Max–stable Random Sketches: Estimation of Distances, Norms and Dominance Norms

Stilian Stoev (sstoev@umich.edu)
Department of Statistics
University of Michigan

Theoretical Computer Science seminar
December 7, 2005

Joint work with Murad S. Taqqu at Boston University

Part I: Motivation and background

Random Sketches and Data Streams: motivation

Signals which cannot be stored and processed with conventional methods are emerging in:

- Telephone call monitoring (unmanageably large data bases).
- Internet traffic monitoring (vast, rapidly growing amounts of data).
- Sensing applications (restrictions on power, bandwidth, etc.).
- Online transactions (Banking, Stock market data, E-commerce, etc.).

The streaming model – a general framework. Let

$$f(i), \quad 1 \leq i \leq N$$

be a signal with large domain size N. Starting with $f(i) = 0$, the signal f is updated sequentially in time:

- cash register Observe data pairs $(i, a(i))$, and update f:

$$f(i) := f(i) + a(i)$$

- aggregated Observe data pairs $(i, f(i))$ directly.

Random Sketches

Instead of storing the signal $f(i)$, $1 \leq i \leq N$, keep:

$$S_j(f) = \sum_{i=1}^{N} f(i) X_j(i), \quad 1 \leq j \leq K,$$

where $K << N$ and where $X_j(i)$ are independent random variables.

The $S_j(f)$'s $1 \leq j \leq N$ is called the random sketch of f.

- Important: Can generate $X_j(i)$'s "on the fly" from small $\mathcal{O}(K \log_2(N))$ set of seeds in main memory.
- Can update the sketch sequentially in the most general, cash register model.
- Can approximate norms and inner products!

If $\text{Var}(X_j(i)) = \sigma^2$, then $\text{Var}(S_j(f)) = \|f\|_{\ell_2}^2 \sigma^2, \quad \|f\|_{\ell_2}^2 = \sum_i f(i)^2$.

- So

$$\|f\|_{\ell_2}^2 \approx \frac{1}{K} \sum_j S_j(f)^2.$$

- For two signals f and g, by linearity,

$$\|f - g\|_{\ell_2}^2 \approx \frac{1}{K} \sum_{j=1}^{K} |S_j(f) - S_j(g)|^2,$$

Max–stable Sketches

Let $f(i)$ be a non–negative signal.

Definition The max–stable sketch $E_j(f)$, $1 \leq j \leq K$ is:

$$E_j(f) = \bigvee_{1\leq i\leq N} f(i)Z_j(i) = \max_{1\leq i\leq N} f(i)Z_j(i),$$

where $Z_j(i)$'s are independent, standard $\alpha-$Fréchet random variables.

∘ The r.v. Z is $\alpha-$Fréchet if

$$\mathbb{P}\{Z \leq x\} = \begin{cases} \exp\{-\sigma^\alpha x^{-\alpha}\} & , \text{ if } x > 0 \\ 0 & , \text{ if } x \leq 0, \end{cases}$$

where $\alpha > 0$ and where $\sigma > 0$.

∘ The parameter $\sigma > 0$ is called the scale coefficient of Z. If $\sigma = 1$ then Z is called standard.

Properties:

- scale invariance For all $a > 0$, the r.v. aZ is $\alpha-$Fréchet with scale coefficient $a\sigma$. Indeed, $\forall x > 0$:

  $$\mathbb{P}\{aZ \leq x\} = \exp\{-\sigma^\alpha (x/a)^{-\alpha}\} = \exp\{-(a\sigma)^\alpha x^{-\alpha}\}.$$

- max–stability For all $a_i \geq 0$,

  $$\bigvee_i a_i Z(i) \stackrel{d}{=} (\sum_i a(i)^\alpha)^{1/\alpha} Z,$$

  for any independent identically distributed $\alpha-$Fréchet $Z,\ Z(i)$'s.

More on max–stability...

Indeed, by independence, $x > 0$ and $\forall a_i \geq 0$,

$$\begin{aligned} \mathbb{P}\{\vee_i a_i Z(i) \leq x\} &= \mathbb{P}\{a_i Z(i) \leq x, \text{ for all i }\} \\ \prod_i \mathbb{P}\{a_i Z(i) \leq x\} &= \exp\{-\sum_i (a(i)\sigma)^\alpha x^{-\alpha}\}. \end{aligned}$$

The last equals $\mathbb{P}\{cZ \leq x\}$ where

$$c = (\sum_i a(i)^\alpha)^{1/\alpha}.$$

• Max–stability: Maxima of independent $\alpha-$Fréchet is $\alpha-$Fréchet.

