j=1}^{n}(F_p(A_j))^{q-1}\right)\sum_{j=1}^{n}(F_p(A_j))^{q}\\ &\le \left(\max_{j=1}^{n}(F_p(A_j))^{q}\right)^{(q-1)/q} F_{p,q}(A)\\ &\le \left(\sum_{j=1}^{n}(F_p(A_j))^{q}\right)^{(q-1)/q} F_{p,q}(A)\\ &= F_{p,q}^{2-1/q} \quad .\end{aligned}$$

Therefore,

$$\mathcal{E}_2 \le 3\left(\frac{n^{1-1/q}(F_{p,q})^2}{\bar\epsilon k}\right)^{1/2} \le \epsilon F_{p,q}$$

provided,

$$k = \frac{36\cdot n^{1-1/q}}{\bar\epsilon\epsilon^2} = \frac{36\cdot n^{1-1/q}}{q\cdot\epsilon^3} \quad .$$

This proves the lemma. □

As is usual in most calculations involving the Hss technique, the dominant error is the variance of $\bar{\Psi}$, whereas, the error $\mathcal{E}_2$ is minor. The same property is seen in this instance as well.

Lemma 10. *If $k \geq n^{1-1/q}$ and $\bar{\epsilon} \leq \epsilon/(4q)$, then, $\Pi_1 \leq \epsilon F_{p,q}$ and $\Pi_2 \leq \epsilon F_{p,q}$.*

Proof. Recall that the function $\pi : [n] \to \mathbb{R}$ is defined as follows.

$$\pi_i = \begin{cases} \Delta_{l(i)} \cdot |\psi'(\xi_i(f_i, \Delta_l))| & \text{if } i \in G_0 - \mathit{lmargin}(G_0) \text{ or } i \in \text{mid}(G_l) \\ \Delta_{l(i)} \cdot |\psi'(\xi_i(f_i, \Delta_l))| & \text{if } i \in \mathit{lmargin}(G_l), \text{ for some } l > 1 \\ \Delta_{l(i)-1} \cdot |\psi'(\xi_i(f_i, \Delta_{l-1}))| & \text{if } i \in \mathit{rmargin}(G_l) \end{cases}$$

where, the notation $\xi_i(f_i, \Delta_l)$ returns the value of t that maximizes $|\psi'(t)|$ in the interval $[f_i - \Delta_l, f_i + \Delta_l]$.

Therefore, if $A_j \in G_l$, then,

$$\begin{aligned} \pi_{A_j} &\leq \Delta_{l-1}(F_p(A_j)(1+\bar{\epsilon}))^{q-1} \\ &\leq 2\bar{\epsilon} F_p(A_j)(F_p(A_j))^{q-1} \leq \epsilon F_p(A_j) \end{aligned}$$

since, $(1+\bar{\epsilon})^{q-1} \leq 2$ by the choice of $\bar{\epsilon} = \epsilon/(4q)$. Therefore,

$$\Pi_1 \leq \epsilon F_{p,1}(A) \ .$$

Similarly, if $A_j \in G_l$, then,

$$\pi^2_{A_j} \leq 2\bar{\epsilon}^2 \frac{F_{p,1}(A)}{2^l \cdot k} (F_p(A_j))^{2q-1} \cdot 2^{l+1} \leq 4\bar{\epsilon}^2 \frac{F_{p,1}(A)(F_p(A_j))^{2q-1}}{k} \ .$$

Therefore,

$$\begin{aligned} \Pi_2 &\leq \Big(\sum_{\substack{j\in[n] \\ A_j \notin \mathit{lmargin}(G_0)}} \pi^2_{A_j} 2^{l(i)+1} \Big)^{1/2} \\ &\leq 2\bar{\epsilon} \left(\frac{F_{p,1}(A) F_{p,2q-1}(A)}{k} \right)^{1/2} \\ &\leq 2\bar{\epsilon} \frac{n^{1-1/q} F_{p,q}(A)}{k} \leq \epsilon F_{p,q}(A) \end{aligned}$$

since, $k \geq n^{1-1/q}$. □

We therefore have the following theorem. An additional factor of $\log n + \log(1/\epsilon)$ arises due to the derandomization using Nisan's PRG [20] in the manner used by Indyk [14].

Theorem 2. *For each $p \in (0, 2]$ and $q \in (1, 2]$, there exists an algorithm that estimates $F_{p,q}(A)$ to within relative accuracy of ϵ using space*

$$O\left(\frac{n^{1-1/q}}{\epsilon^3} (\log n)^2 (\log(n/\epsilon)) \right)$$

with probability at least $7/8$. □

Lower Bounds. Some lower bounds may be obtained quite simply for the problem of estimating $F_{p,q}$ by reducing the problem of estimating the pqth one-dimensional moment $F_{p \cdot q}$ to $F_{p,q}$ as follows [22]. Consider an n-dimensional vector a and view it as the first row of the $n \times n$ matrix A, the rest of whose entries are zeros. Then, by definition, $F_{p,q}(A) = F_{p \cdot q}(a)$. Since, it is known that $F_{pq}(a)$ has a space lower bound of $\Omega(n^{1-2/(pq)})$ for $pq > 2$, the same holds for $F_{p,q}$ as well.

