Sketching, Moment Estimation, and the Lévy-Khintchine Representation Theorem

Seth Pettie
University of Michigan, Ann Arbor, MI, USA

Dingyu Wang
University of Michigan, Ann Arbor, MI, USA

Abstract

In the d-dimensional turnstile streaming model, a frequency vector $\mathbf{x} = (\mathbf{x}(1), \ldots, \mathbf{x}(n)) \in (\mathbb{R}^d)^n$ is updated entry-wisely over a stream. We consider the problem of *f-moment estimation* for which one wants to estimate

$$f(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{v\in[n]} f(\mathbf{x}(v))$$

with a small-space sketch. A function f is *tractable* if the f-moment can be estimated to within a constant factor using polylog(n) space.

The f-moment estimation problem has been intensively studied in the $d = 1$ case. Flajolet and Martin estimate the F_0-moment ($f(x) = \mathbb{1}(x > 0)$, incremental stream); Alon, Matias, and Szegedy estimate the L_2-moment ($f(x) = x^2$); Indyk estimates the L_α-moment ($f(x) = |x|^\alpha$), $\alpha \in (0, 2]$. For $d \geq 2$, Ganguly, Bansal, and Dube estimate the $L_{p,q}$ hybrid moment ($f : \mathbb{R}^d \to \mathbb{R}, f(x) = (\sum_{j=1}^d |x_j|^p)^q$), $p \in (0, 2], q \in (0, 1)$. For tractability, Bar-Yossef, Jayram, Kumar, and Sivakumar show that $f(x) = |x|^\alpha$ is *not* tractable for $\alpha > 2$. Braverman, Chestnut, Woodruff, and Yang characterize the class of tractable one-variable functions except for a class of *nearly periodic functions*.

In this work we present a simple and generic scheme to construct sketches with the novel idea of hashing indices to *Lévy processes*, from which one can estimate the f-moment $f(\mathbf{x})$ where f is the *characteristic exponent* of the Lévy process. The fundamental *Lévy-Khintchine representation theorem* completely characterizes the space of all possible characteristic exponents, which in turn characterizes the set of f-moments that can be estimated by this generic scheme.

The new scheme has strong explanatory power. It unifies the construction of many existing sketches (F_0, L_0, L_2, L_α, $L_{p,q}$, etc.) and it implies the tractability of many nearly periodic functions that were previously unclassified. Furthermore, the scheme can be conveniently generalized to multidimensional cases ($d \geq 2$) by considering multidimensional Lévy processes and can be further generalized to estimate *heterogeneous moments* by projecting different indices with different Lévy processes. We conjecture that the set of tractable functions can be characterized using the Lévy-Khintchine representation theorem via what we called the *Fourier-Hahn-Lévy* method.

2012 ACM Subject Classification Theory of computation → Sketching and sampling

Keywords and phrases Streaming Sketches, Lévy Processes

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.77

Related Version *Full Version*: https://arxiv.org/abs/2410.17426

Funding This work was supported by NSF Grant CCF-2221980.

1 Introduction

Streaming algorithms have been studied for over 40 years, beginning with Morris's [25] *approximate counter* and Munro and Patterson's [26] 1- and 2-pass *selection* algorithms, both from 1978. In 1983 Flajolet and Martin [14] designed and analyzed the first "modern" sketch called *Probabilistic Counting with Stochastic Averaging* (PCSA). PCSA works only in *incremental* streams whereas all linear sketches, such as the celebrated AMS sketch [1] for estimating L_2-moments, maintain a vector of elements subject to both increments and decrements. In this work we consider a generalized turnstile model, where each element has a value in $\mathbb{R}^d$ for some dimension $d \in \mathbb{Z}_+$.

▶ **Definition 1** ($\mathbb{R}^d$-turnstile model)**.** *Let* $[n] = \{1, 2, \ldots, n\}$ *be the universe. The frequency vector* $\mathbf{x} = (\mathbf{x}(1), \ldots, \mathbf{x}(n)) \in (\mathbb{R}^d)^n$ *is initialized as all zeros and gets updated by a stream of pairs of the form of* (v, y)*, where* $v \in [n]$ *and* $y \in \mathbb{R}^d$*.*

- Update(v, y)*,* $\mathbf{x}(v) \leftarrow \mathbf{x}(v) + y$*.*

In this paper we consider the following generic streaming estimation problem.

▶ **Problem 1** (f-moment estimation in the $\mathbb{R}^d$-turnstile model)**.** Fix $d \in \mathbb{Z}_+$. Let $f : \mathbb{R}^d \to \mathbb{C}$ with $f(0) = 0$. Let $\mathbf{x} \in (\mathbb{R}^d)^n$ be the frequency vector of the current stream. Estimate the f*-moment* $f(\mathbf{x})$, where

$$f(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{v \in [n]} f(\mathbf{x}(v)).$$

The f-moment estimation problem was originally posed by Alon, Matias, and Szegedy [1].[1] They estimated the L_2-moment ($f(x) = |x|^2$) with a data structure now commonly known as the AMS sketch, and then asked in general what f-moments can be estimated in polylog space. Indyk [17] designed a class of sketches for estimating the L_p moments, with $p \in (0, 2]$. When $p > 2$, Bar-Yossef, Jayram, Kumar, and Sivakumar [2] proved that estimating the L_p-moment requires $\Omega(\text{poly}(n))$ space. The L_0-moment ($f(x) = \mathbb{1}(x \neq 0)$) has been estimated in two distinct ways. Cormode, Data, Indyk, and Muthukrishnan [10] approximate L_0 by L_α with very small $\alpha > 0$. Kane, Nelson, and Woodruff [20] project each element onto $\mathbb{Z}_p$, $p > \epsilon^{-1} \log n$ being a random prime, which effectively reduces L_0-estimation to cardinality estimation over incremental streams (F_0-estimation).

Apart from the frequency moments L_p, $p \in [0, 2]$, there are other function moments that can be estimated over streams. In particular, a function f is called *tractable* if the f-moment can be approximated to within a $1 \pm \epsilon$ factor in $\text{poly}(\epsilon^{-1}, \log n)$ space. Characterizing the class of tractable functions (the *tractability problem*) has been a central theoretical problem in the streaming literature. Braverman and Ostrovsky [6] managed to characterize all tractable functions $f : \mathbb{Z} \to \mathbb{R}_+$ that are symmetric ($f(x) = f(-x)$) and increasing on $[0, \infty)$. Braverman and Chestnut [3] then charaterized all tractable functions that are symmetric, non-negative, and *decreasing* on $[1, \infty)$. Braverman, Chestnut, Woodruff, and Yang [4] extended the characterization to all symmetric functions, except for a class of "nearly periodic" functions.

The tractability question is closely related to the problem of designing *universal sketches*. A universal sketch specifies a family of functions $\mathcal{F}$ such that for *any* $f \in \mathcal{F}$, the f-moment can be estimated. Sketches in [6, 7, 4] are all universal for their respective classes; they are based on sampling L_2-heavy hitters from subsampled versions of the input vector, then applying f to the samples. Though in this work we focus on the polylog-space regime, there have also been intensive studies for sketching problems in the polynomial-space regime as well, e.g., L_p-moment with $p > 2$ [1, 5, 18, 8, 28, 15], cascaded norms [19], and Schatten norms [24].

[1] Unless specified otherwise, we assume the stream is one-dimensional and integral, i.e., $\mathbf{x} \in \mathbb{Z}^n$, when discussing prior works.

While most of the literature considers the one-dimensional case where $d = 1$, Ganguly, Bansal, and Dube [16] estimate the $L_{p,q}$ hybrid moment, which is defined to be $f(x_1, \ldots, x_d) = (\sum_{j=1}^{d} |x_j|^p)^q$, where $p \in [0,2], q \in [0,1]$. The $L_{p,q}$-moment is a meaningful class of multidimensional moments which include, e.g., the Euclidean length $f(x_1, \ldots, x_d) = \sqrt{\sum_{j=1}^{d} x_j^2}$, corresponding to the $L_{2,1/2}$-moment. If the elements are points in $\mathbb{R}^d$ subject to a stream of updates of the form of `(key, displacement)`, then the $L_{2,1/2}$-moment is equal to the sum of the Euclidean distances $\sum_{v\in[n]} |\mathbf{x}(u)|$. Multidimensional inputs can arise when querying *multiple* one-dimensional streams. For example, suppose we have two $\mathbb{R}$-turnstile streams with frequency vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. A query like "how many elements are non-zero in $\mathbf{x}$ and at least 3 in $\mathbf{y}$" can be written as a two-dimensional f-moment query with $f(x,y) = \mathbb{1}(x \neq 0)\, \mathbb{1}(y \geq 3)$.

Main Contribution

In this work, we uncover an intimate relation between turnstile sketches and *Lévy processes*. Lévy processes have been studied since the early 20th century, and have been used to model phenomena in various fields, e.g., physics (how does a gas particle move?) and finance (how does the stock price change?). We will show that all Lévy processes have *algorithmic* interpretations in the context of streaming sketches, and that the fundamental *Lévy-Khintchine representation theorem* leads to a unified view of sketching for moment estimation.

We give a detailed technical synopsis of Lévy processes in Section 2. For the time being, a Lévy process $(X_t)_{t\geq 0}$, where $X_t \in \mathbb{R}^d$, is defined by having independent, stationary increments. I.e., $(X_{t_1+t_2} - X_{t_1}) \sim X_{t_2}$, for any $t_1, t_2 \in \mathbb{R}_+$, and $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ are mutually independent for any $t_1 < \cdots < t_k$. Some common one-dimensional Lévy processes are

Linear drift. $X_t = \gamma t$ with drift rate γ.
Brownian motion. $X_t \sim \mathcal{N}(0, t\sigma^2)$ with variance σ^2.
Poisson process. $X_t \sim \text{Poisson}(\lambda t)$ with rate λ.
α-stable process. Defined by $X_t/t^{1/\alpha} \sim X_1 \sim \alpha$-stable. (For $\alpha \in (0,2]$, the α-stable random variable X_1 on $\mathbb{R}$ is defined by its characteristic function $\mathbb{E}e^{izX_1} = e^{-|z|^\alpha}$, for any $z \in \mathbb{R}$.)

These can all be generalized to higher dimensions. The Lévy-Khintchine representation theorem identifies every Lévy process X with its *characteristic exponent* $f_X : \mathbb{R}^d \to \mathbb{C}$, where for any $z \in \mathbb{R}^d$, $\mathbb{E}\exp(i\langle X_t, z\rangle) = \exp(-tf_X(z))$.

