Max–stable sketches: estimation of $\ell_\alpha-$norms, dominance norms and point queries for non–negative signals

Stilian A. Stoev
Department of Statistics
University of Michigan, Ann Arbor
sstoev@umich.edu

Murad S. Taqqu
Department of Mathematics and Statistics
Boston University
murad@bu.edu

February 10, 2006

Abstract

Let $f : \{1, 2, \ldots, N\} \to [0, \infty)$ be a non–negative signal, defined over a very large domain and suppose that we want to be able to address approximate aggregate queries or point queries about f. To answer queries about f, we introduce a new type of random sketches called *max–stable sketches*. The (ideal precision) max–stable sketch of f, $E_j(f)$, $1 \le j \le K$, is defined as:

$$E_j(f) := \max_{1 \le i \le N} f(i) Z_j(i), \quad 1 \le j \le K,$$

where the KN random variables $Z_j(i)$'s are independent with standard $\alpha-$Fréchet distribution, that is, $\mathbb{P}\{Z_j(i) \le x\} = \exp\{-x^{-\alpha}\}$, $x > 0$, where α is an arbitrary positive parameter. Max–stable sketches are particularly natural when dealing with maximally updated data streams, logs of record events and dominance norms or relations between large signals. By using only max–stable sketches of relatively small size $K << N$, we can compute in small space and time: *(i)* the $\ell_\alpha-$norm, $\alpha > 0$, of the signal *(ii)* the distance between two signals in a metric, related to the $\ell_\alpha-$norm, and *(iii)* dominance norms, that is, the norm of the maxima of several signals. In addition, we can also derive point queries about the signal.

As is the case of $p-$stable, $0 < p \le 2$, (sum–stable) additive sketches, see Indyk (2000), max–stability ensures that $E_j(f) \stackrel{d}{=} \|f\|_{\ell_\alpha} Z_1(1)$, with $\|f\|_{\ell_\alpha} = (\sum_{1 \le i \le N} f(i)^\alpha)^{1/\alpha}$, where $\stackrel{d}{=}$ means equal in distribution. We derive $\epsilon - \delta$ probability bounds on the relative error for distances and dominance norms. This can be implemented by efficient algorithms requiring small space even when the computational precision is finite and a limited amount of random bits are used. Our approach in approximating dominance norms improves considerably on existing techniques in the literature.

1 Introduction and motivation

Random sketches have become an important tool in building unusually efficient algorithms for approximate representation of large data sets. One of their major applications is to data streams. To put our work into perspective, we start by describing briefly the data streaming context. We then list some of the major contributions and discuss our results.

Consider an integer–valued signal $f : \{1, \ldots, N\} \to \{-M, \ldots, M\}$, defined over a "very large" domain, so that it is not feasible to store and/or process it in real time. The signal is updated or acquired sequentially in time, starting with the zero signal at time zero. Following, for example, Gilbert *et al.* (2001*b*) (see also Muthukrishnan (2003) and the seminal work of Henzinger, Raghavan and Rajagopalan (1998)), we focus on two streaming models: *(i) cash register*, and *(ii) aggregate*. In case *(i)*, data pairs $(i, a(i))$ are observed successively (in arbitrary order in i) and on each data arrival, the $i-$th component of the signal is updated incrementally (like a cash register): $f(i) := f(i) + a(i)$. In case *(ii)*, the data pairs $(i, f(i))$ are observed directly (again in arbitrary order in i). In this case, a given index i appears at most once and the corresponding $f(i)$'s are not updated incrementally multiple times. Model *(i)* is more general and more widely used. Both models, however, have found important applications in many areas such as on-line processing of large data bases, network traffic monitoring, computational geometry, etc. (see, e.g. Gilbert, Kotidis, Muthukrishnan and Strauss (2002, 2001*a*), Cormode and Muthukrishnan (2003*b*), Indyk (2003)). Much of the work on sketches was motivated by the seminal paper of Alon, Matias and Szegedy (1996). For a detailed review of methodologies and applications in this emerging area in theoretical computer science see Muthukrishnan (2003).

Random sketches are statistical summaries of the signal f, which can be updated sequentially (as the stream is observed) using little processing time, processing space and computations. Many algorithms involving random sketches have been proposed, see e.g. Muthukrishnan (2003) and the references therein. They provide as a common feature, approximations to various queries (functionals) on the signal f, within a factor of $(1 \pm \epsilon)$, with probability at least $(1-\delta)$, where $\epsilon > 0$ and $\delta > 0$ are "small" error and probability parameters chosen in advance. Typically, this is realized by algorithms consuming an amount

$$\mathcal{O}\Big(\log_2 M (\log_2 N)^{\mathcal{O}(1)} \ln(1/\delta)/\epsilon^{\mathcal{O}(1)} \Big)$$

of storage, and even smaller order of processing time per stream item f. Here M denotes the size of the range of the values $a(i)$ in the cash register model or $f(i)$ in the aggregate model. In many applications, one may be willing to sacrifice deterministic approximations at the expense of stochastic approximations, which are valid with high probability and are easy to compute.

Indyk (2000) has pioneered the use of $p-$stable distributions in random sketches (see also, Feingenbaum, Kannan, Strauss and Viswanathan (1999), Fong and Strauss (2000), Cormode (2003) and Cormode and Muthukrishnan (2003*a*)). The $p-$stable $(0 < p \leq 2)$ sketch of the signal f is defined as:

$$S_j(f) := \sum_{i=1}^{N} f(i) X_j(i), \quad j = 1, \ldots, K, \tag{1.1}$$

where the KN random variables $X_j(i)$ are independent with a $p-$stable distribution. The

stability (sum–stability, see the Appendix) property of the $p-$stable distribution implies that

$$S_j(f) \stackrel{d}{=} \Big(\sum_{i=1}^{N} |f(i)|^p\Big)^{1/p} X_1(1) = \|f\|_{\ell_p} X_1(1), \quad j = 1, \ldots, K, \tag{1.2}$$

where $\stackrel{d}{=}$ means equal in distribution. Also, by the linearity of the inner product (1.1), the sketch $S_j(f),\ j = 1, \ldots, K$ of f can be updated sequentially, in both, *cash register* and *aggregate* streaming models with $\mathcal{O}(K)$ operations per pair $(i, a(i))$ or $(i, f(i))$, where the $X_j(i)$'s are generated efficiently, on demand (see below). In his seminal work, Indyk (2000), used $p = 1$ (Cauchy distributions for the $X_j(i)$'s) and the median statistic

$$\text{median}\{|S_j(f)|,\ 1 \le j \le K\}$$

to estimate the norm $\|f\|_{\ell_1} := \sum_{i=1}^{N} |f(i)|$ of the signal. It was shown that for any $\epsilon > 0$ and $\delta > 0$, the norm $\|f\|_{\ell_1}$ is estimated within a factor of $(1 \pm \epsilon)$ with probability at least $(1 - \delta)$, provided that

$$K \ge \mathcal{O}\Big(\frac{1}{\epsilon^2} \log(1/\delta)\Big). \tag{1.3}$$

Moreover, by using the results of Nisan (1990), these estimates were shown to be realized with $\mathcal{O}(\log_2 M \log_2(N/\delta) \log(1/\delta)/\epsilon^2)$ bits of storage, needed primarily to store *truly random bits* or *seeds* for the pseudorandom generator. Roughly speaking, these seed bits are used to efficiently generate any one of the KN variables $Z_j(i)$'s on demand, when a data pair $(i, a(i))$ or $(i, f(i))$ is observed.

Exploiting the linearity of sketches and the properties of the stable distributions, Indyk (2000) also developed approximate embeddings of high–dimensional signals $f \in \ell_1^N$ in ℓ_1^m, where $m << N$. This allowed efficient approximate solutions to difficult nearest neighbor search algorithms in high dimensions, see e.g. Datar *et al.* (2004). Other authors also used stable distributions to construct efficient stochastic approximation algorithms, see e.g. Cormode and Muthukrishnan (2003*a*).

Here, we propose a novel type of random sketches, called *max–stable sketches*. Namely, consider a *non–negative* signal $f : \{1, \ldots, N\} \to [0, \infty)$. The $\alpha-$max–stable sketch of f is defined as:

$$E_j(f) := \max_{1 \le i \le N} f(i) Z_j(i), \quad j = 1, \ldots, K, \tag{1.4}$$

where the KN random variables $Z_j(i)$ are independent standard $\alpha-$Fréchet , that is,

$$\mathbb{P}\{Z_j(i) \le x\} = \Phi_\alpha(x) := \begin{cases} \exp\{-x^{-\alpha}\} & ,\ x > 0 \\ 0 & ,\ x \le 0, \end{cases} \tag{1.5}$$

and where $\alpha > 0$ is an arbitrary positive parameter.

The max–stable sketches can only be maintained in the *aggregate* streaming model, that is, when any given index value i is observed at most once in the stream. This is so because the operation "max" is not linear. Indeed, if the signal values $f(i)$ were incremented sequentially (as in the cash register model), then to be able to update the max–stable sketch of f, one would have to know the whole signal thus far, which is not feasible. Other than the *aggregate* model,

a natural streaming context for max–stable sketches is when the *cash register* is updated in a max–incremental fashion:

$$f(i) := \max\{f(i), a(i)\}.$$

In this setting, max–stable sketches can be maintained sequentially.