In particular, for $p = q = 2$, this reduction of $F_{pq}(a)$ to $F_{p,q}(A)$ implies a lower bound of space $\tilde{O}(\sqrt{n})$, which is the space required by the Hss algorithm of Section 5 (ignoring $(1/\epsilon^{O(1)})$ and poly-logarithmic factors).

For $pq \in [0, 2]$, F_{pq} has a lower bound of $\Omega(1/\epsilon^2)$ [21]. This implies that the bilinear stable sketches technique presented for the range $p \in [0, 2]$ and $q \in [0, 1]$ is close to optimal, up to polynomial factors in $1/\epsilon$ and poly-logarithmic factors in n and $F_{1,1}(A)$.

Recently, Jayram and Woodruff [16] have shown a space lower bound of $\Omega(n^{1-1/q})$ for estimating $F_{1,q}$ and $F_{0,q}$, whenever $q \geq 1$. This shows that the Hss algorithm described in this section for estimating $F_{p,q}$ is nearly space optimal for $p = 0$ or 1 and $q \in [1, 2]$. The problem of obtaining lower bounds for estimating $F_{p,q}$ for $p \in (0, 2)$ and $q \in (0, 2)$, (with the exception of the above cases) is open.

References

1. Noga Alon, Yossi Matias, and Mario Szegedy. "The space complexity of approximating frequency moments". *J. Comp. Sys. and Sc.*, 58(1):137–147, 1998.
2. Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. "An information statistics approach to data stream and communication complexity". In *Proceedings of ACM STOC*, pages 209–218, Princeton, NJ, 2002.
3. Z. Bar-Yossef, T.S. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan. "Counting distinct elements in a data stream". In *Proceedings of International Workshop on Randomization and Computation (RANDOM)*, Cambridge, MA, 2002.
4. L. Bhuvanagiri and S. Ganguly. "Estimating Entropy over Data Streams". In *Proceedings of European Symposium on Algorithms (ESA)*, pages 148–159, 2006.
5. A. Chakrabarti, S. Khot, and X. Sun. "Near-Optimal Lower Bounds on the Multi-Party Communication Complexity of Set Disjointness". In *Proceedings of International Conference on Computational Complexity (CCC)*, Aarhus, Denmark, 2003.
6. Moses Charikar, Kevin Chen, and Martin Farach-Colton. "Finding frequent items in data streams". *Theoretical Computer Science*, 312(1):3–15, 2004. Preliminary version appeared in Proceedings of ICALP 2002, pages 693-703.
7. Graham Cormode and S. Muthukrishnan. "An Improved Data Stream Summary: The Count-Min Sketch and its Applications". *J. Algorithms*, 55(1):58–75, April 2005.
8. Graham Cormode and S. Muthukrishnan. "Space Efficient Mining of Multigraph Streams". In *Proceedings of ACM International Symposium on Principles of Database Systems (PODS)*, 2005.
9. P. Flajolet and G.N. Martin. "Probabilistic Counting Algorithms for Database Applications". *J. Comp. Sys. and Sc.*, 31(2):182–209, 1985.
10. S. Ganguly and L. Bhuvanagiri. "Hierarchical Sampling from Sketches: Estimating Functions over Data Streams". *Algorithmica*, 53:549–582, 2009.
11. S. Ganguly and A. Majumder. "CR-precis: A Deterministic Summary Structure for Update Streams". In *Proceedings of Int'l Symp. on Algorithms, Probabilistic and Experimental Methodologies (ESCAPE), LNCS 4614*, pages 48–59, 2007.
12. Sumit Ganguly and Graham Cormode. "On Estimating Frequency Moments of Data Streams". In *Proceedings of International Workshop on Randomization and Computation (RANDOM)*, 2007.
13. P. B. Gibbons and S. Tirthapura. "Estimating simple functions on the union of data streams". In *Proceedings of ACM SPAA*, pages 281–291, Heraklion, Crete, Greece, 2001.

14. Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. Preliminary Version appeared in Proceedings of IEEE FOCS 2000, pages 189-197.
15. Piotr Indyk and David Woodruff. "Optimal Approximations of the Frequency Moments". In *Proceedings of ACM Symposium on Theory of Computing STOC*, pages 202–298, Baltimore, Maryland, USA, June 2005.
16. T.S. Jayram and David P. Woodruff. "The Data StreamSpace Complexity of Cascaded Norms". In *Proceedings of IEEE Foundations of Computer Science (FOCS)*, 2009.
17. Daniel M. Kane, Jelani Nelson, and David P. Woodruff. "On the Exact Space Complexity of Sketching and Streaming Small Norms". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, 2010.
18. Ping Li. Estimators and tail bounds for dimension reduction in ℓ_α $(0 < \alpha \leq 2)$ using stable random projections. In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, pages 10–19, 2008.
19. A. McGregor. "Open problems in data streams and related topics: IITK Workshop on Algorithms for Data Streams, 2006". http://www.cse.iitk.ac.in/users/sganguly/openproblems.pdf.
20. N. Nisan. "Pseudo-Random Generators for Space Bounded Computation". In *Proceedings of ACM Symposium on Theory of Computing STOC*, pages 204–212, May 1990.
21. David P. Woodruff. "Optimal space lower bounds for all frequency moments". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, pages 167–175, 2004.
22. David P. Woodruff. *Personal Communication*, August 2008.