1.1 Lévy Processes and Sketching

We briefly explain why Lévy processes naturally lie at the heart of linear sketching. Suppose X is the random memory state of a *linear* sketch[2] of input vector $\mathbf{x}$ and $\hat{f}(X)$ is an estimate of the f-moment $f(\mathbf{x})$. Now consider the situation where the input is repeated w times over disjoint domains. Then by construction, the f-moment becomes $w \cdot f(\mathbf{x})$ and the estimate becomes $\hat{f}(X^{(1)} + \cdots + X^{(w)})$ where $X^{(j)}$ are i.i.d. copies of X.[3] In other words, we should have

$$\frac{\hat{f}(X^{(1)} + \cdots + X^{(w)})}{w} \approx f(\mathbf{x}),$$

[2] It suffices to consider linear sketches in the turnstile model [23].
[3] One can think of this as increasing the vector length from n to wn and letting $X^{(j)}$ be the sketch of $\mathbf{x}^{(j)}$, which is a copy of $\mathbf{x}$ on the index interval $[(j-1)n, jn)$.

for *any* $w \in \mathbb{Z}_+$ and also for the limiting case as $w \to \infty$. Therefore, no matter how complicated the distribution of the linear sketch may be, the sum $X^{(1)} + \cdots + X^{(w)}$ will converge[4] to some well-behaved limiting random process. For the AMS sketch [1], X happens to be sub-gaussian and therefore the normalized sum converges to a Gaussian. For Indyk's [17] stable sketch, X is α-stable, and therefore the normalized sum remains α-stable. These are merely two special cases of Lévy processes. In general, if $X = (X_t)_{t\in\mathbb{R}_+}$ is a Lévy process, then $X^{(1)} + \cdots + X^{(w)} = (X_t^{(1)} + \cdots + X_t^{(w)})_{t\in\mathbb{R}_+} \sim (X_{wt})_{t\in\mathbb{R}_+}$. Thus summing i.i.d. Lévy processes is equivalent to simply *rescaling time*. After normalizing the time scale,[5] we see Lévy processes are stable under i.i.d. sums. Lévy processes therefore form a mathematical closure of linear sketches in terms of their limiting distributions.

In practice it usually suffices to construct some algorithmically simple random projection in the *domain of attraction* of the limiting process. For example, the AMS sketch for estimating L_2 does not need to explicitly use Gaussians in its projection; Rademacher ($\{-1, 1\}$) random variables suffice. For a *prototypical solution*, it is convenient to consider linear sketches with projection sampled directly from Lévy processes. The distribution of the sketch will always lie in the space of Lévy processes and is therefore easier to track.

1.2 Random Projection with Lévy Processes

We now consider linear sketching with Lévy processes and show how different processes can be used to track different function moments. Throughout the paper we work in the *random oracle* model, in which we can evaluate uniformly random hash functions $H : [n] \to [0, 1]$.

Given any Lévy process $X = (X_t)_{t\in\mathbb{R}_+}$ on $\mathbb{R}^d$, by the Lévy-Khintchine representation theorem (Theorem 7), there exists a function $f_X : \mathbb{R}^d \to \mathbb{C}$ such that for any $t \in \mathbb{R}_+$ and $z \in \mathbb{R}^d$,

$$\mathbb{E}e^{i\langle z, X_t\rangle} = e^{-tf_X(z)}. \qquad \text{(Lévy-Khintchine)}$$

Suppose now we aggregate all the input in one register in the following way

$$C_t = \sum_{v\in[n]} \left\langle \mathbf{x}(v), X_t^{(v)} \right\rangle,$$

where the $X_t^{(v)}$ are i.i.d. copies of X_t. Clearly C_t is a linear sketch which can be maintained over a distributed stream. We thus have

$$\begin{aligned}
\mathbb{E}e^{iC_t} &= \mathbb{E}e^{i\sum_{v\in[n]}\left\langle X_t^{(v)}, \mathbf{x}(v)\right\rangle} \\
&= \prod_{v\in[n]} \mathbb{E}e^{i\left\langle X_t^{(v)}, \mathbf{x}(v)\right\rangle} && \text{(by independence)} \\
&= \prod_{v\in[n]} e^{-tf_X(\mathbf{x}(v))} && \text{(by Lévy-Khintchine)} \\
&= e^{-tf_X(\mathbf{x})}, && \text{(recall that } f(\mathbf{x}) = \textstyle\sum_{v\in[n]} f(\mathbf{x}(v)))
\end{aligned}$$

from which the f_X-moment $f_X(\mathbf{x})$ can be recovered by choosing a suitable time $t \approx \Theta(1/|f(\mathbf{x})|)$. Of course, we do not always know a suitable t, and therefore maintain samples for many t, evenly spaced on a logarithmic scale. This is the key observation that lets us

[4] with a proper normalization depending on X and w

[5] To see how this "time normalization" generalizes the typical scalar normalization of stable variables, note that for α-stable variables, the normalization is $w^{-1/\alpha}(X^{(1)} + \cdots + X^{(w)})$, which is equivalent to scale the time down by w since $w^{-1/\alpha}X_{wt} \sim X_t$ if $(X_t)_{t\in\mathbb{R}_+}$ is α-stable.

estimate every f-moment, where f is the characteristic exponent of a Lévy process. In addition, note that the statistic e^{iC_t} only depends on the value of $C_t \mod 2\pi$. Therefore, it suffices to store $C_t \mod 2\pi$ for the estimation scheme above, which is a *compact value* in the range $[0, 2\pi)$. See Figure 1.[6]

1.3 Stable Projections

A d-dimensional Lévy process X is called α*-stable* if for any $t \in \mathbb{R}_+$, $X_t \sim t^{1/\alpha} X_1$. If the Lévy process X in $C_t = \sum_{v\in[n]} \left\langle \mathbf{x}(v), X_t^{(v)} \right\rangle$ is α-stable, then we have

$$C_t = \sum_{v\in[n]} \left\langle \mathbf{x}(v), X_t^{(v)} \right\rangle \sim \sum_{v\in[n]} \left\langle \mathbf{x}(v), t^{1/\alpha} X_1^{(v)} \right\rangle \sim t^{1/\alpha} C_1.$$

In other words, the projection C_t is also α-stable. Due to this self-similarity, it suffices to consider the projection only at time 1 and there is no need for using multiple time scales. Now let the characteristic exponent be $f_X(z)$. Then we have the characteristic function

$$\mathbb{E}e^{izC_1} = \mathbb{E}e^{iz\sum_{v\in[n]}\left\langle \mathbf{x}(v), X_1^{(v)}\right\rangle} = \prod_{v\in[n]} \mathbb{E}e^{i\left\langle \mathbf{x}(v), zX_1^{(v)}\right\rangle} \qquad \text{(by independence)}$$

Since $X^{(v)}$ is α-stable, we have $zX_1^{(v)} \sim X_{z^\alpha}^{(v)}$

$$\begin{aligned}
&= \prod_{v\in[n]} \mathbb{E}e^{i\left\langle \mathbf{x}(v), X_{z^\alpha}^{(v)}\right\rangle} \\
&= \prod_{v\in[n]} e^{-z^\alpha f(\mathbf{x}(v))} && \text{(by Lévy-Khintchine)} \\
&= e^{-z^\alpha f(\mathbf{x})} = e^{-(zf(\mathbf{x})^{1/\alpha})^\alpha} \\
&= \mathbb{E}e^{izf(\mathbf{x})^{1/\alpha} Y_\alpha}, && \text{(by Lévy-Khintchine)}
\end{aligned}$$

where Y_α is a *one*-dimensional α-stable random variable with $\mathbb{E}e^{izY_\alpha} = e^{-|z|^\alpha}$. Since the characteristic function uniquely identifies the distribution, we know $C_1 \sim f(\mathbf{x})^{1/\alpha} Y_\alpha$. Note that a register in Indyk's [17] L_α-stable sketch distributes as $(\sum_{v\in[n]} |\mathbf{x}(v)|^\alpha)^{1/\alpha} Y_\alpha$. Thus m i.i.d. copies of C_1 *emulate* the distribution of Indyk's L_α-sketch where the L_α-moment is replaced by the f-moment. Such function moments can be estimated by Indyk's estimator [17] or Li's estimators [22] with the same error guarantee.

Note that such stable projection takes less space than the generic projection considered in Section 1.2 since in general C_t needs to be stored for for many t while stable projection only requires to store C_1. Next, we formally state the new results.

1.4 New Results

The main theorem connects the Lévy-Khintchine representation theorem for generic Lévy processes with f-moment estimation in the $\mathbb{R}^d$-turnstile model. See Table 1 for some notation. Note that for a vector $\mathbf{x} \in (\mathbb{R}^d)^n$ with multidimensional entries ($d \geq 2$), the L_1 and L_∞ norms use the Euclidean norm per entry. For example, $\|\mathbf{x}\|_1 = \sum_{v\in[n]} |\mathbf{x}(v)|$, where $|\mathbf{x}(v)| = \sqrt{\sum_{j=1}^d \mathbf{x}(v)_j^2}$.

---

[6] Strictly speaking, by wrapping C_t in $[0, 2\pi)$, the sketch becomes non-linear. It is still additive, and therefore allows simple merging of two sketches, but does not permit scalar multiplication.

Table 1 Notation.

Notation	Definition	Notes
$\lvert x\rvert$	$(\sum_{j=1}^{d} x_j^2)^{1/2}$	$x \in \mathbb{R}^d$, Euclidean norm
$\lvert z\rvert$	$(z\bar{z})^{1/2}$	$z \in \mathbb{C}$, modulus
$\lVert\mathbf{x}\rVert_0$	$\sum_{v\in[n]} \mathbb{1}(\mathbf{x}(v) \neq 0)$	$\mathbf{x} \in (\mathbb{R}^d)^n$, L_0-moment
$\lVert\mathbf{x}\rVert_1$	$\sum_{v\in[n]} \lvert\mathbf{x}(v)\rvert$	$\mathbf{x} \in (\mathbb{R}^d)^n$, L_1-moment
$\mathbb{T}$	complex unit circle	identified by $[0, 2\pi)$
$\mathbb{S}_{d-1}$	$\{x \in \mathbb{R}^d : \lvert x\rvert = 1\}$	unit sphere in $\mathbb{R}^d$
$\mathbb{V}Z$	$\mathbb{E}(Z - \mathbb{E}Z)\overline{(Z - \mathbb{E}Z)}$	variance of Z

▶ **Theorem 2** (generic Lévy-Tower, Section 3)**.** *Let $f : \mathbb{R}^d \to \mathbb{C}$ be any function of the form*

$$f(z) = \frac{1}{2}\langle z, Az\rangle - i\langle \gamma, z\rangle + \int_{\mathbb{R}^d} \left(1 + i\langle z, s\rangle \mathbb{1}(|s| < 1) - e^{i\langle z,s\rangle}\right) \nu(ds),$$

where A is a covariance matrix, $\gamma \in \mathbb{R}^d$, and ν is a positive measure over $\mathbb{R}^d$ with $\int_{\mathbb{R}^d} \min(|x|^2, 1)\, \nu(dx) < \infty$. There exists a mergeable sketch of $O(\epsilon^{-2} \log n)$ words such that for any input stream $\mathbf{x} \in (\mathbb{R}^d)^n$ with $|f(\mathbf{x})| \in [1, \mathrm{poly}(n)]$, it returns an estimate $\widehat{f(\mathbf{x})}$ that with probability 99/100, where

$$\left|\widehat{f(\mathbf{x})} - f(\mathbf{x})\right| \leq O(\epsilon|f(\mathbf{x})|).$$

▶ Remark 3. The Lévy-Tower sketch improves the understanding of tractability of one-dimension function moments. In particular, it implies the tractability of many nearly periodic functions that were not previously classified. See Section 5 for a discussion of the new result and the existing tractability results. It also implies the tractability of a large class of *multidimensional moments* which have not been considered before.