In the spirit of $p-$stable sketches, the max–stability of the $Z_j(i)$'s implies (see the Appendix):

$$E_j(f) \stackrel{d}{=} \Big(\sum_{i=1}^{N} f(i)^{\alpha}\Big)^{1/\alpha} Z_1(1) = \|f\|_{\ell_\alpha} Z_1(1), \quad j = 1, \ldots, K.$$

Therefore, as in Indyk (2000), for any $\epsilon > 0$ and $\delta > 0$, we can estimate the norm $\|f\|_{\ell_\alpha}$ within a factor of $(1 \pm \epsilon)$, with probability at least $(1 - \delta)$, if

$$K \geq \mathcal{O}\Big(\frac{1}{\epsilon^2} \ln(1/\delta)\Big).$$

Following the ideas in Indyk (2000), we show by using results of Nisan (1990), that this can be realized with real algorithms of space

$$\mathcal{O}\Big(\log_2(M) \log_2(N/\delta) \ln(1/\delta)/\epsilon^2\Big), \quad \text{and} \quad \mathcal{O}\Big(\log_2(M) \ln(1/\delta)/\epsilon^2\Big)$$

processing time per stream item $(i, f(i))$. Note that $\alpha > 0$ can be chosen arbitrarily large, whereas one is limited to $0 < p \leq 2$ in the $p-$stable case. Since the max–stable sketches are non–linear, being able to approximate $\|f\|_{\ell_\alpha}$, for any $\alpha > 0$, does not imply approximation of the distance $\|f - g\|_{\ell_\alpha}$ in the $\ell_\alpha-$norm of two signals based on their sketches. Therefore, our results do not contradict the findings of Saks and Sun (2002). The recent paper of Indyk and Woodruff (2005) provides algorithms for approximating $\ell_\alpha-$norms for $\alpha > 2$ which essentially match the theoretical lower bounds on the complexity in Saks and Sun (2002). The strengths of max–stable sketches lie in approximating max–linear functionals.

One of the key advantages of max–stable sketches is in handling dominance norms. Cormode and Muthukrishnan (2003*a*), consider the problem of estimating the norm of the dominant of several signals, that is, *dominance norms.* Given non–negative signals $f_r(i)$, $1 \leq i \leq N$, $r = 1, \ldots, R$, the goal is to estimate the norm $\|f^*\|_{\ell_\alpha}$, where

$$f^*(i) := \max_{1 \leq r \leq R} f_r(i), \quad 1 \leq i \leq N. \tag{1.6}$$

Such type of problems are of interest when monitoring Internet traffic, for example, where $f_r(i)$ stands for the number of packets transmitted by IP address i in its $r-$th transmission. The signal f^* then represents *worst case scenario* in terms of traffic load on the network and its norm or various other characteristics are of interest to network administrators. Other applications of dominance norms arise when studying electric grid loads and in finance (for more details, see Cormode and Muthukrishnan (2003*a*) and the references therein). A novel area of applications of max–stable sketches arises in privacy, see Ishai, Malkin, Strauss and Wright (2006).

Suppose that we only have access to the max–stable sketches $E_j(f_r),\ 1 \le j \le K$ of the signals f_r in (1.6). In view of *max–linearity*, we then have

$$E_j(f^*) = \max_{1 \le r \le R} E_j(f_r), \quad 1 \le j \le K.$$

That is, one can recover the max–stable sketch of the dominant signal f^* by taking component–wise maxima of the sketches of the signals $f_r,\ 1 \le r \le R$. Therefore, the quantity $\|f^*\|_\alpha$, which is the dominance norm of the signals $f_r,\ 1 \le r \le R$, can be readily estimated from the sketch $E_j(f^*)$ by using medians or sample moments. Moreover, this can be done with precision within a factor of $(1 \pm \epsilon)$ and with probability at least $(1 - \delta)$, provided that $K \ge \mathcal{O}(\ln(1/\delta)/\epsilon^2)$. In practice, one deals with finite precision calculations and pseudo–random number generators of bounded space. In this setting, as in the case of approximating plain norms, one can compute dominance norms by using an algorithm with processing space $\mathcal{O}(\log_2(M)\log_2(N/\delta)\ln(1/\delta)/\epsilon^2)$, and $\mathcal{O}(\log_2(M)\ln(1/\delta)/\epsilon^2)$ per item processing time. Our approach consumes less randomness, space and improves significantly on the processing time in the method of Cormode and Muthukrishnan (2003*a*) (see Theorem 2 therein).

In addition to norms and dominance norms, one can use max–stable sketches to recover large values of the signal exactly. We construct a point estimate $\widehat{f}(i_0)$ of $f(i_0)$, $i_0 \in \{1, \ldots, N\}$, based on an *ideal precision* $\alpha-$max–stable sketch of size K, such that for any $\epsilon > 0$ and $\delta > 0$,

$$\mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} \ge 1 - \delta, \quad \text{if} \quad f(i_0) > \epsilon \|f\|_{\ell_\alpha} \quad \text{and} \quad K \ge \frac{\ln(1/\delta)}{\epsilon^\alpha}. \tag{1.7}$$

Real algorithms of space

$$\mathcal{O}(\log_2(DN/\delta\epsilon)\log_2(N/\delta)\log_2(1/\delta)/\epsilon^\alpha) = \mathcal{O}((\log_2(DN/\delta\epsilon))^{\mathcal{O}(1)}\log_2(1/\delta)/\epsilon^\alpha)$$

and smaller order of per stream item processing time exist. Here $D = 1 + f(i_0)/(\sum_{i \ne i_0} f(i)^\alpha)^{1/\alpha}$ and the signal f is supposed to take integer values. Observe that one can easily maintain the largest $1/\epsilon^\alpha$ values of the signal exactly. Although, our method does not improve on the naive approach, the proposed estimator may be useful when the signal is not directly observable but its max–stable sketch is available. This is particularly useful in applications related to privacy, see the forthcoming paper of Ishai, Malkin, Strauss and Wright (2006).

Important ideas exploiting min–stability have been successfully used in the literature. Cohen (1997) assigns to the items of a positive signal independent Exponential variables with parameters equal to the signal values. The minima of independent exponentials is an exponential variable with parameter equal to the sum of the parameters of the individual components. Therefore, by keeping only the minima of such exponential variables corresponding to certain ranges of the signal values, one can estimate the sum of the signal values in these ranges. This can be done efficiently, in small space and time, by taking independent copies of such minima, see Theorem 2.3 in Cohen (1997). This approach can be viewed as a dual approach to that of the max–stable sketches. Indeed, the reciprocal of an Exponential variable is $\alpha-$Fréchet with $\alpha = 1$. We provide here a more general framework where α can be arbitrary positive parameter and therefore we can estimate not only sums but $\ell_\alpha-$norms. In fact, going a step further, we estimate efficiently *dominance norms* of several signals and show that relatively large values can be recovered exactly with high probability.

Alon, Duffield, Lund and Thorup (2005) suggest an interesting random priority sampling scheme. It assigns random priority $q_i = w_i/U_i$, to an item i which has weight $w_i > 0$, where the U_i's are independent uniformly distributed random numbers in $(0,1)$. In our scenario w_i corresponds to the signal value $f(i)$. However, instead of taking maxima of the priorities q_i over i, these authors consider the top$-k$ largest priority items. By using a statistic, relative to the $(k+1)-$st largest priority, they can estimate efficiently and with high probability the sum of the weights w_i for relatively small k's. This is an interesting approach and it differs from that of the max–stable sketches in two major aspects: *(i)* the random priorities q_i have Pareto distribution with heavy–tail exponent 1, whereas we employ Fréchet distributions to be able to use their max–stability property; *(ii)* in priority sampling, one keeps the top$-k$ values, whereas max–stable sketches keep different realizations of the maximal "priority". The second difference between max–stable sketches and the priority sampling scheme of Alon, Duffield, Lund and Thorup (2005) is crucial since the top$-k$ priorities are *dependent* random variables. This fact, we believe, makes the rigorous analysis of the variance in the priority sampling scheme rather difficult (see Conjecture 1 in the last reference). Nevertheless, priority sampling involves generating only N independent realizations, where N is the size of the signal. It is thus computationally more efficient than our method and the method of Cohen (1997) (see also Section 1.5.6 in Alon, Duffield, Lund and Thorup (2005) for a discussion).

In summary, the max–stable sketches proposed here are natural when dealing with *dominance norms* and $\ell_\alpha-$norms for arbitrarily large $\alpha > 0$. Their properties, moreover, can be established precisely and rigorously related to the nature of the signal f. Max–stable sketches complement and improve on existing techniques, and can offer a new "non–linear" dimension in stochastic approximation algorithms.

2 Approximating $\ell_\alpha-$norms and distances

We show here that max–stable sketches can be used to estimate norms and certain distances between two signals. For simplicity, we deal here with ideal precision sketches. In an extended version of the paper we show that efficient real algorithms exist by using the results of Nisan (1990).

We first focus on estimating the norm $\|f\|_{\ell_\alpha} = (\sum_i f(i)^\alpha)^{1/\alpha}$ from the $\alpha-$max–stable sketch $E_j(f),\ 1 \le j \le K$ of f (see (1.4)). Introduce the quantities

$$\ell_{\alpha,r}(f) := \Big(\frac{1}{\Gamma(1-r/\alpha)K}\sum_{j=1}^{K} E_j(f)^r\Big)^{1/r}$$

for some $0 < r < \alpha$, and

$$\ell_{\alpha,\mathrm{med}}(f) := (\ln 2)^{1/\alpha}\mathrm{median}\{E_j(f),\ 1 \le j \le K\},$$

see (4.13) and (4.14) below for motivation. By using the max–stability property of $\alpha-$Fréchet distributions, we obtain the following result.

Theorem 2.1 *Let $\epsilon > 0$ and $\delta > 0$. Then, for all $0 < r < \alpha/2$, we have*

$$\mathbb{P}\{|\ell_{\alpha,r}(f)/\|f\|_{\ell_\alpha} - 1| \le \epsilon\} \ge 1-\delta, \quad \textit{and} \quad \mathbb{P}\{|\ell_{\alpha,\mathrm{med}}(f)/\|f\|_{\ell_\alpha} - 1| \le \epsilon\} \ge 1-\delta,$$

provided that $K \geq C\log(1/\delta)/\epsilon^2$. Here the constant $C > 0$ depends only on α and r, in the case of the $\ell_{\alpha,r}(f)$ estimator.

The idea of the proof is given in the Appendix. It relies on the fact that $E_j(f)^r$ have finite variances for all $0 < r < \alpha/2$ and uses the Central Limit Theorem. More general results where $\alpha/2 \leq r < \alpha$ will be given in an extended version of the paper.

The last result indicates that the quantities $\ell_{\alpha,r}(f),\ 0 < r < \alpha/2$ and $\ell_{\alpha,\mathrm{med}}$ approximate $\|f\|_{\ell_\alpha}$ up to a factor of $(1 \pm \epsilon)$ with probability at least $(1-\delta)$. This can be achieved with an ideal–precision sketch of size $K = \mathcal{O}(\log(1/\delta)/\epsilon^2)$.