• Sum–stability: Sums of independent $p-$stable is $p-$stable.

Analogies between max– and sum–stability can be drawn.

Notation: If Z is $\alpha-$Fréchet with scale coefficient σ, then

$$\|Z\|_\alpha := \sigma.$$

Caution: $\|Z\|_\alpha$ does not denote the "norm" $(\mathbb{E}Z^\alpha)^{1/\alpha}$.

• heavy tails For $\alpha-$Fréchet Z,

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

• moments $\mathbb{E}Z^r = \infty$ if $r \geq \alpha$ and

$$\mathbb{E}Z^r = \|Z\|_\alpha^r \Gamma(1 - r/\alpha) < \infty, \ \forall r \in (0, \alpha).$$

• medians

$$\text{med}(Z) = \|Z\|_\alpha (\ln 2)^{-1/\alpha}.$$

Back to max–stable sketches...

Max–stability implies that $E_j(f),\ 1 \le j \le K$ are independent $\alpha-$Fréchet variables. For their scale coefficients, we have:

$$\|E_j(f)\|_\alpha = (\sum_{i=1}^{N} f(i)^\alpha)^{1/\alpha} = \|f\|_{\ell_\alpha}.$$

Thus, by the moment and median properties, one can estimate the norm $\|f\|_{\ell_\alpha}$.

- For any $0 < r < \alpha$:

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

-

$$\|f\|_{\ell_\alpha} \approx (\ln 2)^{1/\alpha} \text{median}\{E_j(f),\ 1 \le j \le K\}.$$

Limitations and strengths

1. limitation Max–stable sketches are non–linear. Do not work for the cash register model.

2. limitation Signal must be non–negative.

3. strength Work for any $\alpha > 0$ and not only for $0 < \alpha \le 2$ (unlike the sum–stable sketches).

4. strength Natural when working with dominance relations. If $f(i) \le g(i)$, then $E_j(f) \le E_j(g)$.

Part II: Some results

Approximation of ℓ_α–norms

Let $f(i),\ 1 \le i \le N$ be a non–negative signal with α–max–stable sketch $E_j(f),\ 1 \le j \le K$.

Define the estimators:

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

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

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

We have the following result.

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

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

and

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

provided

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

where $C = C(\alpha, r) > 0$.

Caveats: $E_j(f)$ is not linear in f, so $E_j(\cdot)$ is not a linear embedding. Thus $\alpha > 2$ does not contradict known lower bounds on complexity.

Usual linearity is not a natural framework for max–stable sketches.

Max–stable sketches are max–linear.

Dominance ℓ_α–norms

Let $f_1(i), \ldots, f_m(i)$ be m non–negative signals.

Goal: Estimate the max–dominance norm

$$\mathrm{dom}_{\max,\alpha}(f_1, \ldots, f_m) := (\sum_{i=1}^{N} \max\{f_k(i),\ 1 \le k \le m\}^\alpha)^{1/\alpha},$$

that is,

$$\mathrm{dom}_{\max,\alpha}(f_1, \ldots, f_m) = \| \vee_{1 \le k \le m} f_k(\cdot)\|_{\ell_\alpha}.$$

Relevance:

- Data stream: $(i, k, f_k(i))$ where i is an IP address and where $k = 1, \ldots, m$ is the k-th transmission. Here $f_k(i)$ is the number of bytes transmitted by i on its k-th transmission.

  The dominance norm yields a measure of the "maximum possible utilization of the network" (Cormode and Muthukrishnan, 2002).

- Electrical grid load.

- Financial applications.

Dominance ℓ_α–norms (cont'd)

Cormode and Muthukrishnan (2002) estimate the max–dominance norm by using p–stable sketches of the signals $f_k(\cdot)$ for a sequence of p's.

Letting $p \to 0+$, they show that $\mathrm{dom}_{\max,1}$ is approximated within a factor of $(1 \pm \epsilon)$ with probability at least $(1 - \delta)$, if

$$\text{Space} = \mathcal{O}(\log(M) + \epsilon^{-1} \log(N) \log\log(N)) \log(1/\delta)/\epsilon^2)$$

and

$$\text{Per item time} = \mathcal{O}(\log(f_k(i)) \log(N) \log(1/\delta)/\epsilon^4).$$

Here $f_k : \{1, \ldots, N\} \to \{0, 1, \ldots, M\}$.

Problems:

- Multiple per item operations.
- p–stable variables with $p \approx 0$ have extremely high variability. Thus, problems with simulation and computational error.
- Need to take a sequence of p–stable sketches with p's converging to 0!
- Can only approximate $\mathrm{dom}_{\max,\alpha}$ for $\alpha = 1$!

