AT THE CONFLUENCE OF STREAMS; ORDER, INFORMATION & SIGNALS

S. Guha
UPenn

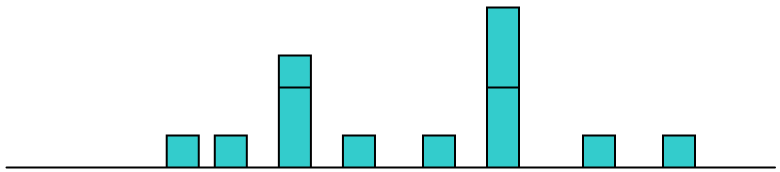
Joint work with P. Indyk and A. McGregor.

Data Streams

- We are given a sequence of input $x_1,\ldots,x_i,\ldots,x_m$ and have to compute some function f
- Computation proceeds in passes
- Space is restricted
- Any x_i not explicitly remembered: inaccessible in the same pass

Example

- Sitting next to the wireless router - delay suffered by every packet
- Interested in the distribution of the delay
  - Pkt 1, 0.2 ms..
  - Pkt 2, 0.3
  - …
  - Pkt ?, 0.2 $\Rightarrow$ 2 pkts at 0.2 …



- Stream is specified by Updates.
- Every stream item is a <i (=delay),+1>
- Assume we normalize somehow. No deletions (this talk).

- Of course we cannot store an explicit vector … $i \in [n]$
- Space in o(n) & input is given is a piece meal fashion.

Data Streams (this talk) …

- Understanding the impact of the order of the input on computation.

- A view of 2 well known problems

  - (dis) Similarity of streams -> 1996
  - Order statistic -> 1978

Distances between 2 streams

- Channel 9 similar to channel 1 ?
  - Distributions X & Y
- "I believe the distribution is the same as last Thursday"

- $(1+\varepsilon)$ approximation; i.e., $(1+\varepsilon)$ D(X,Y)
- Alon, Matias & Szegedy
- Johnson Lindenstrauss
- Feigenbaum, Kannan, Strauss & Vishwanathan ℓ_1 but in an "aggregate model" $\Rightarrow$ … (i,# of packets) …
- Indyk ℓ_k for $0<k\leq 2$ …
- Tight results for $k \geq 3$ have since been achieved…

Random Projections

- [Johnson, Lindenstrauss] 1984
- Given a matrix A whose elements are iid Gaussian, and any vector x, with high prob.

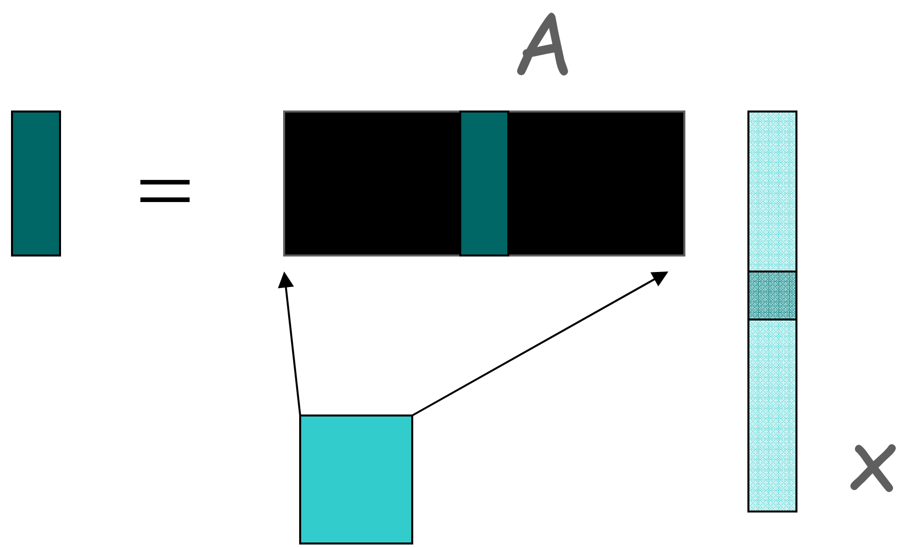
$$\|x\|_2 \leq \|Ax\|_2 \leq (1+\varepsilon)\|x\|_2$$

if $x \in R^n$ then $A \in R^{n \times O(\log n)}$
$\Rightarrow Ax \in R^{O(\log n)}$.

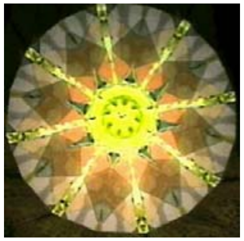
Dimensionality reduction, nearest nbr searches.