We now focus on approximating distances. Consider two signals $f,\ g : \{1,\ldots,N\} \to [0,\infty)$ and let $E_j(f),\ E_j(g),\ j = 1,\ldots,K$ be their *ideal precision* max–stable sketches. Observe that the max–stable sketches are non–linear and therefore even if $f(i) \leq g(i),\ 1 \leq i \leq N$, the sketch $E_j(g-f)$ does not equal $E_j(g) - E_j(f)$. Nevertheless, one can introduce a distance between the signals f and g, other than the norm $\|f-g\|_{\ell_\alpha}$ which can be computed by using the sketches $E_j(f)$ and $E_j(g)$.

Consider the functional

$$\rho_\alpha(f,g) := \|f^\alpha - g^\alpha\|_{\ell_1} = \sum_i |f(i)^\alpha - g(i)^\alpha|.$$

One can verify that $\rho_\alpha(f,g)$ is a metric on $\mathbb{R}_+^N$. This metric, rather than the norm $\|f-g\|_{\ell_\alpha}$, is more natural when dealing with max–stable sketches (see, Stoev and Taqqu (2005)).

Observe that

$$\begin{aligned}\|f^\alpha - g^\alpha\|_{\ell_1} &= \sum_i (f(i)^\alpha \vee g(i)^\alpha - f(i)^\alpha) + \sum_i (f(i)^\alpha \vee g(i)^\alpha - g(i)^\alpha) \\ &= 2\|f \vee g\|_{\ell_\alpha}^\alpha - \|f\|_{\ell_\alpha}^\alpha - \|g\|_{\ell_\alpha}^\alpha.\end{aligned}$$

By the max–linearity of max–stable sketches, we get $\mathbb{E}_j(f \vee g) = E_j(f) \vee E_j(g)$ (see (4.12), below). Therefore, the terms in the last expression can be estimated in terms the estimators $\ell_{\alpha,r}(f)$ and $\ell_{\alpha,\mathrm{med}}(f)$ above. Namely, we define

$$\widehat{\rho}_{\alpha,r}(f,g) := 2\ell_{\alpha,r}(f \vee g)^\alpha - \ell_{\alpha,r}(f)^\alpha - \ell_{\alpha,r}(g)^\alpha,$$

for some $0 < r < \alpha$, and

$$\widehat{\rho}_{\alpha,\mathrm{med}}(f,g) := 2\ell_{\alpha,\mathrm{med}}(f \vee g)^\alpha - \ell_{\alpha,\mathrm{med}}(f)^\alpha - \ell_{\alpha,\mathrm{med}}(g)^\alpha.$$

Theorem 2.2 *Let $\epsilon > 0$, $\delta > 0$ and $\eta > 0$. If*

$$\rho_\alpha(f,g) \geq \eta \|f \vee g\|_{\ell_\alpha}^\alpha, \tag{2.1}$$

then, for all $0 < \rho < \alpha/2$, we have

$$\mathbb{P}\Big\{|\widehat{\rho}(f,g)/\rho(f,g) - 1| \leq \mathcal{O}(\epsilon/\eta)\Big\} \geq 1 - 3\delta, \tag{2.2}$$

provided that $K \geq C\ln(1/\delta)/\epsilon^2$. Here the constant $C > 0$ depends only on α and r, in the case of the $\widehat{\rho}_{\alpha,r}(f,g)$ estimator. Here $\widehat{\rho}(f,g)$ stands for either $\widehat{\rho}_{\alpha,\mathrm{med}}(f,g)$ or $\widehat{\rho}_{\alpha,r}(f,g)$.

The idea of the proof is given in the Appendix. The condition (2.1) is essential. Indeed, by taking indicator signals $f(i) = 1_A(i)$ and $g(i) = 1_B(i)$, we get that

$$\rho_\alpha(f, g) = \sum_i 1_A(i) 1_B(i) = |A \cap B|.$$

Therefore, if condition of type (2.1) was not present, one would be able to efficiently estimate the intersection of the two sets A and B with small relative error, which is proved to be a hard problem (see, Razborov (1992) and also Section 4 in Cormode and Muthukrishnan (2003*a*)).

3 Approximating dominance norms

Let now $f_r(i),\ 1 \le i \le N,\ r = 1, \ldots, R$ be R non–negative signals defined over large domain. Our goal is to approximate their dominance $\ell_\alpha-$norm, that is, the norm $\|f^*\|_{\ell_\alpha},\ \alpha > 0$, of the signal

$$f^*(i) := \max_{1 \le r \le R} f_r(i), \quad 1 \le i \le N.$$

As argued in the introduction, such problems arise in Internet traffic monitoring, electric grid management and also in finance. The seminal paper of Cormode and Muthukrishnan (2003*a*) addresses the problem of dominance norm estimation in the special case $\alpha = 1$. It was shown therein that the problem has a small space and time approximate solution, valid with high probability. The main tool used used in the last work are $p-$(sum)stable sketches of the data where the stability index $p > 0$ is taken to have very small values. Here, we propose an alternative solution to the dominance norm problem, which is superior in terms of time and space consumption and also works when dealing with $\|\cdot\|_{\ell_\alpha}$ for an arbitrary $\alpha > 0$. In the end of this section, we also show the connection between our approach and that of Cormode and Muthukrishnan (2003*a*).

Let $E_j(f_r),\ 1 \le j \le K$ be the $\alpha-$max–stable sketches of the signals $f_r,\ r = 1, \ldots, R$. By max–linearity of the max–stable sketch:

$$E_j(f^*) = \max_{1 \le r \le R} E_j(f_r), \quad \forall j, \tag{3.1}$$

where $f^*(i) = \max_{1 \le r \le R} f_r(i)$, $1 \le i \le N$.
Hence, by using sample medians for example, we get

$$\begin{aligned} \mathrm{dom}_{\max,\alpha}(f_1, \ldots, f_R) &:= \|f^*\|_{\ell_\alpha} \approx \ell_{\alpha,\mathrm{med}}(f^*) \\ &= (\ln 2)^{1/\alpha} \mathrm{median}\{\vee_r E_j(f_r),\ 1 \le j \le K\}. \end{aligned}$$

Our first results on $\ell_\alpha-$norm approximation imply:

Theorem 3.1 *Let $\epsilon > 0$ and $\delta > 0$. For all $0 < r < \alpha/2$:*

$$\mathbb{P}\{|\ell_{\alpha,r}(f^*)/\|f^*\|_{\ell_\alpha} - 1| \le \epsilon\} \ge 1 - \delta, \quad \textit{and} \quad \mathbb{P}\{|\ell_{\alpha,\mathrm{med}}(f^*)/\|f\|_{\ell_\alpha} - 1| \le \epsilon\} \ge 1 - \delta,$$

provided

$$K \ge C \log(1/\delta)/\epsilon^2.$$

Here the constant $C > 0$ depends only on α and r, in the case of the $\ell_{\alpha,r}(f^*)$ estimator.

The proof of this result follows from the max–linearity of the max–stable sketches (see (3.1)) and Theorem 2.1 above. We now make some remarks on the differences between our approach and that of Cormode and Muthukrishnan (2003*a*).

Remarks:

- Note that the statement of Lemma 1 in the last reference is mathematically incorrect. One cannot have α in the right–hand side since limit has been taken as $\alpha \to 0+$ in the left–hand side therein. The correct statement is as follows:

  Let ξ_α be symmetric $\alpha-$stable random variables with constant scale coefficients $\sigma > 0$. Then, as $\alpha \to 0+$, we have

  $$|\xi_\alpha|^\alpha \xrightarrow{d} Z,$$

  where Z is a standard $1-$Fréchet random variable, that is, $\mathbb{P}\{Z \leq x\} = \exp\{-1/x\},\ x > 0$. Observe that $Z = 1/X$, where X is an Exponential random variable with mean 1. See Exercise 1.29, p. 54 in Samorodnitsky and Taqqu (1994).

  Therefore, the continuity of the cumulative distribution function of the limit Z implies that the medians of the distributions of $|\xi_\alpha|^\alpha$ converge, as $\alpha \to 0+$, to the median of Z which is $1/\ln(2)$ (see (4.14) below). (Note that here we use α as in Cormode and Muthukrishnan (2003*a*) whereas the parameter α plays a different role in Relation (4.14) below.)

- The method of Cormode and Muthukrishnan (2003*a*) uses $p-$(sum)stable sketches with very small $p > 0$. The $p-$stable distributions involved in these sketches have infinite moments of all orders greater than p and in practice take extremely large values. This poses a number of practical challenges in storing and in fact precisely generating these random sketches. Our method does involve heavy–tailed random variables but they are not as *extremely heavy–tailed* and have good computational properties. Furthermore, Fréchet distributions can be simulated more efficiently than sum–stable distributions (see the Appendix). Therefore, in practice we expect our method to be more robust than the one of Cormode and Muthukrishnan (2003*a*).

- The storage and per item processing times of our method are significantly less than those of Cormode and Muthukrishnan (2003*a*).

4 Answering point queries with max–stable sketches

Max–stable sketches can be also used to recover relatively large values of the signal exactly with high probability. This problem has in fact a deterministic solution by using a naive algorithm in small space and time. Namely, as the signal is being observed (in the aggregate or time series model) we simply maintain a vector of the top$-K$ largest values. Max–stable sketches however, can be very helpful if no direct access to the signal is available either due to security, computational, power or privacy restrictions (see Ishai, Malkin, Strauss and Wright (2006)).

We first present the main ideas using ideal algorithms which assume infinite precision and random variables which can be perfectly generated. We then remove these idealizations.