As discussed in Section 1.3, stable projections form a special subclass of this linear projection method. We prove the following *emulation theorem*, enabling us to extend any analysis of Indyk's [17] classic L_α-stable sketch to the estimation of more general stable moments.

▶ **Theorem 4** (Lévy-Stable, Section 3.3)**.** *Let f be the characteristic exponent of any d-dimensional α-stable process X. When $\alpha = 2$, $f(x) = e^{-\frac{1}{2}t\langle x,Ax\rangle}$, where A is a covariance matrix, and when $\alpha \in (0, 2)$, $f(x) = \int_{\mathbb{S}_{d-1}} |\langle x, \xi\rangle|^\alpha\, \mu(d\xi)$, where μ is a symmetric, positive measure on $\mathbb{S}_{d-1}$.*[7] *Let the Lévy-Stable sketch be parameterized by X. Given any input vector $\mathbf{x} \in (\mathbb{R}^d)^n$ and $\mathbf{x}' \in \mathbb{R}^n$ such that $f(\mathbf{x}) = \sum_{v\in[n]} |\mathbf{x}'(v)|^\alpha$, Lévy-Stable with input $\mathbf{x}$ and Indyk's L_α-stable sketch with input $\mathbf{x}'$ distribute identically.*

Theorem 4 serves as an illuminating example showing how the Lévy-Khintchine theorem helps to understand streaming sketching. Previously, only two classes of stable moments are considered: one dimensional L_α-moments by Indyk [17], and multidimensional $L_{p,q}$-moments by Ganguly et al. [16], which are sketched by *algorithmic tricks* of combining stable random variables. The Lévy-Stable sketch extends such tricks to *all stable processes* in a systematic way. For example, we can now estimate, using any estimator of Indyk's sketch [17, 22], the f-moment of an $\mathbb{R}^3$-turnstile stream with

[7] For the case $\alpha \in (0, 2)$, only symmetric processes are considered here for simplicity.

$$f(x) = \int_{\mathbb{S}_2} \frac{|\langle \xi, x \rangle|}{|\xi_1|^2 + |\xi_2| + |\xi_3|^{1/2}} \, d\xi,$$

where $\mathbb{S}_2$ is the unit $\mathbb{R}^3$-sphere.

Table 2 Sketches and Lévy processes. The target function f_X is the characteristic exponent of the corresponding Lévy process, i.e., $f_X(x) = -\log \mathbb{E}e^{i\langle x, X_1\rangle}$. Random oracle is assumed (i.e., hash functions are provided externally) and the space is measured in *words* where each word uses $O(\log n)$ bits. $O(\log n)$-size words suffice for the Lévy-Tower and Lévy-Stable sketches under the common assumption that the total number of updates is at most $\text{poly}(n)$. Since F_0 is the cardinality over incremental stream, the space can be compressed to $O(\epsilon^{-2})$ *bits* [20, 21, 27]. In [20], L_0 is approximated by $\sum_{v\in[n]} \mathbb{1}(P|\mathbf{x}(v))$ where P is a random prime. In [10], L_0 is approximated by $\sum_{v\in[n]} |\mathbf{x}(v)|^\alpha$ with a small α. See [17, 20] for space complexity for estimating F_0, L_0, L_α when hash functions are stored explicitly.

Sketch Citation	Target $f(x)$	Space	Corresponding Lévy Process
Flajolet & Martin [14] (F_0)	$\mathbb{1}(x > 0)$	$O(\epsilon^{-2})$	pure killed process
Alon, Matias, and Szegedy [1] (L_2)	$\|x\|^2$	$O(\epsilon^{-2})$	Brownian motion
Indyk [17] (L_α)	$\|x\|^\alpha$	$O(\epsilon^{-2})$	α-stable process
Kane, Nelson, and Woodruff [20] (L_0)	$\mathbb{1}(P\|x)$	$O(\epsilon^{-2}\log n)$	compound Poisson process over $\mathbb{Z}_P$, P random prime
Ganguly, Bansal, and Dube [16] ($L_{p,q}$)	$(\sum_{j\in[d]} \|x_j\|^p)^q$	$O(\epsilon^{-2})$	per-entry p-stable processes subordinated by per-row q-stable subordinators
Lévy-Tower (**new**)	f_X	$O(\epsilon^{-2}\log n)$	any Lévy process X over $\mathbb{R}^d$
Lévy-Stable (**new**)	f_X	$O(\epsilon^{-2})$	any stable Lévy process X over $\mathbb{R}^d$

Heterogeneous moments

It will become clear in the analysis that both Lévy-Tower and Lévy-Stable will work for *heterogeneous moments* $\sum_{v\in[n]} f_v(\mathbf{x}(v))$, where each element v can have its own function f_v. One only needs to simulate v's process $X^{(v)}$ according to f_v. For Lévy-Tower, f_v can be the characteristic exponent of any Lévy process. For Lévy-Stable with parameter $\alpha \in (0, 2]$, f_v can be the characteristic exponent of any α-stable Lévy process. For example, with $d = 2$ there are many 2-dimensional 1-stable processes to choose from, and one could have $f_1(x, y) = |x| + |y|$, $f_2(x, y) = |x + y|$, $f_3(x, y) = \sqrt{x^2 + y^2}$, each of which corresponds to a 2-dimensional 1-stable process.

1.5 Organization

We first review Lévy processes and the Lévy-Khintchine representation theorem in Section 2. Then present the Lévy-Tower and Lévy-Stable sketches in Section 3. In Section 4, we discuss the connection between previous sketches and Lévy processes in a greater detail (see Table 2 for a summary). In Section 5, we discuss the problem of characterizing the set of *tractable* functions, and describe the *Fourier-Hahn-Lévy* method for expanding the range of the Lévy-Tower beyond Lévy-Khintchine-representable functions. We conclude in Section 6 with a conjectured characterization of the class of tractable functions.

2 Preliminary: Lévy Processes

We use Sato's text [29] as our reference for the theory of Lévy processes.

2.1 Lévy Processes on $\mathbb{R}^d$

▶ **Definition 5** (Lévy processes [29, page 3]). *A random process $X = (X_t)_{t\in\mathbb{R}_+}$ on $\mathbb{R}^d$ is a* Lévy process *if it satisfies*

Stationary Increments. $X_{t+s} - X_t \sim X_s$ *for all* $t, s \in \mathbb{R}_+$.

Independent Increments. *for* $0 \le t_1 < t_2 \ldots < t_k$, $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ *are mutually independent*

Stochastic continuity. $X_0 = 0$ *almost surely and* $\lim_{t\searrow 0} \mathbb{P}(|X_t| > \epsilon) = 0$ *for any* $\epsilon > 0$.

▶ Remark 6. A process that satisfies stationary and independent increments is called *memoryless.*

The primary way to study Lévy processes is through their *characteristic functions.* See Table 1 for notations.

▶ **Theorem 7** (Lévy-Khintchine representation [29, page 37]). *Any Lévy process $X = (X_t)_{t\in\mathbb{R}_+}$ on $\mathbb{R}^d$ can be identified by a triplet (A, ν, γ) where A is a covariance matrix, ν is a measure on $\mathbb{R}$ such that*

$$\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} \min(|s|^2, 1)\, \nu(ds) < \infty, \tag{1}$$

and $\gamma \in \mathbb{R}^d$. The identification is through the characteristic function. For any $t \in \mathbb{R}_+$ and $z \in \mathbb{R}^d$,

$$\mathbb{E}e^{i\langle X_t, z\rangle} = \exp\left(-t\left(\frac{1}{2}\langle z, Az\rangle - i\langle \gamma, z\rangle + \int_{\mathbb{R}^d} \left(1 + i\langle z, s\rangle \mathbb{1}\left(|s| < 1\right) - e^{i\langle z,s\rangle}\right) \nu(ds)\right)\right). \tag{2}$$

Conversely, any triplet (A, ν, γ) where A is a covariance matrix, ν satisfies (1), and $\gamma \in \mathbb{R}^d$ corresponds to a Lévy process satisfying (2).

▶ Remark 8. In the one-dimensional case, the matrix A can be identified by the usual variance σ^2. Then we have for $z \in \mathbb{R}$,

$$\mathbb{E}e^{izX_t} = \exp\left(-t\left(\frac{1}{2}\sigma^2 z^2 - i\gamma z + \int_{\mathbb{R}} \left(1 + izs\mathbb{1}\left(|s| < 1\right) - e^{izs}\right) \nu(ds)\right)\right). \tag{3}$$

We call the exponent in (2) the *characteristic exponent*, denoted by $f_X(z) = -\log \mathbb{E}e^{i\langle X_1, z\rangle}$. Let us note some properties of the characteristic exponent f.

▶ **Lemma 9.** *Let X be any Lévy process on $\mathbb{R}^d$ and f be its characteristic exponent.*

- $\mathfrak{Re} f \ge 0$. $\mathfrak{Re} f = 0$ *if and only if X is a deterministic drift, i.e., $X_t = \gamma t$ for all $t \in \mathbb{R}_+$.*
- *For any $x \in \mathbb{R}^d$, $f(-x) = \overline{f(x)}$, the complex conjugate of $f(x)$.*

We list some common one-dimensional Lévy processes in Table 3 that will be frequently used later.

Table 3 Common one-dimensional Lévy processes.

Process	Characteristic Exponent	Notes
linear drift	$i\gamma z$	$\gamma \in \mathbb{R}_+$
α-stable	$\lvert z\rvert^\alpha$	$\alpha \in (0, 2]$
Poisson	$1 - e^{iz}$	
compound Poisson	$\int_{-\infty}^{\infty}(1 - e^{izs})\,\nu(ds)$	$\nu(\mathbb{R}) < \infty$ is the jump rate, $\nu(\mathbb{R})^{-1}\nu$ is the jump distribution
symmetric compound Poisson	$2\int_0^\infty(1 - \cos(zs))\,\nu(ds)$	$\nu(\mathbb{R}) < \infty$

3 Infinitely Divisible Sketches

3.1 Lévy-Tower Sketches

We now present a sketch that is induced by a *generic* Lévy process on $\mathbb{R}^d$, which consists of two parameters.