What it achieves

- Computes Norm when elements arrive out of order.



Note: A proof that such a pseudorandom generator exists is Necessary – and is not always easy.



A Kaleidoscope of questions

- Philosophical …

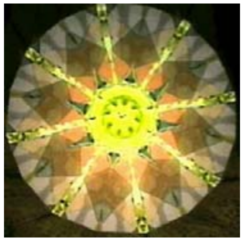
Which other distances are approximable? What property?

(likely) That is it. The only approximable distances are likely to be norms, i.e., function of $\{x_i - y_i\}$.

Stable distributions appear as the key idea in context thing …

- Pragmatic …

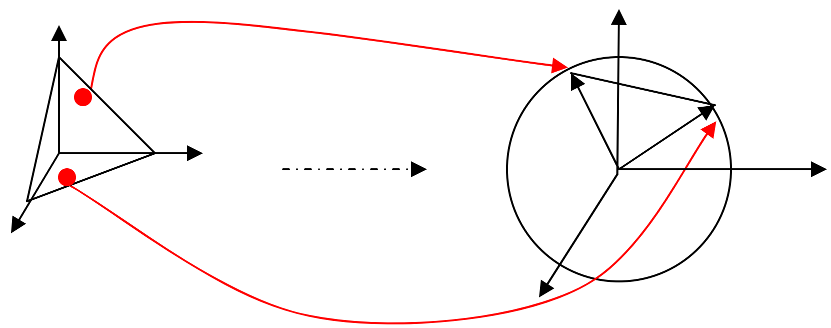
A Kaleidoscope of questions



- Philosophical …

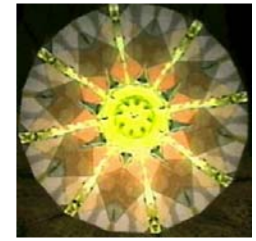
Example … $D^2=\Sigma_i (\sqrt{x_i} - \sqrt{y_i})^2$

(squared) Hellinger distance



- Pragmatic …

A Kaleidoscope of questions



- Philosophical …

  Example … $D^2 = \Sigma_i (\sqrt{x_i} - \sqrt{y_i})^2$

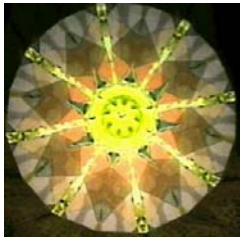
  (squared) Hellinger distance

  "Aggregate" Model (FKSV) easy …
  Hard in update models.

  $\Sigma_i \sqrt{|x_i - y_i|}$ is easy (1/2 stable distribution)

- Pragmatic …

A Kaleidoscope of questions



- Philosophical …

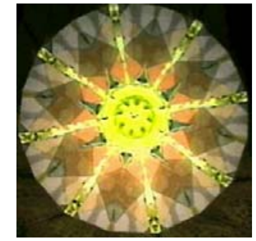
  Example … $D^2 = \Sigma_i (\sqrt{x_i} - \sqrt{y_i})^2$

  (squared) Hellinger distance

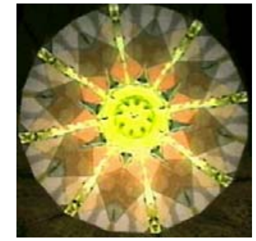
  Small space embedding (into ℓ_2) easy
  Hard for streaming

- Pragmatic …

A Kaleidoscope of questions

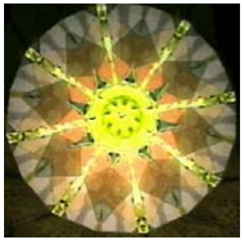


- Philosophical.
- Pragmatic …
  - What measures of distances are meaningful for distributions ?
    - Hypothesis testing:
      - f-divergences or Ali-Silvey-Cziszar divergences
    - Mathematical programming:
      - Bregman divergences
  - Model “Risk” etc.,