Consider a non–negative signal $f : \{1, \ldots, N\} \to [0, \infty)$. Let $\alpha > 0$ and let $E_j(f),\ j = 1, \ldots, K$ be an *ideal* $\alpha-$max–stable sketch of f defined in (1.4). Given an $i_0 \in \{1, \ldots, N\}$, set

$$g_j(i_0) := \frac{E_j(f)}{Z_j(i_0)} = \frac{\max_{1 \le i \le N} f(i) Z_j(i)}{Z_j(i_0)}, \quad j = 1, \ldots, K, \tag{4.1}$$

and define the *point query* estimate $\widehat{f}(i_0)$ as the smallest of the $g_j(i_0),\ j = 1, \ldots, K$. If $g_{(j)}(i_0),\ j = 1, \ldots, K$ denote the sorted $g_j(i_0)$'s:

$$g_{(1)}(i_0) \le g_{(2)}(i_0) \le \cdots \le g_{(K)}(i_0),$$

then the *point query* estimate $\widehat{f}(i_0)$ is

$$\widehat{f}(i_0) := g_{(1)}(i_0) = \min_{1 \le j \le K} g_j(i_0). \tag{4.2}$$

We also introduce the following criterion:

$$\text{criterion}(i_0) := \begin{cases} 1 & , \text{ if } g_{(1)}(i_0) = g_{(2)}(i_0) \\ 0 & , \text{ if } g_{(1)}(i_0) < g_{(2)}(i_0). \end{cases}$$

which serves as a proxy for $\widehat{f}(i_0) = f(i_0)$, as indicated in the following theorem.

Theorem 4.1 *Let $\epsilon \in (0, 1)$, $\delta > 0$ and $i_0 \in \{1, \ldots, N\}$.*
(i) *If $f(i_0) > \epsilon \|f\|_{\ell_\alpha}$ and $K \ge \ln(1/\delta)/\epsilon^\alpha$ (see (1.7)), then*

$$\mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} \ge 1 - \delta. \tag{4.3}$$

(ii) (a) criterion$(i_0) = 1$ *implies* $\widehat{f}(i_0) = f(i_0)$.
(ii) (b) *If for some $\theta > 0$,*

$$f(i_0) > \epsilon \|f\|_{\ell_\alpha} \quad \text{and} \quad K \ge \max\{3, 2C_\theta \ln(2/\delta)/\epsilon^{\alpha+\theta}\},$$

where $C_\theta = \mathcal{O}(1/\theta^{1+\theta/\alpha})$ is given in (4.22), then

$$\mathbb{P}\{\text{criterion}(i_0) = 1\} \ge 1 - \delta. \tag{4.4}$$

We now address the algorithmic implementation of the point query problem and its criterion. This is more involved now than in the case of norms and therefore we present a detailed argument here. Following Indyk (2000), suppose now that the signal is only of finite precision e.g.

$$f : \{1, \ldots, N\} \to \{0, 1, \ldots, L\}$$

and suppose, moreover, that our pseudorandom numbers can only take values in the set $V_L := \{p/q,\ p, q \in \{0, 1, \ldots, L\},\ q \ne 0\}$. Let $U_j(i)$ be infinite precision independent uniform random numbers in $(0, 1)$. We shall base our algorithms on discretized versions of the ideal standard $\alpha-$Fréchet variables $Z_j(i) := \Phi_\alpha^{-1}(U_j(i))$, where $\Phi_\alpha^{-1}(y) := (\ln(1/y))^{-1/\alpha},\ y \in (0, 1)$

is the inverse cumulative distribution function of a standard $\alpha-$Fréchet variable. Fix a small parameter $\gamma > 0$ (to be specified), and suppose that

$$U_j(i) \in [\gamma, 1-\gamma], \quad \text{for all } i = 1, \ldots, N,\ j = 1, \ldots, K. \tag{4.5}$$

This is not a limitation since this happens with probability at least $(1-2\gamma)^{KN}$ which is at least $(1-\delta)$ provided that $\gamma < \delta/(4KN) = \mathcal{O}(\delta/KN)$. This is so because $|\ln(1-2\gamma)| \leq 4\gamma$, $\forall \gamma \in (0, 1/4)$ and since $|\ln(1-\delta)| \geq \delta$, $\forall \delta \in (0,1)$.

Let now $\widetilde{U}_j(i)$ be a multiple of $1/L$, nearest to $U_j(i)$, which is also in the set $[\gamma, 1-\gamma] \cap V_L$. Let $Z_j^*(i) := \Phi_\alpha^{-1}(\widetilde{U}_j(i))$ and let $\widetilde{Z}_j(i)$ be a multiple of $1/L$ in the set V_L, nearest to $Z_j^*(i)$. Observe that $|\widetilde{Z}_j(i) - Z_j^*(i)| \leq 1/L$ and, as in Indyk (2000), by the mean value theorem,

$$|Z_j(i) - Z_j^*(i)| \leq \frac{1}{L} \sup_{y \in [\gamma, 1-\gamma]} \Big| d\Phi_\alpha^{-1}(y)/dy \Big| = \mathcal{O}\Big(\frac{1}{L\gamma^{1+1/\alpha}}\Big). \tag{4.6}$$

and therefore,

$$|\widetilde{Z}_j(i) - Z_j(i)| \leq \beta := \mathcal{O}(1/L\gamma^{1+1/\alpha}), \quad \text{for all } i = 1, \ldots, N,\ j = 1, \ldots, K. \tag{4.7}$$

Theorem 4.2 *Let $\epsilon \in (0, (\alpha/(\alpha+1))^{1/\alpha^2})$, $\delta \in (0,1)$ and $D > 0$. Suppose that*

$$2(1 + C_\alpha)\epsilon(\sum_{1 \leq i \leq N} f(i)^\alpha)^{1/\alpha} \leq f(i_0) \leq D(\sum_{i \neq i_0} f(i)^\alpha)^{1/\alpha}, \tag{4.8}$$

where $C_\alpha = \alpha(1+1/\alpha)^{1+1/\alpha} e^{-(1+1/\alpha)}$.

If the precision β in (4.7) is such that $\beta \leq \epsilon^\alpha/(D+1)$, then there exists an algorithm, implementing the point estimator $\widehat{f}(i_0)$ so that

$$\mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} \geq 1 - 3\delta, \tag{4.9}$$

holds. This can be done in space $\mathcal{O}(\log_2(DN/\epsilon\delta)\log_2(N)\ln(1/\delta)/\epsilon^\alpha)$ with the same order of bit–wise operations per stream item.

The proof is given in the Appendix.

The infinite precision was essential in proving that $\{\text{criterion}(i_0) = 1\}$ implies $\{\widehat{f}(i_0) = f(i_0)\}$. We cannot expect this when using real algorithms where the $Z_j(i)$'s have finite precision. The following result shows, however, that there is nevertheless an algorithm such that $\{\text{criterion}(i_0) = 1\}$ implies that $\{\widehat{f}(i_0) = f(i_0)\}$ holds with high probability, independently of the nature of the signal f.

Theorem 4.3 *Let the point estimator $\widehat{f}(i_0)$ and its criterion be based on a max–stable sketch in terms of the finite precision variables $\widetilde{Z}_j(i)$ as in (4.7). If $\beta \leq C(\delta/(K^2[\ln(NK^2/\delta)]^{1/\alpha}))$, then*

$$\mathbb{P}(\{\widehat{f}(i_0) \neq f(i_0)\} \cap \{\text{criterion}(i_0) = 1\}) \leq \delta, \tag{4.10}$$

where the constant C does not depend on the signal f. The last probability bound is also valid for an algorithm requiring storage $\mathcal{O}((\log_2(N)\ln(1/\delta)/\epsilon^\alpha)^{\mathcal{O}(1)})$ and the same order of operations per stream item.

The proof is given in the Appendix.

Remarks:

1. Relation (4.10) shows that our criterion may falsely indicate that $\widehat{f}(i_0) = f(i_0)$ only with small probability.

2. Our point query and its criterion algorithms have features of both Las Vegas and Monte Carlo randomized algorithms. Namely, they give *exact* results, as Las Vegas algorithms do, however their computational time is fixed and sometimes (with low probability) they fail to give correct results. As in Monte Carlo algorithms, the probability of getting exact results grows with the size of the max–stable sketch.

Acknowledgments

We would like to thank Anna Gilbert, Martin Strauss and Joel Tropp for their remarks and encouragement. Martin Strauss helped us understand better the results of Nisan (1990).

Appendix

Background on max–stable distributions

Definition 4.1 A random variable Z is said to be max–stable if, for any $a,\ b > 0$, there exist $c > 0$ and $d \in \mathbb{R}$, such that

$$\max\{aZ', bZ''\} \stackrel{d}{=} cZ + d, \tag{4.11}$$

where Z' and Z'' are independent copies of Z.

This definition resembles the definition of sum–stability where the operation "max" is the summation. Recall that X is sum–stable if for any $a,\ b > 0$, there exist $c > 0$ and $d \in \mathbb{R}$, such that

$$aX' + bX'' \stackrel{d}{=} cX + d.$$

Both sum–stable and max–stable distributions arise as the limit distributions when taking sums and maxima, respectively, of independent and identically distributed (iid) random variables. For more details on sum–stable and max–stable variables, see e.g. Samorodnitsky and Taqqu (1994) and Resnick (1987). We will only review in more detail the class of $\alpha-$Fréchet max–stable variables.

Definition 4.2 A random variable Z is said to be $\alpha-$Fréchet, for some $\alpha > 0$, if

$$\mathbb{P}\{Z \leq x\} = \exp\{-\sigma^\alpha x^{-\alpha}\}, \quad \text{for } x > 0,$$

and zero otherwise (for $x \leq 0$), where $\sigma > 0$. If $\sigma = 1$, then Z is said to be *standard* $\alpha-$Fréchet.

We now list some key features of the $\alpha-$Fréchet variables

• The parameter σ plays the role of a scale coefficient. Indeed, for all $a > 0$,

$$\mathbb{P}\{aZ \le x\} = \mathbb{P}\{Z \le x/a\} = \exp\{-(a\sigma)^\alpha x^{-\alpha}\}, \quad x > 0,$$

and therefore aZ is $\alpha-$Fréchet with scale coefficient $a\sigma$.