Dominance ℓ_α–norms (cont'd)

Our solution:

Let $E_j(f_k),\ 1 \le j \le K$ be the α–max–stable sketch of $f_k, k = 1, \dots, K$.

By max–linearity of the max–stable sketch:

$$E_j(f^*) = \max_{1 \le k \le m} E_j(f_k), \quad \forall j,$$

where

$$f^*(i) = \max_{1 \le k \le m} f_k(i).$$

Hence

$$\begin{aligned} \mathrm{dom}_{\max,\alpha}(f_1, \dots, f_m) &= \|f^*\|_{\ell_\alpha} \approx \ell_{\alpha,\mathrm{med}}(f^*) \\ &= (\ln 2)^{1/\alpha} \mathrm{median}\{\vee_k E_j(f_k),\ 1 \le j \le K\}. \end{aligned}$$

Our first results on ℓ_α–norm approximation imply:

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

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

provided

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

where $C = C(\alpha, r) > 0$.

Dominance ℓ_α–norms (cont'd)

Benefits:

- Faster per item processing. No need to take multiple sketches and let a parameter converge to 0.
- Max–dominance norm approximation for any $\alpha > 0$!
- Easier to implement, i.e. no danger of large simulation and computational error.

Why does the Cormode and Muthukrishnan method work?

Let X_p be a standard (symmetric*) p–stable variable with $p > 0$. By Exercise 1.29, p. 54 in Samorodnitsky and Taqqu (1994),

$$|X_p|^p \xrightarrow{d} 1/E, \quad \text{where } E \text{ is Exponential.}$$

Note: If E is Exponential, i.e. $\mathbb{P}\{E \le x\} = 1 - e^{-x}$, $x > 0$, then

$$\mathbb{P}\{1/E \le x\} = \exp\{-1/x\} = \exp\{-x^{-1}\}.$$

That is $Z = 1/E$ is α–Fréchet with $\alpha = 1$.

• Cormode and Muthukrishnan's algorithm is an approximation to our algorithm in the special case $\alpha = 1$.

*can be relaxed

Approximation of certain distances

Let f and g be non–negative signals. For any $\alpha > 0$,

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

is a distance. It is a norm only when $\alpha = 1$.

- Note that for any $a,\ b \geq 0$:

$$|a - b| = 2a \vee b - a - b.$$

- Thus,

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

- Can approximate $\rho_\alpha(f,g)$ in terms of the max–stable sketches of f and g, since $E_j(f \vee g) = E_j(f) \vee E_j(g)$.

Replacing $\|f\|_{\ell_\alpha}$ by $\ell_{\alpha,\mathrm{med}}(f)$, we get

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

Theorem *Let f and g be separated, i.e.*

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

for some $\eta > 0$.

Then, for all $\epsilon,\ \delta > 0$,

$$\mathbb{P}\{|\widehat{\rho}/\rho - 1| \leq \epsilon/\eta\} \geq 1 - 3\delta,$$

provided $K = \mathcal{O}(\ln(1/\delta)/\epsilon^2)$. Here $\widehat{\rho} = \widehat{\rho}_{\alpha,\mathrm{med}}(f,g)$ and $\rho = \rho_{\alpha,\mathrm{med}}(f,g)$.

Approximation of certain distances (cont'd)

Idea of Proof:

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

Since $\rho = 2\|f \vee g\|_{\ell_\alpha}^\alpha - \|f\|_{\ell_\alpha}^\alpha - \|g\|_{\ell_\alpha}^\alpha$, we get:

$$|\widehat{\rho} - \rho| \leq \mathcal{O}(\epsilon\|f \vee g\|_{\ell_\alpha}^\alpha) \leq \mathcal{O}(\epsilon/\eta)\rho. \quad \square$$

• Note: The separation condition $\rho(f,g) \geq \eta\|f \vee g\|_{\ell_\alpha}^\alpha$ cannot be removed!

Indeed, if $1_A(i) = f(i)$ and $1_B(i) = g(i)$ are indicator signals, then

$$\rho_\alpha(f,g) = |A \cap B|.$$

• Estimation of the size of the intersection $A \cap B$ is provably hard problem: Razborov (1992) (also Cormode and Muthukrishnan, 2002), if no separation condition is imposed.

Point queries

Given is a non–negative signal f through its α–max–stable sketch $E_j(f)$, $1 \le j \le K$.

Goal: Estimate $f(i_0)$ for some $1 \le i_0 \le N$ using only the max–stable sketch.