- An accuracy parameter $m \in \mathbb{Z}_+$ which corresponds to the number of subsketches in classic settings.
- A Lévy process $X = (X_t)_{t\in\mathbb{R}_+}$ on $\mathbb{R}^d$ with characteristic exponent $f(z) = -\log \mathbb{E}e^{i\langle X_1, z\rangle}$.

▶ **Definition 10** ((f, m)-Lévy-Tower)**.** *Let f be the characteristic exponent of a Lévy process X on $\mathbb{R}^d$ and $m \in \mathbb{Z}_+$. An (f, m)-*Lévy-Tower *is an infinite vector $S = (S_k^{(j)})_{k\in\mathbb{Z},j\in[m]} \subset \mathbb{T}^{\mathbb{Z}}$, initialized as all zero. For any element $v \in [n]$ and $y \in \mathbb{R}^d$, a vector update $\mathbf{x}(v) \leftarrow \mathbf{x}(v) + y$ is effected by:*

Update(v, y) **:** *For each $k \in \mathbb{Z}$ and $j \in [m]$, $S_k^{(j)} \leftarrow S_k^{(j)} + \left\langle y, X_{2^{-k}}^{(v,j)}\right\rangle \mod 2\pi$, where $X^{(v,j)} = (X_t^{(v,j)})_{t\in\mathbb{R}_+}$ is an i.i.d. copy of the Lévy process X with characteristic exponent f.*

▶ Remark 11. If $|f(\mathbf{x})|$ is in the range $[1, \mathrm{poly}(n)]$, then it suffices to store levels $k \in [0, O(\log n)]$. Therefore a Lévy-Tower takes $O(\log n)$ words of space and each word stores a number in $[0, 2\pi)$.

The construction is standard in sketch design: subsample, randomly project, and then sum, but with an important and powerful tweak that the two steps "subsample and random project" are replaced by a single step: sampling at time 2^{-k} of the Lévy process whose characteristic exponent is f. Consequently, such sketches can be used to estimate f-moments. We characterize the distribution of the sketch as follows.

▶ **Lemma 12.** *Fix a frequency vector $\mathbf{x} = (\mathbf{x}(v))_{v\in[n]}$. For any $k \in \mathbb{Z}$, we have*

$$S_k \sim \sum_{v\in[n]} \left\langle X_{2^{-k}}^{(v)}, \mathbf{x}(v)\right\rangle$$

$$\mathbb{E}e^{iS_k} = e^{-2^{-k} f(\mathbf{x})}.$$

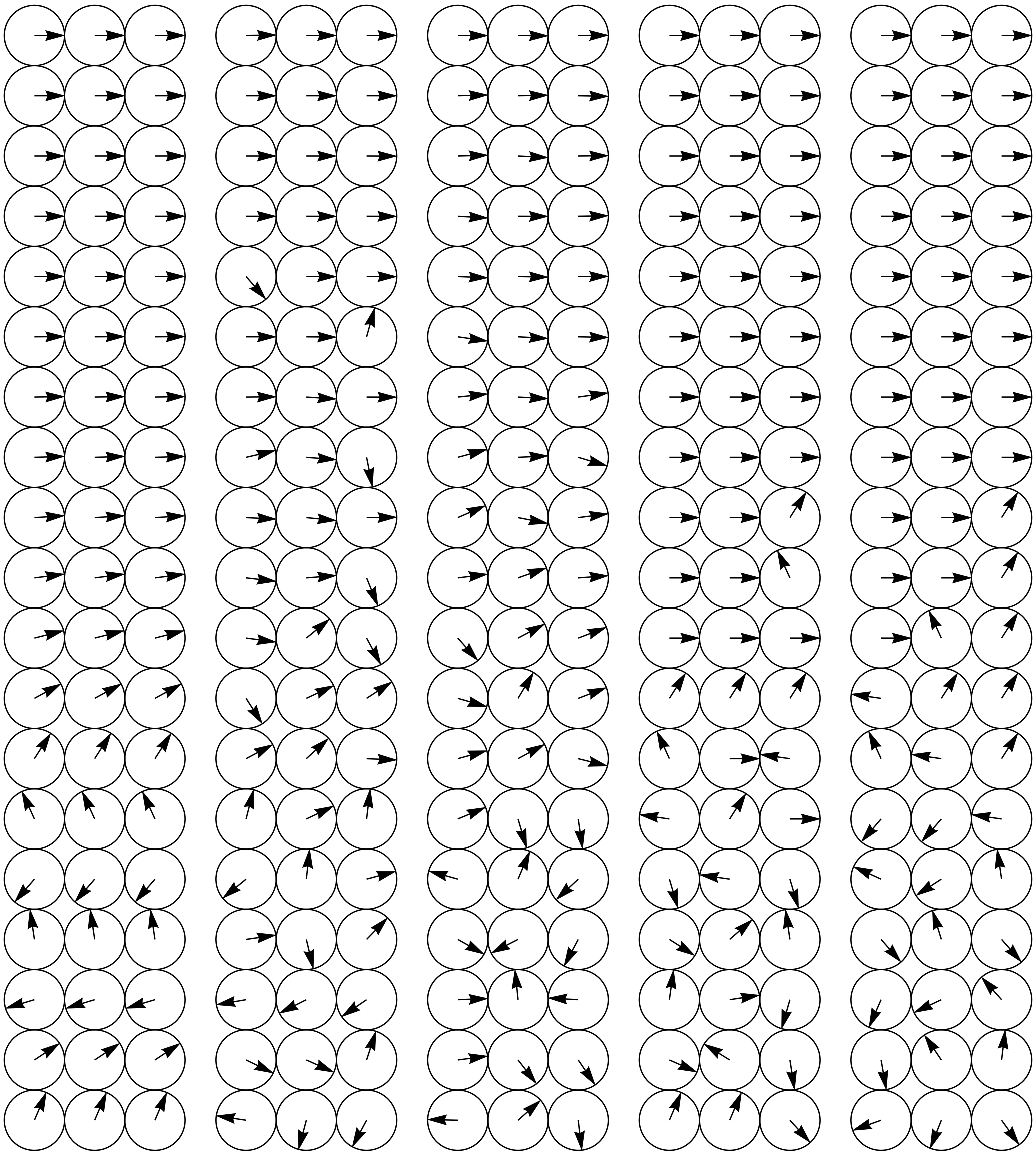

Figure 1 Lévy-Tower with $m = 3$ and $\mathbf{x} = (1, 0, \ldots, 0)$. From left to right: linear drift, Cauchy process, Brownian motion, Poisson process with rate 1, Poisson process with rate 2. Different Lévy processes have different "sensitivities" for a target function-moment. For example, linear drift is only sensitive to the sum of the vector and insensitive to how the values are distributed. The Cauchy process is only sensitive to the L_1-moment, while Brownian motion is only sensitive to the L_2-moment. Poisson processes are sensitive to the *support size* of $\mathbf{x}$ and at the same time leaking information about other f-moments.

Proof. The first statement is trivially true since the sketch is linear. For the second, note that the $X^{(v)}$ are i.i.d., and we have

$$\begin{aligned}
\mathbb{E}e^{iS_k} &= \mathbb{E}e^{i\sum_{v\in[n]}\left\langle X^{(v)}_{2^{-k}},\mathbf{x}(v)\right\rangle} \\
&= \prod_{v\in[n]} \mathbb{E}e^{i\left\langle X^{(v)}_{2^{-k}},\mathbf{x}(v)\right\rangle} && \text{(by independence)} \\
&= \prod_{v\in[n]} e^{-2^{-k}f(\mathbf{x}(v))} && \text{(by Lévy-Khintchine)} \\
&= e^{-2^{-k}f(\mathbf{x})}. && \text{(definition of } f(\mathbf{x})\text{)}
\end{aligned}$$

◀

Note that the f-moment lies exactly in the exponent of $\mathbb{E}e^{iS_k}$. We now formally present the estimation method.

3.2 Estimation of Lévy-Tower

As discussed in Section 1.2, the estimator needs to find a suitable level t and infer the f-moment from C_t. In the Lévy-Tower, we store $S_k = C_{2^{-k}}$ for all $k \in \mathbb{Z}$. We first prove the concentration of the empirical mean at each level.

▶ **Lemma 13.** *For any* $t \in \mathbb{R}_+$,

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m} e^{iC_t^{(j)}} - e^{-tf(\mathbf{x})}\right| > \eta\right) \le \frac{2t|f(\mathbf{x})|}{m\eta^2}.$$

Proof. By Chebyshev's inequality, we have

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^{m} e^{iC_t^{(j)}} - e^{-tf(\mathbf{x})}\right| > \eta\right) \le \frac{\mathbb{V}e^{iC_t}}{m\eta^2}.$$

Furthermore, since $\mathbb{E}e^{iC_t} = e^{-tf(\mathbf{x})}$, we have

$$\mathbb{V}e^{iC_t} = \mathbb{E}|e^{iC_t}|^2 - |\mathbb{E}e^{iC_t}|^2 = 1 - e^{-2t\mathfrak{Re}f(\mathbf{x})} \le 2t\mathfrak{Re}f(\mathbf{x}) \le 2t|f(\mathbf{x})|.$$ ◀

▶ **Lemma 14.** *With probability at least* $99/100$, *for any* $k \ge \log_2 |f(\mathbf{x})|$, *we have*

$$\left|\frac{1}{m}\sum_{j=1}^{m} e^{iS_k^{(j)}} - e^{-tf(\mathbf{x})}\right| \le O(1/\sqrt{m}).$$

Proof. By the union bound and Lemma 13,

$$\begin{aligned}
&\mathbb{P}\left(\exists k \ge \log_2 |f(\mathbf{x})|, \left|\frac{1}{m}\sum_{j=1}^{m} e^{iS_k^{(j)}} - e^{-tf(\mathbf{x})}\right| > \eta\right) \\
&\le \sum_{k=\log_2|f(\mathbf{x})|}^{\infty} \frac{2\cdot 2^{-k}|f(\mathbf{x})|}{m\eta^2} \\
&= O\left(\frac{1}{m\eta^2}\right).
\end{aligned}$$

Thus it suffices to choose $\eta = O(1/\sqrt{m})$ for the probability above at most $1/100$. ◀

We consider the complex logarithm defined on $\mathbb{C}$ except for the negative real line.