• One can check by using independence that (4.11) holds for any $\alpha-$Fréchet Z. More generally, let $Z, Z(1), \ldots, Z(n)$ be iid $\alpha-$Fréchet with scale coefficients $\sigma > 0$, and let $f(i) \ge 0$. Then, by independence, for any $x > 0$,

$$\mathbb{P}\{\vee_{1\le i\le n} f(i)Z(i) \le x\} = \prod_{1\le i\le n} \mathbb{P}\{Z_i \le x/f(i)\} = \exp\{-\sum_{i=1}^n f(i)^\alpha \sigma^\alpha x^{-\alpha}\},$$

and thus

$$\xi := \bigvee_{1\le i\le n} f(i)Z(i) \stackrel{d}{=} \sigma_\xi Z, \quad \text{where } \sigma_\xi = (\sum_i f(i)^\alpha)^{1/\alpha} = \|f\|_{\ell_\alpha},$$

and where Z is a standard $\alpha-$Fréchet variable. That is, the weighted maxima ξ is an $\alpha-$Fréchet variable with scale coefficient σ_ξ equal to $\|f\|_{\ell_\alpha}$.

This last property is one motivation to consider max–stable sketches. The max–stable sketch defined in (1.4) can be viewed as a collection of independent realizations of an $\alpha-$Fréchet variable with scale coefficient equal to $\|f\|_{\ell_\alpha}$.

• The max–stable sketches are *max–linear*. That is, if $f,\ g \in \mathbb{R}_+^N$ are two signals, then for any $a,\ b \ge 0$, we have:

$$E_j(af \vee bg) = aE_j(f) \vee bE_j(g), \quad \text{for all } j = 1, \ldots, K. \tag{4.12}$$

Indeed,

$$E_j(af \vee bg) = \bigvee_{1\le i\le N} (af(i) \vee bg(i))Z_j(i) = a \bigvee_{1\le i\le N} f(i)Z_j(i) \vee b \bigvee_{1\le i\le N} g(i)Z_j(i) = aE_j(f) \vee bE_j(g).$$

• The $\alpha-$Fréchet variables are heavy–tailed. Namely, by using the Taylor series expansion of $1 - e^{-z}$, one can show that

$$\mathbb{P}\{Z > x\} \sim \sigma^\alpha x^{-\alpha}, \quad \text{as } x \to \infty,$$

where $a_n \sim b_n$ means $a_n/b_n \to 1,\ n \to \infty$. Thus, the moments $\mathbb{E}Z^p,\ p > 0$ of Z are finite only if $0 < p < \alpha$. However, when $0 < p < \alpha$, these moments can be easily evaluated. We have

$$\mathbb{E}Z^p = \int_0^\infty z^p d\exp\{-\sigma^\alpha z^{-\alpha}\} = \sigma^p \int_0^\infty u^{-p/\alpha} e^{-u} du = \sigma^p \Gamma(1 - p/\alpha), \tag{4.13}$$

where in the last integral we used the change of variables $u = \sigma^\alpha x^{-\alpha}$ and where $\Gamma(a) := \int_0^\infty u^{a-1} e^{-u} du,\ a > 0$ denotes the Gamma function.

• One can also easily express the median $\mathrm{med}(Z)$ of an $\alpha-$Fréchet variable Z. Indeed, $\mathbb{P}\{Z \leq \mathrm{med}(Z)\} = 1/2$ and by solving $\exp\{-\sigma^\alpha \mathrm{med}(Z)^{-\alpha}\} = 1/2$, one obtains:

$$\mathrm{med}(Z) = \frac{\sigma}{(\ln 2)^{1/\alpha}}. \tag{4.14}$$

In Section 2, we used Relations (4.13) and (4.14) to estimate norms and distances of signals based on their max–stable sketches.

• The $\alpha-$Fréchet variables can be easily simulated. If $U_j,\ j \in \mathbb{N}$ are independent uniformly distributed variables in $(0,1)$, then $Z_j := \Phi_\alpha^{-1}(U_j) = (\ln(1/U_j))^{-1/\alpha},\ j \in \mathbb{N}$ are independent standard $\alpha-$Fréchet. Indeed, for all $x > 0$,

$$\mathbb{P}\{(\ln(1/U))^{-1/\alpha} \leq x\} = \mathbb{P}\{\ln(1/U) \geq x^{-\alpha}\} = \mathbb{P}\{U \leq e^{-x^{-\alpha}}\} = e^{-x^{-\alpha}}.$$

Proofs for Sections 2 and 3

Proof of Theorem 2.1: Observe that

$$\frac{\ell_{\alpha,r}(f)^r}{\|f\|_{\ell_\alpha}^r} = \frac{1}{\Gamma(1-r/\alpha)K} \sum_{j=1}^{K} \xi_j^r,$$

and

$$\frac{\ell_{\alpha,\mathrm{med}}(f)}{\|f\|_{\ell_\alpha}} = (\ln(2))^{1/\alpha} \mathrm{median}\{\xi_j,\ 1 \leq j \leq K\},$$

where $\xi_j,\ j = 1, \ldots, K$ are independent standard $\alpha-$Fréchet variables.

Therefore, the result in the case of the sample–median based estimator follows from Lemma 2 in Indyk (2000), for example, since the derivative of $\Phi_\alpha^{-1}(y) = (\ln(1/y))^{-1/\alpha}$ at $y = 1/2$ is bounded. The result in the case of the moment estimator follows from the Central Limit Theorem, since the variables $\xi_j^r - 1/\Gamma(1-r/\alpha)$ have zero expectations and finite variances. □

We will provide more detailed bounds in the above proof with absolute constants in an extended version of this paper.

Proof of Theorem 2.2: We consider only $\widehat{\rho}(f,g) = \widehat{\rho}_{\alpha,r}(f,g)$. The argument for the estimator $\widehat{\rho}_{\alpha,\mathrm{med}}(f,g)$ is similar. Suppose that K is as in Theorem 2.1, so that with probabilities at least $(1-\delta)$, we have $\ell_{\alpha,r}(f)^\alpha = \|f\|_{\ell_\alpha}^\alpha(1+\mathcal{O}(\epsilon))$, $\ell_{\alpha,r}(g)^\alpha = \|g\|_{\ell_\alpha}^\alpha(1+\mathcal{O}(\epsilon))$, and $\ell_{\alpha,r}(f \vee g)^\alpha = \|f \vee g\|_{\ell_\alpha}^\alpha(1+\mathcal{O}(\epsilon))$.

Thus, by the union bound of probabilities, with probability at least $(1-3\delta)$ we have

$$\widehat{\rho}_{\alpha,r}(f,g) = 2\|f \vee g\|_{\ell_\alpha}^\alpha(1+\mathcal{O}(\epsilon)) - \|f\|_{\ell_\alpha}^\alpha(1+\mathcal{O}(\epsilon)) - \|g\|_{\ell_\alpha}^\alpha(1+\mathcal{O}(\epsilon)). \tag{4.15}$$

Now, since

$$\rho_\alpha(f,g) = 2\|f \vee g\|_{\ell_\alpha}^\alpha - \|f\|_{\ell_\alpha}^\alpha - \|g\|_{\ell_\alpha}^\alpha \geq \eta\|f \vee g\|_{\ell_\alpha}^\alpha \geq \eta \max\{\|f\|_{\ell_\alpha}^\alpha, \|g\|_{\ell_\alpha}^\alpha\},$$

we get, from (4.15), that

$$\widehat{\rho}_{\alpha,r}(f,g) = (2\|f \vee g\|_{\ell_\alpha}^\alpha - \|f\|_{\ell_\alpha}^\alpha - \|g\|_{\ell_\alpha}^\alpha)(1+\mathcal{O}(\epsilon/\eta)).$$

The last relation is valid with probability at least $(1-3\delta)$, which implies (2.2). □

Proofs for Section 4

Proof of Theorem 4.1: Observe that by (1.4), (4.1) and (4.2),

$$\widehat{f}(i_0) = \min_{1\le j\le K} E_j(f)/Z_j(i_0) = \min_{1\le j\le K} f(i_0) \vee Z_j(i_0)^{-1} \bigvee_{i\ne i_0} f(i)Z_j(i),$$

where $\vee$ denotes "max". Now $\vee_{i\ne i_0} f(i)Z_j(i)$ is independent of $Z_j(i_0) \stackrel{d}{=} Z_j(1)$ and, by max–stability (see Appendix), it equals in distribution $(\sum_{i\ne i_0} f(i)^\alpha)^{1/\alpha} Z_j(2)$. Hence

$$\widehat{f}(i_0) \stackrel{d}{=} f(i_0) \vee \min_{1\le j\le K} \Big\{ Z_j(1)^{-1} (\sum_{i\ne i_0} f(i)^\alpha)^{1/\alpha} Z_j(2) \Big\} =: f(i_0) \vee \min_{1\le j\le K} c_f(i_0) Z_j(2)/Z_j(1), \tag{4.16}$$

where $c_f(i_0) = (\sum_{i\ne i_0} f(i)^\alpha)^{1/\alpha}$. By using again the independence in j, we get

$$\begin{aligned} \mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} &= \mathbb{P}\{f(i_0) \ge \min_{1\le j\le K} c_f(i_0) Z_j(2)/Z_j(1)\} = 1 - \mathbb{P}\{f(i_0) < c_f(i_0) Z_1(2)/Z_1(1)\}^K \\ &= 1 - \mathbb{P}\{Z_1(1)/Z_1(2) < c_f(i_0)/f(i_0)\}^K. \end{aligned} \tag{4.17}$$

By Lemma 4.1, the probability in (4.17) equals $1/(1 + f(i_0)^\alpha/c_f(i_0)^\alpha)$, and hence

$$\mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} = 1 - \Big(\frac{f(i_0)^\alpha}{f(i_0)^\alpha + c_f(i_0)^\alpha}\Big)^K = 1 - \Big(\frac{\sum_{i\ne i_0} f(i)^\alpha}{\sum_i f(i)^\alpha}\Big)^K = 1 - \Big(1 - \frac{f(i_0)^\alpha}{\|f\|_{\ell_\alpha}^\alpha}\Big)^K. \tag{4.18}$$

Now $f(i_0) > \epsilon \|f\|_{\ell_\alpha}$ implies $\mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} \ge 1 - (1-\epsilon^\alpha)^K$. By choosing $\delta \ge (1-\epsilon^\alpha)^K$, we get $K \ge \ln(\delta)/\ln(1-\epsilon^\alpha)$. Since $|\ln(1-x)| \ge x$, for all $x \in (0,1)$, we get that $K \ge \ln(1/\delta)/\epsilon^\alpha \ge \ln(\delta)/\ln(1-\epsilon^\alpha)$, for all $\epsilon \in (0,1)$, which completes the proof of part **(i)**.