Consider the quantities:

$$g_j(i_0) = E_j(f)/Z_j(i_0), \quad 1 \le j \le K.$$

and sort them:

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

- Define the point estimator:

$$\widehat{f}(i_0) := g_{(1)}(i_0) = \min_{1 \le j \le K} \frac{E_j(f)}{Z_j(i_0)}.$$

and a criterion

$$\text{criterion}(i_0) = 1, \quad \text{if } g_{(1)}(i_0) = g_{(2)}(i_0).$$

- Intuition:

$$g_j(i_0) = \frac{f(i_0)Z_j(i_0) \vee \max_{i \ne i_0} f(i)Z_j(i)}{Z_j(i_0)} = f(i_0) \vee \xi_j(i_0),$$

where

$$\xi_j(i_0) = \bigvee_{i \ne i_0} f(i)Z_j(i)/Z_j(i_0).$$

Point queries (cont'd)

Thus:

$$\widehat{f}(i_0) = g_{(1)}(i_0) = f(i_0) \vee \min_j \xi_j(i_0).$$

- Note that $\widehat{f}(i_0)$ equals $f(i_0)$ if

$$\min_j \xi_j(i_0) < f(i_0).$$

- Why does the criterion work?

∘ Recall criterion$(i_0) = 1$, if:

$$g_{(1)}(i_0) = f(i_0) \vee \xi_{(1)}(i_0) = f(i_0) \vee \xi_{(2)}(i_0) = g_{(2)}(i_0),$$

then

Either $f(i_0) < \xi_{(2)}(i_0) = \xi_{(1)}(i_0)$ Or $f(i_0) \geq \xi_{(2)}(i_0) = \widehat{f}(i_0)$.

∘ The first possibility

$$\xi_{(2)}(i_0) = \xi_{(1)}(i_0)$$

happens only with probability zero.

∘ Hence, criterion$(i_0) = 1$, implies $\widehat{f}(i_0) = f(i_0)$, with probability 1.

Point queries (cont'd)

One can show that

$$\widehat{f}(i_0) = f(i_0) \vee c_f Z'/Z'', \quad \text{with } c_f = (\sum_{i \neq i_0} f(i)^\alpha)^{1/\alpha}$$

where Z' and Z'' are independent, standard α−Fréchet.

- Thus,

$$\mathbb{P}\{\widehat{f}(i_0) = f(i_0)\} = \mathbb{P}\{Z'/Z'' \leq f(i_0)/c_f\}.$$

- One can easily show that

$$\mathbb{P}\{Z'/Z'' \leq x\} = \frac{1}{x^{-\alpha} + 1}, \quad \forall x > 0$$

and therefore:

Theorem *If* $f(i_0) \geq \epsilon \|f\|_{\ell_\alpha} \geq \epsilon c_f$, *then*

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

provided $K \geq \ln(1/\delta)/\epsilon^\alpha$.

- Intuition: Relatively large $f(i_0)$'s, can be recovered exactly with high probability.

- We have similar $\epsilon - \delta$ guarantees for

$$\mathbb{P}\{\text{criterion}(i_0) = 1\}.$$

Point queries: some limitations

Let $\alpha = 1$, fix $\delta > 0$ and $\epsilon > 0$.

∘ We can recover, with probability $(1 - \delta)$ at most $1/\epsilon$ of the largest signal values $f(i)$.

∘ We need space $\log(1/\delta)/\epsilon$.

• A naive method: maintain the top $1/\epsilon$ signal values. It needs space $1/\epsilon$.

• Max–stable sketches do not provide exponential improvement over the naive method.

• Possible advantages:

1. Direct access to the signal may not be possible. If we have only max–stable sketches, we can recover exactly signal values.

2. Point queries can be extended to range queries.

3. Max–stable sketches are universal: norms, distances, dominance norms.

4. Very easy (fast) to simulate: If U uniformly distributed on (0,1), then

$$Z = \ln(1/U)^{-1/\alpha} \quad \text{is } \alpha\text{–Fréchet.}$$

Conclusion

Max–stable sketches provide a new non–linear perspective in random sketching.

Strengths

- Efficient approximation of $\|\cdot\|_{\ell_\alpha}$.
- Efficient and natural way to estimate dominance norms.
- Added benefit: can recover large signal values exactly, with high probability.
- Easy/fast simulation of the $Z_j(i)$'s.

A lot more to be done!

Future work

- Range queries: Estimate the signal over a range by using interpolation?
- Count dominance: Theory almost there.
- Inclusion queries: Theory almost there.
- Real algorithms: Prove that real algorithms exist in all theoretical scenarios. Some work done, more to do!
- Possible, natural applications?

Thank you!