▶ **Lemma 15.** *Let $x, y \in \{z \in \mathbb{C} : |1 - z| < 1/2\}$, then $|\log x - \log y| < 2|x - y|$.*

▶ **Theorem 16.** *With probability at least 99/100, the following procedure returns an estimate $\widehat{f(\mathbf{x})}$ such that $|\widehat{f(\mathbf{x})} - f(\mathbf{x})| \le O(|f(\mathbf{x})|/\sqrt{m})$. Let K be the maximum level and assume $K \ge \log_2 |f(\mathbf{x})| \ge 0$.*

- *Enumerate $k = K, K-1, \ldots, 0$ in the decreasing order, let $Y_k = \frac{1}{m}\sum_{j=1}^{m} e^{iS_k^{(j)}}$.*
  - *If $|1 - Y_k| > 0.2$, then return $-2^k \log Y_k$ and halt.*

Proof. By Lemma 14, with probability at least 99/100, we have $\left|Y_k - e^{-tf(\mathbf{x})}\right| \le O(1/\sqrt{m})$ for all $k \in [\log_2 |f(\mathbf{x})|, K]$. We assume this event happens. Let W be the index selected in the procedure. First note that for $z \in \mathbb{C}$ with $\mathfrak{Re} z \ge 0$, one has $|1 - e^{-z}| \le |z|$. In addition, by Lemma 9, we know $\mathfrak{Re} f(\mathbf{x}) \ge 0$. Thus

$$|1 - Y_k| = |1 - e^{-tf(\mathbf{x})}| + O(1/\sqrt{m}) \le t|f(\mathbf{x})| + O(1/\sqrt{m}).$$

Let m be large enough that $O(1/\sqrt{m}) < 0.01$. Since the times are base 2, one has

$$\begin{aligned} 2^{-W} &\in [(0.2 - 0.01)/|f(\mathbf{x})|, 2(0.2 + 0.01)/|f(\mathbf{x})|] \\ &\subset (0.2/|f(\mathbf{x}))|, 0.45/|f(\mathbf{x}))|). \end{aligned}$$

Thus we have $|1 - Y_W| < 0.45 + O(1/\sqrt{m}) < 0.5$. By Lemma 15, we have

$$|-\log Y_W - 2^{-W} f(\mathbf{x})| \le 2|Y_W - e^{-2^{-W} f(\mathbf{x})}| \le O(1/\sqrt{m}).$$

We conclude that

$$|-2^W \log Y_W - f(\mathbf{x})| \le O(2^W/\sqrt{m}) = O(|f(\mathbf{x})|/\sqrt{m}).$$ ◀

3.3 Lévy-Stable Sketches

Recall that a d-dimensional Lévy process X is *α-stable* if for any $t \in \mathbb{R}_+$, $X_t \sim t^{1/\alpha} X_1$. We call the characteristic exponents of stable processes *stable moments*. Stable moments are of special interest in the context of streaming sketches since there is no need to store the whole tower; only m i.i.d. samples suffice to return an estimate with relative error $O(1/\sqrt{m})$.[8] The only symmetric one-dimensional stable processes are α-stable random processes, which correspond to Indyk's sketches [17]. On the other hand, there is a rich class of higher dimensional stable processes, the one implicit in Ganguly et al. [16] for estimating $L_{p,q}$ hybrid moments being just a single special case.

▶ **Theorem 17** (Lévy-Khintchine for stable processes [29, page 86])**.** *Let X be a Lévy process on $\mathbb{R}^d$ and $x \in \mathbb{R}^d$. X is 2-stable if and only if $\mathbb{E}e^{i\langle x, X_t\rangle} = e^{-\frac{1}{2}t\langle x, Ax\rangle}$ for some covariance matrix A. If X is symmetric, then it is α-stable for $\alpha \in (0, 2)$ if and only if there is a finite, positive measure μ on the sphere $\mathbb{S}_{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$ such that*

$$\mathbb{E}e^{i\langle x, X_t\rangle} = \exp\left(-t \int_{\mathbb{S}_{d-1}} |\langle x, \xi\rangle|^\alpha \, \mu(d\xi)\right).$$

[8] As discussed in Section 1.3, the Lévy-Towers induced by stable processes are *self-similar*: all registers in the tower are identically distributed with a proper normalization.

▶ Remark 18. For simplicity we only consider symmetric processes here when $\alpha \in (0, 2)$ so that the characteristic exponent $f_X(x_1, \ldots, x_d) = f_X(|x_1|, \ldots, |x_d|)$ for any $x \in \mathbb{R}^d$. See [29] for the full characterization.

As discussed in Section 1.3, any multidimensional α-stable moment is as simple to estimate as the one-dimensional L_α-moment, since it is possible to maintain a random variable distributed as $f(\mathbf{x})^{1/\alpha} Y_\alpha$ where Y_α is a unit α-stable random variable. We now give the formal definition of the Lévy-Stable sketch for completeness.

▶ **Definition 19** ((f, m)-Lévy-Stable). *Let f be the characteristic exponent of an α-stable Lévy process X on $\mathbb{R}^d$ and $m \in \mathbb{Z}_+$. An (f, m)-Lévy-Stable sketch is a vector of m registers $T = (T_1, \ldots, T_m) \in \mathbb{R}^m$, initialized as all zero. For any element $v \in [n]$ and $y \in \mathbb{R}^d$, the vector update $\mathbf{x}(v) \leftarrow \mathbf{x}(v) + y$ is effected by:*

Update(v, y) **:** *For each $k \in [m]$, $T_k \leftarrow T_k + \left\langle y, X_k^{(v)} - X_{k-1}^{(v)} \right\rangle$, where $X^{(v)} = (X_t^{(v)})_{t \in \mathbb{R}_+}$ is an independent copy of the Lévy process X with characteristic exponent f, indexed by v.*

▶ Remark 20. By construction, the Lévy-Stable sketches emulates Indyk's sketch, storing m words where each word stores a real number of magnitude at most poly(n) with high probability, given $f(\mathbf{x}) \leq \text{poly}(n)$.

▶ Remark 21. We remark here on the differences between the Lévy-Tower sketch and the Lévy-Stable sketch.

- Lévy-Tower estimates generic function moment f_X where X is the characteristic exponent of *any* Lévy process. Lévy-Tower stores projections at multiple times of the process and each projection is wrapped around within $[0, 2\pi)$.
- Lévy-Stable estimates function moment f_X where X is the characteristic exponent of a *stable* Lévy process. Lévy-Stable stores projections at unit time of the process and each projection is in $\mathbb{R}$. Lévy-Stable generalizes the stable sketches of Indyk [17] and Ganguly et al. [16].

4 Previous Sketches and Lévy Processes

In this section, we will discuss in greater detail how previous sketching techniques are equivalent to sampling Lévy processes. The guiding question is: what will happen if the input vector is repeated a large number of times on disjoint supports?

4.1 Single-Level Aggregation and the Central Limit Theorem

We start from the AMS sketch by Alon, Matias, and Szegedy [1], where a cell Q stores

$$Q = \sum_{v \in [n]} \mathbf{x}(v) \xi_v.$$

Here $\mathbf{x}$ is a $\mathbb{Z}$-stream and the $\xi_v \in \{-1, 1\}$ are i.i.d. Rademacher random variables.[9] Now suppose the input vector is repeated w times on disjoint supports. Then the final state would be

$$M_w = Q_1 + Q_2 + \cdots + Q_w,$$

[9] Again, in [1]'s analysis, 4-wise independence hashing suffices to guarantee a good enough estimate but we need to assume they are i.i.d. here to talk about the exact distribution of the final state.

where the (Q_j) are i.i.d. copies of Q. Note that $\mathbb{E}Q = 0$ and $\mathbb{V}Q = \sum_{v\in[n]} \mathbf{x}(v)^2$. Thus as $w \to \infty$ the normalized final state $M_w/\sqrt{w}$ converges to a centered Gaussian random variable with variance $\sum_{v\in[n]} \mathbf{x}(v)^2$ in distribution, by the central limit theorem. Thus, in the regime as $w \to \infty$, the AMS sketch eventually becomes the same as a cell in the Lévy-Tower $\sum_{v\in[n]} \left\langle X_1^{(v)}, \mathbf{x}(v) \right\rangle$, where the $(X^{(v)})$ are i.i.d. one dimensional Brownian motion and $X_1^{(v)}$ is the value of the process at time $t = 1$. Of course, since Brownian motions are self-similar, one does not need to sample the Lévy processes at exponentially spaced intervals; samples from m independent processes at time $t = 1$ suffice to approximate $\sum_{v\in[n]} \mathbf{x}(v)^2$ with $O(1/m)$ relative variance. The characteristic exponent of Brownian motion is the target function for the L_2-moment: $f(x) = |x|^2$.

Another illuminating example is Ganguly's [15] L_k-moment estimator for $k \geq 3$. Ganguly randomly projects the elements and stores $S = \sum_{v\in[n]} \mathbf{x}(v)Z^{(v)}$ where each $Z^{(v)}$ is a random root of $x^k = 1$ on the complex unit circle. The statistic S^k seems to be a good estimator for F_k-moment since $\mathbb{E}S^k = \sum_{v\in[n]} \mathbf{x}(v)^k$. However, since this random projection has finite variance, the normalized sum $S/\sqrt{\sum_{v\in[n]} |\mathbf{x}(v)|^2}$ will converge to a complex Gaussian as the input vector is duplicated on disjoint supports. In the limit, the only information remaining in the sketch pertains to the L_2-moment. Indeed, one in fact needs the number of i.i.d. registers to grow polynomially in the support-size for it to work for F_k [15].

By the generalized central limit theorem, for whatever random variable Q that is produced by the current input $\mathbf{x}$, if for some sequences $(a_w)_{w\in\mathbb{N}}$ and $(b_w)_{w\in\mathbb{N}}$, $(M_w - b_w)/a_w$ converges to some non-degenerate random variable Y as $w \to \infty$, then Y has to be α-stable for $\alpha \in (0, 2]$. If in addition, $\mathbb{V}Q < \infty$, then Y is Gaussian. Non-Gaussian stable distributions are discussed later in Section 4.3.