We now prove part **(ii) (a)**. Observe that

$$(g_{(1)}(i_0), g_{(2)}(i_0)) \stackrel{d}{=} (f(i_0) \vee \xi_{(1)}(i_0)), f(i_0) \vee \xi_{(2)}(i_0)), \tag{4.19}$$

where $\xi_{(j)}(i_0) \le \xi_{(j+1)}(i_0),\ j \le K-1$ is the sorted sample of independent random variables $\xi_j(i_0) := c_f(i_0) Z_j(2)/Z_j(1),\ j = 1, \ldots, K$. Since the joint distribution of $\xi_{(1)}(i_0)$ and $\xi_{(2)}(i_0)$ has a density, it follows that $\mathbb{P}\{\xi_{(1)}(i_0) = \xi_{(2)}(i_0)\} = 0$. Hence, in view of (4.19), with probability 1, we have

$$\{\text{criterion}(i_0) = 1\} \equiv \{g_{(1)}(i_0) = g_{(2)}(i_0)\} = \{f(i_0) \ge \xi_{(2)}(i_0)\}. \tag{4.20}$$

The right–hand side of (4.20) occurs only if $\{\widehat{f}(i_0) = f(i_0)\}$, which completes the proof of **(ii) (a)**.

We now turn to part **(ii) (b)** and estimate $\mathbb{P}\{\text{criterion}(i_0) = 1\} = \mathbb{P}\{f(i_0) \ge \xi_{(2)}(i_0)\}$. We have,

$$\begin{aligned} \mathbb{P}\{f(i_0) \ge \xi_{(2)}(i_0)\} &= \mathbb{P}\{f(i_0) \ge c_f(i_0) Z_j(2)/Z_j(1), \ \text{ for at least two } j\text{'s}\} \\ &= 1 - \binom{K}{0} p^K - \binom{K}{1} p^{K-1}(1-p) \ge 1 - (K+1)p^{K-1}, \end{aligned}$$

where $p = \mathbb{P}\{f(i_0)Z_1(1) < c_f(i_0)Z_1(2)\}$. Reasoning as in part **(i)**, we get by Lemma 4.1, $p = (1 - f(i_0)^\alpha/\|f\|_{\ell_\alpha}^\alpha) < 1 - \epsilon^\alpha$, since $f(i_0) > \epsilon\|f\|_{\ell_\alpha}$. We thus need to choose K's which satisfy the inequality $\delta \geq (K+1)(1-\epsilon^\alpha)^{K-1} \geq (K+1)p^{K-1}$. For $K \geq 3$, we have $\widetilde{K} := K - 1 \geq (K+1)/2$, and hence it suffices to have $\delta \geq 2\widetilde{K}(1-\epsilon^\alpha)^{\widetilde{K}}$, or simply,

$$\widetilde{K} \geq \frac{\ln(\widetilde{K})}{\epsilon^\alpha} + \frac{\ln(2/\delta)}{\epsilon^\alpha}, \tag{4.21}$$

where we used that $|\ln(1-\epsilon^\alpha)| \geq \epsilon^\alpha,\ \epsilon \in (0,1)$. Let $\theta > 0$ and $\widetilde{K} \geq 2\ln(2/\delta)C_\theta/\epsilon^{\alpha+\theta}$, for some $C_\theta \geq 1$ (to be specified). Then, since $\widetilde{K} \geq 2\ln(2/\delta)C_\theta/\epsilon^{\alpha+\theta} \geq 2\ln(2/\delta)/\epsilon^\alpha$, it follows that (4.21) holds if $\widetilde{K} \geq 2\ln(\widetilde{K})/\epsilon^\alpha$. Since $\widetilde{K}^{\alpha/(\alpha+\theta)} \geq (2C_\theta)^{\alpha/(\alpha+\theta)}/\epsilon^\alpha$, we get that (4.21) holds if

$$\widetilde{K} \geq \frac{\widetilde{K}^{\alpha/(\alpha+\theta)}}{(2C_\theta)^{\alpha/(\alpha+\theta)}} 2\ln(\widetilde{K}) \geq \frac{2\ln(\widetilde{K})}{\epsilon^\alpha}, \quad \text{or if,} \quad (2C_\theta)^{\alpha/(\alpha+\theta)}\widetilde{K}^{\theta/(\alpha+\theta)} \geq 2\ln(\widetilde{K}).$$

The last is equivalent to

$$u^\gamma \geq \ln(u), \quad \text{where } u = \widetilde{K}^{2^{\theta/(\alpha+\theta)}C_\theta^{-\alpha/(\alpha+\theta)}},$$

and $\gamma = (\theta/(\alpha+\theta))2^{-\theta/(\alpha+\theta)}C_\theta^{\alpha/(\alpha+\theta)}$. We have that $u^\gamma \geq \ln(u),\ u > 1$, for all $\gamma \geq 1/e$, and thus for C_θ we obtain:

$$\gamma \equiv \frac{\theta}{\alpha+\theta}2^{-\theta/(\alpha+\theta)}C_\theta^{\alpha/(\alpha+\theta)} \geq 1/e \quad \text{or} \quad C_\theta \geq \frac{2^{\theta/\alpha}}{e^{1+\theta/\alpha}}\Big(1+\frac{\alpha}{\theta}\Big)^{1+\theta/\alpha} + 1, \tag{4.22}$$

where we add 1 in the last relation to ensure that $C_\theta \geq 1$. □

Proof of Theorem 4.2: We will first show that

$$\mathbb{P}\{\widetilde{f}(i_0) = f(i_0)\} \geq 1 - 2\delta, \tag{4.23}$$

where $\widetilde{f}(i_0)$ is defined as $\widehat{f}(i_0)$ but based on *truly independent* variables $\widetilde{Z}_j(i)$ which satisfy (4.7). Recall that Relation (4.7) holds if (4.5) holds. Choose $\gamma > 0$ and L so that $\mathbb{P}(B) \geq 1-\delta$, where B denotes the event {Condition (4.5) holds}. As in the proof of Theorem 4.1,

$$\{\widetilde{f}(i_0) = f(i_0)\} = \{f(i_0) \geq \min_{1\leq j\leq K} \widetilde{Z}_j(i_0)^{-1} \vee_{i\neq i_0} f(i)\widetilde{Z}_j(i)\}.$$

Observe that, since B holds, by (4.7), $\vee_{i\neq i_0} f(i)\widetilde{Z}_j(i) \geq \vee_{i\neq i_0} f(i)Z_j(i) + \beta s_f(i_0)$, where $s_f(i_0) := \max_{i\neq i_0} f(i)$. Thus, by using also that $\widetilde{Z}_j(i_0) \geq Z_j(i_0) - \beta$, we get

$$\{\widetilde{f}(i_0) = f(i_0)\} \supset B \cap \Big\{f(i_0) \geq \min_{1\leq j\leq K}(Z_j(i_0)-\beta)^{-1}\Big(\vee_{i\neq i_0} f(i)Z_j(i) + \beta s_f(i_0)\Big)\Big\} =: B \cap C. \tag{4.24}$$

Since $\mathbb{P}(B) \geq 1-\delta$ by Relation (4.24) it follows that, to establish (4.23), it is enough to show that $\mathbb{P}(C) \geq 1-\delta$, where C denotes the second event in the right–hand side of (4.24). Indeed, $\mathbb{P}\{\widetilde{f}(i_0) = f(i_0)\} \geq \mathbb{P}(B\cap C) \geq 1 - \mathbb{P}(B') - \mathbb{P}(C') \geq 1 - 2\delta$. The event C, however, involves

ideal precision Fréchet random variables where their corresponding $U_j(i)$'s are not bound to satisfy Condition (4.5). We can therefore manipulate C as in the proof of Theorem 4.1 **(i)**.

Since $\vee_{i\neq i_0} f(i)Z_j(i) \stackrel{d}{=} c_f(i_0)Z_1(2)$, where $c_f(i_0) = (\sum_{i\neq i_0} f(i)^\alpha)^{1/\alpha}$, by independence:

$$\begin{aligned}\mathbb{P}(C) &\geq 1 - \mathbb{P}\{(Z_1(1)-\beta)f(i_0) < c_f(i_0)Z_1(2) + \beta s_f(i_0)\}^K \\ &= 1 - \Big(\mathbb{E}\exp\{-\Big(\widetilde{c}_f(i_0)Z_1(2) + \beta(1+\widetilde{s}_f(i_0))\Big)^{-\alpha}\}\Big)^K,\end{aligned}$$

where $\widetilde{c}_f(i_0) = c_f(i_0)/f(i_0)$ and $\widetilde{s}_f(i_0) = s_f(i_0)/f(i_0)$. We now bound above the expectation in the last relation. Note that $\widetilde{c}_f(i_0)z + \beta(1+\widetilde{s}_f(i_0)) \leq 2\widetilde{c}_f(i_0)z$, for all $z > \beta(1+\widetilde{s}_f(i_0))/\widetilde{c}_f(i_0)$. Thus, by (1.5) and as in Relation (4.30) below,

$$\begin{aligned}1-\mathbb{P}(C) &\leq \Big(\mathbb{P}\{Z_1(2) \leq \beta(1+\widetilde{s}_f(i_0))/\widetilde{c}_f(i_0)\} + \mathbb{E}\exp\{-(2\widetilde{c}_f(i_0)Z_1(2))^{-\alpha}\}\Big)^K \\ &= \Big(p(\beta) + \frac{1}{(2\widetilde{c}_f)^{-\alpha}+1}\Big)^K, \qquad (4.25)\end{aligned}$$

where $p(\beta) = \mathbb{P}\{Z_1(2) \leq \beta(1+\widetilde{s}_f(i_0))/\widetilde{c}_f(i_0)\} = \exp\{-(\beta(1+\widetilde{s}_f(i_0))/\widetilde{c}_f(i_0))^{-\alpha}\}$.

Let now K be such that $(1-\epsilon^\alpha)^K \leq \delta$, that is, $K = \mathcal{O}(\ln(1/\delta)/\epsilon^\alpha)$. Then, in view of (4.25), $\mathbb{P}(C) \geq 1-\delta$, provided that

$$p(\beta) + \frac{1}{(2\widetilde{c}_f(i_0))^{-\alpha}+1} \leq 1-\epsilon^\alpha,$$

or, equivalently, if $f(i_0)^\alpha/(f(i_0)^\alpha + 2^\alpha\sum_{i\neq i_0} f(i_0)^\alpha) \geq (\epsilon^\alpha + p(\beta))$. The last inequality holds if

$$f(i_0) \geq 2(\epsilon^\alpha + p(\beta))^{1/\alpha}\|f\|_{\ell_\alpha}.$$

Thus, to prove (4.23), it remains to show that $(\epsilon^\alpha + p(\beta))^{1/\alpha} \leq (1+C_\alpha)\epsilon$, if (4.8) holds and $\beta \leq \epsilon^\alpha/(D+1)$, where β is the "precision" parameter in (4.6). We have that $(1+\widetilde{s}_f(i_0))/\widetilde{c}_f(i_0) \leq 1 + f(i_0)/c_f(i_0) \leq 1+D$, and thus

$$p(\beta) \leq \Phi_\alpha(\beta(D+1)) \leq \Phi_\alpha(\epsilon^\alpha),$$

since $\beta(D+1) \leq \epsilon^\alpha$. One can show that $\Phi''_\alpha(x) \geq 0$, $\forall x \in (0,(\alpha/(\alpha+1))^{1/\alpha})$, and hence $\Phi_\alpha(x) \leq \Phi'_\alpha(1+1/\alpha)x = C_\alpha x$, $\forall x \in (0,(\alpha/(\alpha+1))^{1/\alpha})$. Hence, $p(\beta) \leq C_\alpha\epsilon^\alpha$, for all $\epsilon \in (0,(\alpha/(\alpha+1))^{1/\alpha^2})$, which implies (4.23).

Now, consider a point estimator $\widehat{f}(i_0)$, defined as $\widetilde{f}(i_0)$, but with $\widetilde{Z}_j(i)$ replaced by pseudo–random variables, with the same precision (i.e. taking values in the set $V_L = \{p/q,\ p,q = 0,1,\ldots,L\}$). We will argue that a pseudo–random number generator exists so that

$$\mathbb{P}\{\widetilde{f}(i_0) = \widehat{f}(i_0)\} \geq 1-\delta, \qquad (4.26)$$

for some $\widetilde{f}(i_0)$ based on independent $\widetilde{Z}_j(i)$'s. This, in view of (4.23), would imply (4.9).

We first need (4.5) to hold with probability at least $(1-\delta)$ with $\gamma > 0$ such that $\beta = \mathcal{O}(1/L\gamma^{1+1/\alpha}) \leq \epsilon^\alpha/D$. This can be achieved by taking $L = \mathcal{O}((DNK/\delta\epsilon)^{\mathcal{O}(1)})$. Now, to ensure that (4.23) holds, it suffices to take $K = \mathcal{O}(\ln(1/\delta)/\epsilon^\alpha)$. Therefore, one needs $\log_2(L) = \mathcal{O}(\ln(DN/\delta\epsilon))$ bits to represent each $\widetilde{Z}_j(i)$.

Hence our algorithm uses $\mathcal{O}(\log_2(L)N\ln(1/\delta)/\epsilon^\alpha)$ random bits. As in Indyk (2000), by using the results of Nisan (1990), for each j, $1 \le j \le K$, one can generate pseudo–random $\overline{U_j}(i)$'s, which are "very close" to some independent $\widetilde{U}_j(i)$'s by using only $\mathcal{O}(\log_2(L)\log_2(N/\delta)\ln(1/\delta)/\epsilon^\alpha)$ truly random seeds. These $\overline{U}_j(i)$'s would "fool" our algorithm with probability at least $(1-\delta)$ since it uses only $\mathcal{O}(\log_2(L)\ln(1/\delta)/\epsilon^\alpha)$ space for computations with random bits. That is, one has (4.26). In summary, we need to store $\mathcal{O}(K\log_2(L)\log_2(N/\delta)) = \mathcal{O}(\log_2(DN/\delta\epsilon)\log_2(N/\delta)\log_2(1/\delta)/\epsilon^\alpha)$ bits, needed primarily for the truly random seeds, and to perform about $\mathcal{O}(K\log_2(L)) = \mathcal{O}(\log_2(DN/\delta\epsilon)\log_2(1/\delta)/\epsilon^\alpha)$ of bit–wise operations per each stream item, in order to maintain the sketch. $\square$

Proof of Theorem 4.3: Let, as in (4.1), $\widetilde{g}_j(i_0) := \vee_i f(i)\widetilde{Z}_j(i)/\widetilde{Z}_j(i_0) =: f(i_0)\vee\widetilde{\xi}_j(i_0)$, where $\widetilde{\xi}_j(i_0) := \vee_{i\neq i_0} f(i)\widetilde{Z}_j(i)/\widetilde{Z}_j(i_0)$, $j = 1,\ldots,K$. Since $\{\text{criterion}(i_0) = 1\} = \{f(i_0)\vee\widetilde{\xi}_{(2)}(i_0) = f(i_0)\vee\widetilde{\xi}_{(1)}(i_0)\}$, it follows that

$$\{\widehat{f}(i_0) \neq f(i_0)\} \cap \{\text{criterion}(i_0) = 1\} \subset \{\widetilde{\xi}_{(2)}(i_0) = \widetilde{\xi}_{(1)}(i_0)\},$$

where $\widetilde{\xi}_{(1)}(i_0) \le \widetilde{\xi}_{(2)}(i_0) \le \cdots$ is the ordered sample of $\widetilde{\xi}_j(i_0)$, $j = 1,\ldots,K$.

Therefore, the probability in (4.10) is bounded above by:

$$\begin{aligned}\mathbb{P}\{\widetilde{\xi}_{(2)}(i_0) = \widetilde{\xi}_{(1)}(i_0)\} &\le \mathbb{P}\{\widetilde{\xi}_{j_1} = \widetilde{\xi}_{j_2}, \text{ for some } j_1 \neq j_2,\ j_1,\ j_2 = 1,\ldots,K\} \\ &\le \binom{K}{2}\mathbb{P}\{\widetilde{\xi}_1(i_0) = \widetilde{\xi}_2(i_0)\}. \end{aligned} \tag{4.27}$$

We now focus on bounding the last probability. Since the $\widetilde{\xi}_j(i_0)$'s are independent and discrete random variables, we have

$$\mathbb{P}\{\widetilde{\xi}_1(i_0) = \widetilde{\xi}_2(i_0)\} = \sum_{x:\ x \text{ is an atom}} \mathbb{P}\{\widetilde{\xi}_1(i_0) = x\}^2. \tag{4.28}$$

Let $\eta > 0$ and observe that $(z-\beta) \ge (1-\eta)z$ and $(z+\beta) \le (1+\eta)z$, for all $z \ge \beta/\eta$. Thus, in view of (4.7),

$$\widetilde{\xi}_1(i_0) \le \vee_{i\neq i_0} f(i)\frac{(Z_1(i)+\beta)}{(Z_1(i_0)-\beta)} \le \vee_{i\neq i_0} f(i)\frac{(1+\eta)}{(1-\eta)}\frac{Z_1(i)}{Z_1(i_0)}, \quad \text{if } Z_1(i) \ge \beta/\eta,\ \forall i.$$

Since the $Z_1(i)$'s are ideal precision, independent and $\alpha-$Fréchet, $\vee_{i\neq i_0} f(i)\frac{(1+\eta)}{(1-\eta)}Z_1(i)/Z_1(i_0) \stackrel{d}{=} c_f(i_0)\frac{(1+\eta)}{(1-\eta)}Z_1(2)/Z_1(1)$, where $c_f(i_0) = (\sum_{i\neq i_0} f(i)^\alpha)^{1/\alpha}$. Therefore,

$$\widetilde{\xi}_1(i_0) \stackrel{d}{\le} c_f(i_0)\frac{(1+\eta)}{(1-\eta)}\frac{Z_1(2)}{Z_1(1)} =: \xi^*, \quad \text{if } Z_1(i) \ge \beta/\eta,\ \forall i,$$

where $\stackrel{d}{\le}$ denotes dominance in distribution. Similarly $\widetilde{\xi}_1(i_0) \stackrel{d}{\ge} \xi_*$, where $\xi_* := c_f(i_0)\frac{(1-\eta)}{(1+\eta)}Z_1(2)/Z_1(1)$. Thus,

$$\begin{aligned}\mathbb{P}\{\widetilde{\xi}_1(i_0) = x\} &\le \mathbb{P}(\{\xi_1(i_0) = x\} \cap \{Z_1(i) \ge \beta/\eta,\ \forall i\}) + \mathbb{P}\{Z_1(i) \not\ge \beta/\eta, \text{ for some } i\} \\ &\le F_{\xi_*}(x) - F_{\xi^*}(x) + (1 - \mathbb{P}\{Z_1(1) > \beta/\eta\}^N), \end{aligned}$$

where the last inequality follows from the fact that ξ_* and ξ^* have continuous cumulative distribution functions $F_{\xi_*}(x) := \mathbb{P}\{\xi_* \leq x\}$.