4.2 Multi-Level Subsampling and the Poisson Limit Theorem

Another important sketching technique is *subsampling.* The idea is to devise a sketch that works for $\Theta(m)$ elements and then solve the generic case by subsampling the stream at rates 2^{-k} for $k \in \mathbb{N}$, one of which reduces it down to $\Theta(m)$ elements. Without loss of generality, suppose now we have m non-zero elements with distinct values $\mathbf{x}(1), \ldots, \mathbf{x}(m)$. Let w be the number of times we repeat the input stream and denote the repeated stream as $\mathbf{x}^w$. To obtain a $\Theta(m)$-size set of subsamples, one needs to subsample $\mathbf{x}^w$ with rate $1/w$. Let $Y_{w,j}$ be the indicator that the jth copy of $\mathbf{x}(1)$ is sampled where $\mathbb{E}Y_{w,j} = 1/w$. The number of elements with value $\mathbf{x}(1)$ in the subsampled set is

$$\sum_{j=1}^{w} Y_{w,j} \to \text{Poisson}(1) \text{ in distribution as } w \to \infty,$$

where we have invoked the Poisson limit theorem (see, e.g., [12, Theorem 3.6.1]) because $\mathbb{E}\sum_{j=1}^{w} Y_{w,j} = 1$ and $\max_{j=1}^{w} \mathbb{P}(Y_{w,j} \neq 0) = \frac{1}{w} \to 0$ as $w \to \infty$. Similarly, the number of elements with value $\mathbf{x}(j)$ is also Poisson(1) and the occurrences of different values are independent. It is well known that such a limiting distribution can be simulated algorithmically by duplicating every element a Poisson(1) number of times (see [14, 13, 27, 30]). On the other hand, such limit distributions can be equivalently simulated by hashing each element to a Poisson process and then sampling at different times. Thus all sketches based on subsampling can be simulated by the Lévy-Tower (without mapping to the complex unit circle $\mathbb{T}$) with the corresponding (compound) Poisson processes.

4.3 Stable Random Variables and Stable Processes

Whenever Q has a finite variance, $(Q_1 + Q_2 + \ldots + Q_w)/\sqrt{w}$ goes to Gaussian as $w \to \infty$. Nevertheless, with infinite variance, Q can lie in the domain of attraction of an α-stable distribution for any $\alpha \in (0, 2)$. Indeed, the use of stable random variables is another sketching technique that directly corresponds to Lévy processes: α-stable processes. Similar to the Gaussian case (Section 4.1), α-stable processes are self-similar so there is no need to store the whole tower with exponentially spaced sample times 2^{-k}. Rather, it suffices to sample m independent processes at time $t = 1$.

Indyk [17] uses one-dimensional α-stable random variables for $\alpha \in (0, 2]$ to estimate the L_p-moment, and indeed, the characteristic exponent of the α-stable process is $f(x) = |x|^\alpha$. Ganguly, Bansal, and Dube [16] consider the higher dimensional case, where each element has d attributes. We now show how to reconstruct [16]'s sketch from the perspective of Lévy processes. The target function is

$$f : \mathbb{R}^d \to \mathbb{C}, \quad f(x) = \left(\sum_{j=1}^{d} |x_j|^p\right)^q.$$

Such f-moments are called $L_{p,q}$ hybrid moments in [16]. By the subordination theorem [29, page 197], this function is the characteristic exponent of a vector of d independent p-stable processes that are subordinated by a common q-stable subordinator. That is,

$$X_t = (X^{[1]}_{Z_t}, \ldots, X^{[d]}_{Z_t}),$$

where Z is a q-stable subordinator and the $(X^{[j]})$ are i.i.d. p-stable processes. Note that since $X^{[j]}$ is p-stable, we have $X^{[j]}_{Z_t} \sim Z_t^{1/p} X^{[j]}_1$. Thus, we sample X at time 1 and get

$$X_1 \sim Z_1^{1/p} \left(X^{[1]}_1, \ldots, X^{[d]}_1\right).$$

Take the inner product and we have, for $x \in \mathbb{R}^d$,

$$\langle x, X_1 \rangle = Z_1^{1/p} \sum_{j=1}^{d} x_j X^{[j]}_1,$$

which is exactly the random projection defined in [16].

4.4 HyperLogLog, PCSA, and Pure Killed Processes

Cardinality sketches like HyperLogLog [13] and PCSA [14] that only allow increments also correspond to Lévy processes in a surprisingly natural way. We consider the number system $\mathbb{R} \cup \{\infty\}$, where for any $x \in \mathbb{R}$, $x + \infty = \infty + x = \infty$ and $\infty + \infty = \infty$. Such definitions extend to multiplication with natural numbers where $0 \cdot \infty = 0$ and $k \cdot \infty = \infty$ for any $k \in \mathbb{Z}_+$.

▶ **Definition 22** (pure killed processes). *A* pure killed process $X = (X_t)_{t \in \mathbb{R}_+}$ *with* kill rate $c > 0$ *is a Lévy process over* $\mathbb{R} \cup \{\infty\}$ *which can be simulated as follows.*

- *Sample a* kill time $Y \sim \mathrm{Exp}(c)$.
- $X_t = 0$ *if* $t < Y$ *and* $X_t = \infty$ *otherwise.*

In particular, we have $\mathbb{P}(X_t = \infty) = \mathbb{P}(Y \leq t) = 1 - e^{-ct}$.

Assume the insertion stream is $v_1, \ldots, v_T$, where each v_j is an element in the universe $[n]$. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a pure killed process with unit kill rate ($c = 1$). Similar to the turnstile case in Section 1.2, one stores the sum at time t,

$$C_t = \sum_{j=1}^{T} X_t^{(v_j)} = \sum_{v\in[n]} \mathbf{x}(v) \cdot X_t^{(v)},$$

where $X^{(v)}$s are i.i.d. copies of X. It is straightforward to see that S_k perfectly simulates a Poissonized PCSA cell [14, 27] (Table 4).

Table 4 HyperLogLog and PCSA can be considered as Lévy-Tower sketches with pure killed processes where the bit-or operation $(\{0,1\},\vee)$ is simulated with the extended real numbers $(\{0,\infty\},+)$. Note that, of course, such a reinterpretation does not upgrade HyperLogLog and PCSA to work over turnstile streams. The resulting Lévy-Tower is still incremental-only since $\infty - \infty$ is not defined.

bit operation in cardinality sketches	jumps of a pure killed process
$0 \vee 0 = 0$	$0 + 0 = 0$
$0 \vee 1 = 1$	$0 + \infty = \infty$
$1 \vee 0 = 1$	$\infty + 0 = \infty$
$1 \vee 1 = 1$	$\infty + \infty = \infty$

In fact, one computes

$$\mathbb{E}e^{iC_t} = \prod_{v\in[n]} \mathbb{E}e^{i\mathbf{x}(v)\cdot X_t^{(v)}} = \left(\mathbb{E}e^{iX_t}\right)^{\|\mathbf{x}\|_0} \tag{4}$$

since $0 + \cdots + 0 = 0$ and $\infty + \cdots + \infty = \infty$. By Definition 22, $X_t = 0$ with probability e^{-t}, hence $e^{iX_t} = 1$, and $X_t = \infty$ with probability $1 - e^{-t}$, in which case $e^{iX_t} = 0$.[10] Thus we have $\mathbb{E}e^{iX_t} = e^{-t}$. Inserting this back to Equation (4), we have

$$\mathbb{E}e^{iC_t} = e^{-t\|\mathbf{x}\|_0}.$$

This coincides with the observation in Section 1.2, where the estimation target $\|\mathbf{x}\|_0$ lies right in the exponent of $\mathbb{E}e^{iC_t}$.

We note that HyperLogLog and PCSA both correspond to the pure killed process, with different couplings of cells on different levels. In particular, HyperLogLog stores $(C_1, C_{1/2}, C_{1/4}, \ldots)$ where $C_{2^{-j}}$s are sampled from *a single process* (exactly as in the Lévy-Tower) while PCSA stores $(C_1, C_{1/2}, C_{1/4}, \ldots)$ where $C_{2^{-j}}$s are independent. Since a pure killed process will remain at ∞ after its first jump, for HyperLogLog, $(C_1, C_{1/2}, C_{1/4}, \ldots)$ will always be in the form of a prefix of ones (∞s) followed by all zeros. Therefore, HyperLogLog only needs to store the length of the prefix of ones.

4.5 Uniform Random Projection and the Integral Lévy-Tower

Indyk's L_p-stable sketches [17] are able to estimate the L_p-moment for $p \in (0, 2]$ but do not handle the L_0-moment. Cormode, Datar, Indyk, and Muthukrishnan [10] approximate the L_0-moment by the L_α-moment for very small $\alpha > 0$. This scheme needs a bound on $\|\mathbf{x}\|_\infty$, as a single unbounded update can raise the estimate to infinity. Kane, Nelson, and

[10] Here we take the Cesàro limit $e^{i\infty} = \lim_{T\to\infty} \frac{1}{T}\int_0^T e^{it}\,dt = 0$.

Woodruff [20] take another approach: subsample the stream to a suitable level, randomly project each element within the field $\mathbb{Z}_p$ for a random prime $p > \epsilon^{-1}\log(\|\mathbf{x}\|_\infty)$, and then use an (insertion-only) F_0 estimator. The random projection approach [20] also requires a bound on $\|\mathbf{x}\|_\infty$. As an L_0-moment sketch it has a certain failure probability (e.g., $p|\mathbf{x}(v)$ for many v) but it perfectly estimates a related quantity: the $f_{L_0,p}$-moment for $f_{L_0,p}(x) = \mathbb{1}\,(p \nmid x)$.

We show how the Lévy-Tower reconstructs [20]'s *trick* by *pure computation.* The target function $f(k) = \mathbb{1}\,(p \nmid k)$ for $k \in \mathbb{Z}$ can be written as

$$f(x) = \frac{1}{p}\sum_{j=0}^{p-1}(1 - \cos(2\pi jx/p)).$$

The Lévy process with characteristic exponent f is the compound Poisson process with jumps uniformly chosen from $-2\pi(p-1)/p, \ldots, 0, \ldots, 2\pi(p-1)/p$. See Figure 2. The corresponding Lévy-Tower reconstructs [20]'s sketch: subsample through Poisson processes and uniformly project over $\mathbb{Z}_p$ with the uniform random jumps. It is straightforward to see that the resulting Lévy-Tower is *integral* in the sense that $S_k \in \{2j\pi/p : j = 0, \ldots, p\}$ for any k and S_k can be identified by a $\mathbb{Z}_p$-value. Moreover, while [20]'s estimator needs to assume p is a prime, the Lévy-Tower will work no matter whether p is prime or not.

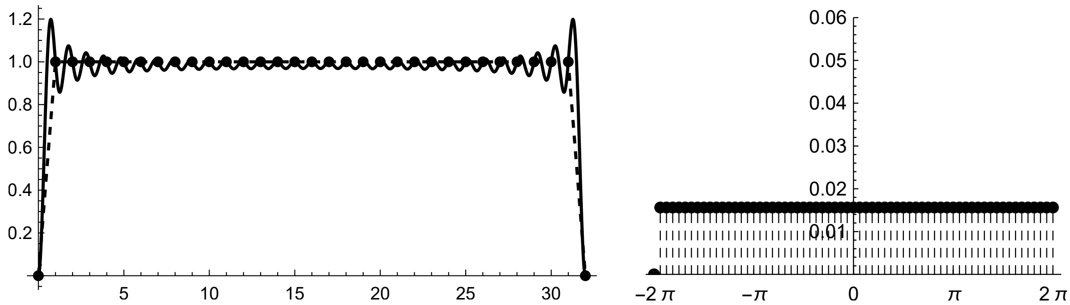

Figure 2 Left: $f_{L_0,32}(x) = \sum_{j=1}^{31}\frac{1}{32}(1-\cos(2\pi jx/32))$. The black dots mark the values at integer xs. Right: The jump distribution of the compound Poisson process X with characteristic exponent $f_{L_0,32}$. The subsampling and uniform random projection tricks used in [20] are recovered from computing the corresponding Lévy process.

5 Tractability and Lévy Processes

A function $f : \mathbb{Z} \to \mathbb{R}$ is *tractable* [4] if the f-moment of any input stream $\mathbf{x} \in \{-M, \ldots, M\}^n$, where $M = O(\text{poly}(n))$, can be $(1 \pm \epsilon)$-approximated with $O(\text{poly}(\epsilon^{-1}, \log n))$ space. Monotonic functions are fully characterized in [6] and all functions except for a class of nearly periodic functions are characterized in [4]. The Lévy-Tower tracks every function f that is the characteristic exponent of a Lévy process, and such exponents are characterized by Lévy-Khintchine. In this section, we investigate the set of tractable functions implied by the Lévy-Khintchine representation.

We start the discussion with a comparison between one-dimensional real characteristic exponents and existing communication complexity lower bounds. Note that by Lévy-Khintchine, such exponents can be written as $f(x) = \frac{1}{2}\sigma^2x^2 + 2\int_0^\infty(1 - \cos(xs))\,\nu(ds)$.

▶ **Lemma 23.** *Let $f(x) = \frac{1}{2}\sigma^2x^2 + 2\int_0^\infty(1 - \cos(xs))\,\nu(ds)$. The following statements are true.*

- *$f(x) \geq 0$ for any $x \in \mathbb{R}$. [Commentary: functions with both positive and negative values require $\Omega(\text{poly}(n))$-size sketches for constant factor approximations [9].]*

- *For any* $z \in \mathbb{Z}$ *and* $x \in \mathbb{R}$, $f(zx) \le z^2 f(x)$. *[Commentary: functions increasing faster than quadratic require* $\Omega(\text{poly}(n))$*-size sketches for constant factor approxiamtions [1, 2].]*
- *For any* $z \in \mathbb{R}_+$, $f(z) = 0$ *if and only if* $f(x+z) = f(x), \forall x \in \mathbb{R}$, *i.e.,* f *is periodic with period* z. *[Commentary: functions that have zeros other than the origin require* $\Omega(\text{poly}(n))$*-size sketches, unless the function is periodic with period* $\min\{z > 0 : f(z) = 0\}$ *[9].]*

Proof. The first statement is obvious. For the second statement, it suffices to prove that with $y = xs$,

$$(1 - \cos(zy)) \le z^2(1 - \cos(y)),$$

for any $y \in \mathbb{R}, z \in \mathbb{Z}$. When $z = 0$, the statement trivially holds. Without loss of generality, assume $z \in \mathbb{Z}_+$ and let $g(y) = z^2(1 - \cos(y)) - (1 - \cos(zy))$. Check that $g'(y) = z^2 \sin(y) - z\sin(zy)$ and $g''(y) = z^2\cos(y) - z^2\cos(zy)$. Clearly for $y \in [0, \pi/z]$, we have $\cos(zy) \le \cos(y)$ and thus $g''(y) \ge 0$. Note $g'(0) = 0$ and therefore $g(y) \ge 0$ for $y \in [0, \pi/z]$. Now consider $y \in [\pi/z, \pi]$, $z^2(1 - \cos(y)) \ge z^2(1 - \cos(\pi/z)) \ge 2 \ge 1 - \cos(zy)$, which implies $g(y) \ge 0$ on $[0, \pi]$. Thus we have $g(y) \ge 0$ for $[\pi, 2\pi]$ too by symmetry. The statement is thus proved since $g(y)$ is periodic in 2π.

For the third statement, $f(z) = 0$ necessarily implies that $\sigma = 0$ and the measure ν concentrates on $2\pi j/z$ for $j \in \mathbb{Z}$. Thus $f(x) = \sum_{j\in\mathbb{Z}}(1 - \cos(2\pi jx/z))\nu(\{2\pi j/z\})$ is periodic with period z. ◀

Braverman, Chestnut, Woodruff, and Yang [4] characterize the tractability of all symmetric functions over $\mathbb{Z}$, except for a class of nearly periodic functions. Specifically, the following example is given in [4, §5] which cannot be solved in the L_2-heavy hitter-based framework, but can be estimated using *ad hoc* algorithmic tricks.

▶ **Definition 24** (g_{np}, a nearly periodic function [4, §5])**.** *For* $x \in \mathbb{N}$, $g_{np}(x) = 2^{-\tau(x)}$, *where* $\tau(x) = \max\{j \in \mathbb{N} : 2^j | x\}$, *i.e., the position of the first "1" in the binary representation of* x.

The reason that g_{np} cannot be tracked by finding L_2-heavy hitters is that $g_{np}(x)$ can occasionally becomes polynomially small in x. In particular, for $k \in \mathbb{N}$, $g_{np}(2^k) = 2^{-k}$. We now demonstrate how the function g_{np} can be estimated *directly* by the Lévy-Tower.

5.1 Estimating g_{np} with the Lévy-Tower

Since it is assumed the element values are at most $\text{poly}(n)$, it suffices to consider a 2^w-periodic version of g_{np} where $w = O(\log(n))$.

▶ **Definition 25** (2^w-periodic g_{np})**.** *For* $x \in \mathbb{N}$,

$$g_{np,w}(x) = \begin{cases} g_{np}(x), & 2^w \nmid x \\ 0, & 2^w \mid x. \end{cases}$$

We show that, indeed, g_{np} corresponds to a Lévy process, specifically a compound Poisson process. The proof is included in the full version.

▶ **Lemma 26.** *Fix* $w \in \mathbb{N}$. *For any* $x \in \mathbb{N}$,

$$g_{np,w}(x) = \sum_{j=1}^{2^w-1} \frac{2^{2\tau(j)+1} + 1}{3 \cdot 2^{2w-1}}(1 - \cos(2\pi jx/2^w)).$$

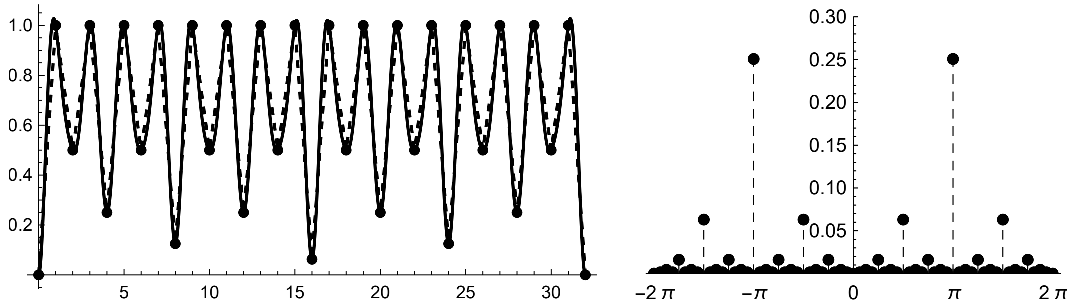

Figure 3 Left: $g_{np,5}(x) = \sum_{j=1}^{31} \frac{2^{2\tau(j)+1}+1}{1536}(1-\cos(2\pi jx/32))$. The black dots mark the values at integer xs. Right: The jump distribution of the compound Poisson process X with characteristic exponent $g_{np,5}$. Such nearly periodic functions do not fit in the L_2-heavy-hitter based framework in [4]. Nevertheless, one may compute the corresponding Lévy process and apply the Lévy-Tower.

The jump rate is

$$\sum_{j=1}^{2^w-1} \frac{2^{2\tau(j)+1}+1}{3\cdot 2^{2w-1}} = \frac{2^{2w}-1}{3\cdot 2^{2w-1}}.$$

By Lévy-Khintchine, $g_{np,w}$ is the characteristic function of a compound Poisson process X with jump rate $\frac{2^{2w}-1}{3\cdot 2^{2w-1}}$ and jump J distributed as

$$\mathbb{P}(J = 2\pi j/2^w) = \mathbb{P}(J = -2\pi j/2^w) = \frac{1}{2}\frac{2^{2\tau(j)+1}+1}{2^{2w}-1}, \quad \text{for } j = 1, \ldots 2^w - 1.$$

Moreover, the parameter w for the random jump induced by $g_{np,w}$ has a neat algorithmic interpretation as the max recursion depth if one implements each random jump by a recursive algorithm[11]. The case where $w \to \infty$ corresponds to the algorithm without depth limit. Let $\mathbb{B}$ be the set of all real fractions ($\in [0,1)$) with finite binary representations. Note that the jump distribution converges pointwisely to

$$\mathbb{P}(J = 2\pi x) = \mathbb{P}(J = -2\pi x) = \frac{1}{2}2^{1-2\tau_*(x)},$$

for any $x \in \mathbb{B}$, where $\tau_*(x)$ returns the location of the first one in the binary representation. For example $\tau_*(1/4) = \tau_*((0.01)_2) = 2$. We may thus simulate such jumps by the following algorithm to handle unbounded input streams with dynamically allocated memory.

▶ **Lemma 27.** *Let A, B be two random bit-tapes, embedded on reals with the binary expansion ($A = \sum_{j=1}^{\infty} A[j]2^{-j}, B = \sum_{j=1}^{\infty} B[j]2^{-j}$), and let $\xi \in \{-1, 1\}$ be a Rademacher random variable. Then*

$$J \sim 2\pi\,\xi \left(\sum_{j=1}^{\tau_*(B)-1} A[j]2^{-j} + 2^{-\tau_*(B)} \right).$$

[11](Halt with prob. 1/2; go left and repeat with prob. 1/4; and go right and repeat with prob. 1/4; see Figure 3).

5.2 The Fourier-Hahn-Lévy Method

The example above shows that it is possible to systematically sketch nearly periodic functions by sketching periodic functions and letting the period go to infinity. A natural conjecture is that the class of tractable functions are precisely those with a Lévy-Khintchine representation. This conjecture is *absolutely false*. We refute it with a simple tractable function having no Lévy-Khintchine representation, then attempt to salvage the conjecture via what we call the *Fourier-Hahn-Lévy* method.