Thus, for the probability in (4.28), we get:

$$\mathbb{P}\{\widetilde{\xi}_1(i_0) = \widetilde{\xi}_2(i_0)\} \leq \sup_{x>0}(F_{\xi_*}(x) - F_{\xi^*}(x)) + (1 - (1 - \exp\{-(\beta/\eta)^{-\alpha}\})^N). \tag{4.29}$$

By taking $\beta/\eta \leq (1/\ln(NK^2/\delta))^{1/\alpha}$, we can make the second term in the right–hand side of (4.29) smaller than δ/K^2 (Lemma 4.2). Indeed, the second term is a monotone increasing function of (β/η). Hence by setting $\epsilon := (1/\ln(NK^2/\delta))^{1/\alpha}$ the upper bound in (4.31) becomes $N\exp\{-(1/[\ln(NK^2/\delta)]^{1/\alpha})^{-\alpha}\} = N\exp\{-\ln(NK^2/\delta)\} = \delta/K^2$. Also, by Lemma 4.3, the first term in the right–hand side of (4.29), is bounded above by $\alpha 2^{4\alpha+3}\eta$, for all $\eta \in (0, 1/2)$. Thus, in view of (4.27), the probability in (4.10) is bounded above by $\mathcal{O}(K^2\eta) + \delta/2$, which can be made smaller than δ by taking $\eta = \mathcal{O}(\delta/K^2)$ and $\beta/\eta = \mathcal{O}((1/\ln(NK^2/\delta))^{1/\alpha})$. This implies that $\beta = \mathcal{O}(\delta/(K^2[\ln(NK^2/\delta)]^{1/\alpha}))$ would ensure that (4.10) holds. Observe that the constant in the last $\mathcal{O}-$bound does not depend on the signal f. $\square$

Auxiliary lemmas

Lemma 4.1 *Let ξ and η be independent, standard $\alpha-$Fréchet variables. Then, for all $x > 0$,*

$$\mathbb{P}\{\xi/\eta \leq x\} = \frac{1}{x^{-\alpha} + 1}.$$

Proof: By independence, and in view of (1.5), we have:

$$\mathbb{P}\{\xi/\eta \leq x\} = \mathbb{E}\exp\{-(\eta x)^{-\alpha}\} = \int_0^\infty e^{-y^{-\alpha}x^{-\alpha}} de^{-y^{-\alpha}} = \int_0^1 u^{x^{-\alpha}} du = \frac{1}{x^{-\alpha} + 1}. \tag{4.30}$$

Here, we used the change of variables $u := e^{-y^{-\alpha}}$. $\square$

Lemma 4.2 *For all $\epsilon \in (0, 1)$ and $N \geq 1$, we have*

$$1 - (1 - \exp\{-(\epsilon)^{-\alpha}\})^N \leq Ne^{-\epsilon^{-\alpha}}. \tag{4.31}$$

Proof: We have that $(1 - x)^N \geq 1 - Nx$, for all $x \in (0, 1)$. Thus, for all $x \in (0, 1)$, $1 - (1 - x)^N \leq 1 - (1 - Nx) = Nx$ and by setting $x := e^{-\epsilon^{-\alpha}}$, we obtain (4.31). $\square$

Lemma 4.3 *Let $\xi^* = c\frac{(1+\eta)}{(1-\eta)}Z'/Z''$ and $\xi_* = c\frac{(1-\eta)}{(1+\eta)}Z'/Z''$, for some $c > 0$ and $\eta \in (0, 1/2)$, where Z' and Z'' are independent standard $\alpha-$Fréchet variables. Then, we have*

$$\sup_{x>0} |F_{\xi_*}(x) - F_{\xi^*}(x)| \leq \alpha 2^{4\alpha+3}\eta, \quad \forall \eta \in (0, 1/2).$$

where $F_{\xi_*}(x) := \mathbb{P}\{\xi_* \leq x\}$ and $F_{\xi^*}(x) := \mathbb{P}\{\xi^* \leq x\}$.

Proof: We have that

$$\Delta F := F_{\xi_*}(x) - F_{\xi^*}(x) = \mathbb{P}\{Z'/Z'' \le C_* x\} - \mathbb{P}\{Z'/Z'' \le C_* x\},$$

where $C_* = (1+\eta)/c(1-\eta)$ and $C^* = (1-\eta)/c(1+\eta)$. By Lemma 4.1, we have that

$$\Delta F = \psi\Big(\frac{(1-\eta)^\alpha}{(1+\eta)^\alpha x^\alpha}\Big) - \psi\Big(\frac{(1+\eta)^\alpha}{(1-\eta)^\alpha x^\alpha}\Big),$$

where $\psi(y) = 1/(c^{-\alpha}y+1)$. By the mean value theorem, since $|\psi'(y)| = c^{-\alpha}/(c^{-\alpha}y+1)^2$ is monotone decreasing in $y > 0$, the last expression is bounded above by:

$$\begin{aligned}
&|\psi'(\frac{(1-\eta)^\alpha}{(1+\eta)^\alpha x^\alpha})|\Big|\frac{(1+\eta)^\alpha}{(1-\eta)^\alpha x^\alpha} - \frac{(1-\eta)^\alpha}{(1+\eta)^\alpha x^\alpha}\Big| \\
&\qquad = \frac{c^{-\alpha}(1+\eta)^{2\alpha}x^\alpha}{(c^{-\alpha}(1-\eta)^\alpha + (1+\eta)^\alpha x^\alpha)^\alpha}\Big(\frac{(1+\eta)^{2\alpha}-(1-\eta)^{2\alpha}}{(1-\eta^2)^\alpha}\Big) \\
&\qquad \le \frac{(1+\eta)^\alpha}{2(1-\eta)^\alpha}\Big(\frac{(1+\eta)^{2\alpha}-(1-\eta)^{2\alpha}}{(1-\eta^2)^\alpha}\Big) = \frac{(1+\eta)^{2\alpha}-(1-\eta)^{2\alpha}}{2(1-\eta)^{2\alpha}}, \qquad (4.32)
\end{aligned}$$

where the last inequality follows from the fact that $ab/(a+b)^2 \le 1/2$, a, $b \in \mathbb{R}$, with $a := c^{-\alpha}(1-\eta)^\alpha$ and $b = (1+\eta)^\alpha x^\alpha$. By using the mean value theorem again, we obtain that, for some $\theta \in [-1,1]$, the right–hand side of (4.32) is bounded above by

$$\frac{2\alpha(1+\theta\eta)^{2\alpha-1}}{(1-\eta)^{2\alpha}}2\eta \le \alpha 2^{2\alpha+2+|2\alpha-1|}\eta \le \alpha 2^{4\alpha+3}\eta,$$

for all $\eta \in (0, 1/2)$. □

References

Alon, N., Duffield, N., Lund, C. & Thorup, M. (2005), Estimating sums of arbitrary selections with few probes, *in* 'Symposium on Principles of Database Systems (PODS)', Baltimore, Maryland, USA.

Alon, N., Matias, Y. & Szegedy, M. (1996), The space complexity of approximating the frequency moments, *in* 'Proceedings of the 28th ACM Symposium on Theory of Computing', pp. 20–29.

Cohen, E. (1997), 'Self–estimation framework with applications to transitive closure and reachability', *J. Comput. System Sci.* **55**, 441–453. Special issue of selected papers from FOCS'94.

Cormode, G. (2003), Stable distributions for stream computations: it's as easy as 0,1,2, *in* 'Workshop on Management and Processing of Massive Data Streams at FCRC'.

Cormode, G. & Muthukrishnan, S. (2003*a*), *Algorithms - ESA 2003*, Vol. 2832 of *Lecture Notes in Computer Science*, Springer, chapter Estimating Dominance Norms of Multiple Data Streams, pp. 148 –160.

Cormode, G. & Muthukrishnan, S. (2003*b*), 'What's hot and what's not: tracking most frequent items dynamically'.

Datar, M., Immorlica, N., Indyk, P. & Mirrokni, V. S. (2004), Locality–sensitive hashing scheme based on p–stable distributions, *in* 'SCG '04: Proceedings of the twentieth annual symposium on Computational geometry', ACM Press, New York, NY, USA, pp. 253–262.

Feigenbaum, J., Kannan, S., Strauss, M. & Viswanathan, M. (1999), An approximate l1-difference algorithm for massive data streams, *in* 'FOCS '99: Proceedings of the 40th Annual Symposium on Foundations of Computer Science', IEEE Computer Society, Washington, DC, USA, p. 501.

Fong, J. H. & Strauss, M. (2000), An approximate lp-difference algorithm for massive data streams, *in* 'STACS '00: Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science', Springer-Verlag, London, UK, pp. 193–204.

Gilbert, A., Kotidis, Y., Muthukrishnan, S. & Strauss, M. (2001*a*), 'QuickSAND: Quick Summary and Analysis of Network Data DIMACS technical report'.

Gilbert, A., Kotidis, Y., Muthukrishnan, S. & Strauss, M. (2001*b*), Surfing wavelets on streams: one-pass summaries for approximate aggregate queries, *in* 'Procedings of VLDB, Rome, Italy'.

Gilbert, A., Kotidis, Y., Muthukrishnan, S. & Strauss, M. (2002), 'How to summarize the universe: Dynamic maintenance of quantiles'.

Henzinger, M., Raghavan, P. & Rajagopalan, S. (1998), 'Computing on data streams'.

Indyk, P. (2000), Stable distributions, pseudorandom generators, embeddings and data stream computation, *in* 'IEEE Symposium on Foundations of Computer Science', pp. 189–197.

Indyk, P. (2003), Better algorithms for high-dimensional proximity problems via asymmetric embeddings, *in* 'SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms', Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 539–545.

Indyk, P. & Woodruff, D. (2005), Optimal approximations of the frequency moments of data streams, *in* 'STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing', ACM Press, New York, NY, USA, pp. 202–208.

Ishai, Y., Malkin, T., Strauss, M. J. & Wright, R. N. (2006), Private multiparty sampling and approximation of vector combinations, In preparation.

Muthukrishnan, S. (2003), Data streams: algorithms and applications, Preprint. `http://www.cs.rutgers.edu/~muthu/`.

Nisan, N. (1990), Pseudorandom generator for space bounded computation, *in* 'Proceedings of the 22nd Symposium on Theory of Computing (STOC)', pp. 204–212.

Razborov, A. (1992), 'On the distributional complexity of disjointness', *Theoretical Computer Science* **106**(2), 385–390.

Resnick, S. I. (1987), *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York.

Saks, M. & Sun, X. (2002), Space lower bounds for distance approximation in the data stream model, *in* 'STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing', ACM Press, New York, NY, USA, pp. 360–369.

Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, London.

Stoev, S. & Taqqu, M. S. (2005), Extremal stochastic integrals: a parallel between max–stable processes and $\alpha-$stable processes, Preprint. `http://math.bu.edu/people/sstoev/tmp/ext-int-7.pdf`.