▶ **Example 28** (the 0-1-5 problem)**.** Consider the function f where $f(0) = 0$, $f(1) = 1$, $f(k) = 5$ for all $k \geq 2$. The f-moment is easy to estimate to within a $1 \pm \epsilon$ factor in $O(\epsilon^{-2}\,\mathrm{polylog}(n))$ space, using ℓ_0-sampling [11]. However, no Lévy process has its characteristic function equal to f, for otherwise we would have $f(2 \cdot 1) \leq 2^2 f(1)$ by Lemma 23, which implies $5 \leq 4$ and is absurd.

We now extend our method to handle the *0-1-5 problem.* To be more concrete, we consider an arbitrary symmetric, real function f defined on $\{-p+1, \ldots 0, \ldots, p-1\}$ where $p > \|\mathbf{x}\|_\infty$ is the range of the values, say 2^{64}. We now describe the three-step *Fourier-Hahn-Lévy method.*

Fourier Transform. Decompose $f(0) - f(x) = \sum_{j=0}^{p-1}(1 - \cos(2\pi xj/p))\hat{f}(j)$ where the coefficient $\hat{f}(j)$ can be computed by taking the Fourier transform of f. Note that $\hat{f}$ is real since we have assumed f to be real and symmetric.

Hahn Decomposition. Decompose $\hat{f} = \hat{f}_+ - \hat{f}_-$ where $\hat{f}_+$ and $\hat{f}_-$ are non-negative real functions.

Lévy Process Simulation. Define functions $f_{\widehat{+}}, f_{\widehat{-}}$ as the inverse Fourier transform of $\hat{f}_+, \hat{f}_-$ respectively. Since the Fourier transform is linear, we have $f = f_{\widehat{+}} - f_{\widehat{-}}$. By construction, $f_{\widehat{+}}, f_{\widehat{-}}$ are all characteristic exponents of some Lévy processes. Estimate those function moments separately with Lévy-Tower sketches and sum to yield an estimate of the f-moment.

▶ **Lemma 29.** *Given any symmetric function $f : \mathbb{Z} \to \mathbb{R}_+$ with $f(0) = 0$ and any stream $\mathbf{x} \in [-p+1, p-1]^n$, the resulting estimator $\widehat{f(\mathbf{x})}$ of the Fourier-Hahn-Lévy method has error*

$$\left|\widehat{f(\mathbf{x})} - f(\mathbf{x})\right| \leq O((|f_{\widehat{+}}(\mathbf{x})| + |f_{\widehat{-}}(\mathbf{x})|)/\sqrt{m}),$$

with probability at least 98/100.

If f is the characteristic exponent of a Lévy process, then $f = f_{\widehat{+}}$ and $f_{\widehat{-}} = 0$. This case was already considered, where we get an asymptotically unbiased estimate with multiplicative error. Clearly, whenever $|f_{\widehat{+}}(\mathbf{x})| + |f_{\widehat{-}}(\mathbf{x})| = O(|f(\mathbf{x})|)$, the Fourier-Hahn-Lévy method *still* returns an estimate with multiplicative error, which handles the *0-1-5 problem*, among others.

6 Conclusion

The design of efficient streaming sketches has traditionally been more *art* than *science*, in which every sketching problem seemed to require a new insight or trick. Some prominent examples are F_0-sketching ($\{0, \infty\}$ projection [14]), L_2-sketching (sub-gaussian projection [1]), L_p-sketching (stable projection [17]), and L_0-sketching ($\mathbb{Z}_p$ uniform projection [20]); see also [16, 13].[12]

[12] There are *universal sketches* [6, 4] which can estimate a large class of function moments with a single sketch. But they consume much more space compared to classic sketches and the ones presented in this paper.

In this work we provide a more systematic method for generating sketches to estimate f-moments, by explicitly using the corresponding Lévy process whose characteristic exponent is f. This scheme mechanically *recreates* many known sketches [14, 1, 17, 16] and provides prototypical solutions to estimating nearly periodic moments, multidimensional moments, and heterogeneous moments. The key idea is to leverage the Lévy-Khintchine representation theorem for Lévy processes in the *design* and *analysis* of sketching algorithms.

Are there any f-moments that are *tractable* [6, 4] (they can be estimated to within a constant factor with a small sketch) but *do not* correspond to Lévy processes? In Section 5 we showed that the answer to this narrow question is *yes*, then introduced the *Fourier-Hahn-Lévy method*, which reduces an f without a Lévy-Khintchine representation to the difference of two functions $f_{\widehat{+}} - f_{\widehat{-}}$ that *can* be estimated by the Lévy-Tower. Does the Fourier-Hahn-Lévy method capture all tractable f-moments? We conjecture the answer is yes.

▶ **Conjecture 30.** *If $f : \mathbb{Z} \to \mathbb{C}$ is tractable, then the f-moment can be estimated to within a constant factor with a* Lévy-Tower*, either directly or after applying the Fourier-Hahn-Lévy method.*

References

1 Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999. doi:10.1006/jcss.1997.1545.

2 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. doi:10.1016/J.JCSS.2003.11.006.

3 Vladimir Braverman and Stephen R Chestnut. Universal sketches for the frequency negative moments and other decreasing streaming sums. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 591, 2015.

4 Vladimir Braverman, Stephen R. Chestnut, David P. Woodruff, and Lin F. Yang. Streaming space complexity of nearly all functions of one variable on frequency vectors. In *Proceedings 35th ACM Symposium on Principles of Database Systems (PODS)*, pages 261–276, 2016. doi:10.1145/2902251.2902282.

5 Vladimir Braverman, Jonathan Katzman, Charles Seidell, and Gregory Vorsanger. An optimal algorithm for large frequency moments using $O(n^{1-2/k})$ bits. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss-Dagstuhl – Leibniz Zentrum für Informatik, 2014. doi:10.4230/LIPIcs.APPROX-RANDOM.2014.531.

6 Vladimir Braverman and Rafail Ostrovsky. Zero-one frequency laws. In *Proceedings 42nd ACM Symposium on Theory of Computing (STOC)*, pages 281–290, 2010. doi:10.1145/1806689.1806729.

7 Vladimir Braverman and Rafail Ostrovsky. Generalizing the layering method of Indyk and Woodruff: Recursive sketches for frequency-based vectors on streams. In *Proceedings 16th International Workshop on Approximation, Randomization, and Combinatorial Optimization (APPROX)*, volume 8096 of *Lecture Notes in Computer Science*, pages 58–70. Springer, 2013. doi:10.1007/978-3-642-40328-6_5.

8 Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *18th IEEE Annual Conference on Computational Complexity, 2003. Proceedings.*, pages 107–117. IEEE, 2003. doi:10.1109/CCC.2003.1214414.

9 Stephen Robert Chestnut. *Stream sketches, sampling, and sabotage.* PhD thesis, Johns Hopkins University, 2015.

10 Graham Cormode, Mayur Datar, Piotr Indyk, and Shanmugavelayutham Muthukrishnan. Comparing data streams using Hamming norms (how to zero in). *IEEE Transactions on Knowledge and Data Engineering*, 15(3):529–540, 2003. doi:10.1109/TKDE.2003.1198388.

11 Graham Cormode and Donatella Firmani. A unifying framework for ℓ_0-sampling algorithms. *Distributed Parallel Databases*, 32(3):315–335, 2014. doi:10.1007/S10619-013-7131-9.

12 Rick Durrett. *Probability: theory and examples.* Cambridge University Press, 2019.

13 Philippe Flajolet, Éric Fusy, Olivier Gandouet, and Frédéric Meunier. HyperLogLog: the analysis of a near-optimal cardinality estimation algorithm. In *Proceedings of the 18th International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA)*, pages 127–146, 2007.

14 Philippe Flajolet and G. Nigel Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31(2):182–209, 1985. doi:10.1016/0022-0000(85)90041-8.

15 Sumit Ganguly. Estimating frequency moments of data streams using random linear combinations. In *Proceedings 8th International Workshop on Randomization and Computation (RANDOM)*, volume 3122 of *Lecture Notes in Computer Science*, pages 369–380, 2004. doi:10.1007/978-3-540-27821-4_33.

16 Sumit Ganguly, Mohit Bansal, and Shruti Dube. Estimating hybrid frequency moments of data streams. *Journal of Combinatorial Optimization*, 23(3):373–394, 2012. doi:10.1007/S10878-010-9339-1.

17 Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. doi:10.1145/1147954.1147955.

18 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, 2005. doi:10.1145/1060590.1060621.

19 T. S. Jayram and David P. Woodruff. The data stream space complexity of cascaded norms. In *Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '09, pages 765–774, USA, 2009. IEEE Computer Society. doi:10.1109/FOCS.2009.82.

20 Daniel M. Kane, Jelani Nelson, and David P. Woodruff. An optimal algorithm for the distinct elements problem. In *Proceedings 29th ACM Symposium on Principles of Database Systems (PODS)*, pages 41–52, 2010. doi:10.1145/1807085.1807094.

21 Kevin J. Lang. Back to the future: an even more nearly optimal cardinality estimation algorithm. *CoRR*, abs/1708.06839, 2017. arXiv:1708.06839.

22 Ping Li. Estimators and tail bounds for dimension reduction in l_α $(0 < \alpha \le 2)$ using stable random projections. In *Proceedings 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 10–19, 2008. URL: http://dl.acm.org/citation.cfm?id=1347082.1347084.

23 Yi Li, Huy L. Nguyen, and David P. Woodruff. Turnstile streaming algorithms might as well be linear sketches. In *Proceedings 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 174–183, 2014. doi:10.1145/2591796.2591812.

24 Yi Li and David P. Woodruff. Tight Bounds for Sketching the Operator Norm, Schatten Norms, and Subspace Embeddings. In Klaus Jansen, Claire Mathieu, José D. P. Rolim, and Chris Umans, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*, volume 60 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 39:1–39:11, Dagstuhl, Germany, 2016. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.APPROX-RANDOM.2016.39.

25 Robert Morris. Counting large number of events in small registers. *Communications of the ACM*, 21(10):840–842, 1978. doi:10.1145/359619.359627.

26 J. Ian Munro and Mike Paterson. Selection and sorting with limited storage. *Theor. Comput. Sci.*, 12:315–323, 1980. doi:10.1016/0304-3975(80)90061-4.

27 Seth Pettie and Dingyu Wang. Information theoretic limits of cardinality estimation: Fisher meets Shannon. In *Proceedings 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 556–569, 2021. doi:10.1145/3406325.3451032.

28 Eric Price and David P Woodruff. Applications of the Shannon-Hartley theorem to data streams and sparse recovery. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 2446–2450. IEEE, 2012. doi:10.1109/ISIT.2012.6283954.

29 Ken-Iti Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

30 Dingyu Wang and Seth Pettie. Better cardinality estimators for HyperLogLog, PCSA, and beyond. In *Proceedings 42nd ACM Symposium on Principles of Database Systems (PODS)*, pages 317–327, 2023. doi:10.1145/3584372.3588680.