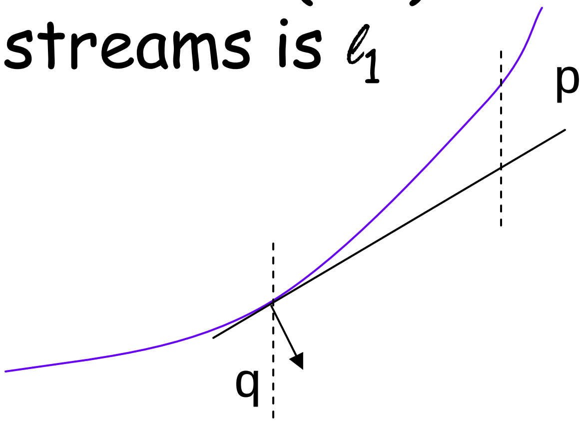
A Kaleidoscope of questions

- Philosophical.
- Pragmatic.
  - f-divergences:
    - Pick a j from x and consider the expected likelihood $D_f(x,y)=E_{x,j}\ f(y_j/x_j)$ provided $f(1)=0$, f convex…
    - ? $KL(x,y) = \sum_j x_j \log (x_j/y_j) \Rightarrow f(u)=- \log u$
    - ? Hellinger^2 $= \sum_j (\sqrt{x_j} - \sqrt{y_j})2 = \sum_j x_j (1-\sqrt{(y_j/x_j)})^2$ or $f(u)=(1-\sqrt{u})^2$.
    - $\ell_1 = \sum_j | x_j - y_j | = \sum_j x_j | 1 - (y_j/x_j)|$ or $f(u)=|1-u|$
    - Also arises from loss functions in learning …

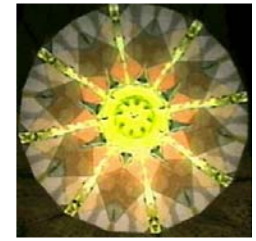


A Kaleidoscope of questions

- Philosophical.
- Pragmatic.
  - The only f-divergence which can be $(1+\varepsilon)$ approximated over update streams is ℓ_1
  - Bregman divergences..
    - Potential field F
    - Convex F



$$B_F(\mathbf{p},\mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - (\nabla F(\mathbf{q})) \circ (\mathbf{p}-\mathbf{q})$$



A Kaleidoscope of questions

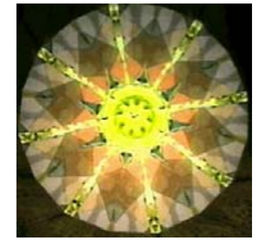
- Philosophical.
- Pragmatic.
  - Example:
    - $F(x)=x^2 \Rightarrow B(x,y)=x^2-y^2-2y(x-y)=(x-y)^2 \Rightarrow \ell_2$!
    - $F(x)=x \lg x \Rightarrow$

      $B(x,y)=x \lg x - y \lg y - (1+\lg y)(x-y) = x \lg (y/x) - x + y$

      $\Rightarrow$ Gen. KL div

$$B_F(\mathbf{p},\mathbf{q})=F(\mathbf{p}) - F(\mathbf{q}) - (\nabla F(\mathbf{q})) \circ (\mathbf{p}-\mathbf{q})$$

A Kaleidoscope of questions



- Philosophical.
- Pragmatic.
  - Bregman div. :
    - ℓ_2 is sketchable/estimable in small space.
    - What about the others? Sorry …

How?

- Lemma (one part): ρ,c >0, distributions x,y
  - For all (in some range) M,δ

    if min $\{ \phi(\delta,2\delta), \phi(2\delta,\delta) \} \geq cM^{\rho}$ $\{\phi(M\delta,(M+1)\delta)+ \phi((M+1)\delta,M\delta) \}$

    then $\exists\ \gamma>0$ such that to get a n^{γ} approximation of $\sum_i \phi(x_i,y_i)$ over [2n] we need $\Omega(n)$ space.

Consequence …

- If f′,f″ exist …

$\phi(M\delta,(M+1)\delta) = M\delta\, f(1+1/M)$

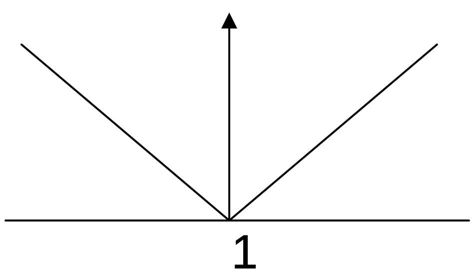
$\leq M\delta\, [\, f(1) + f'(1)/M + f''(\zeta)/(2M^2)]$

Suppose f′(1) exists and $\neq 0$; then consider g(u)=f(u) – f′(1)(u-1).

Well known: the change has no effect …

Lemma applies …

Exceptions?



Proof of the Lemma

- Reduction from Communication Complexity of set disjointness.
- Alice and Bob have $\approx$ n/4 numbers each from [n]. How many bits do they need to exchange to find a common number if such an element exists.
- Θ(n/P) even with P rounds
- If an efficient streaming algorithm existed then they can communicate the "state" of the algorithm.
  - But dimensionality reduction may be possible…
  - Two copies: critical that you are allowed updates

Other results

- Bregman: F″ vanishes or diverges polynomially at the nbd of 0 $\Rightarrow$ Same conclusion. Note F″=constant for ℓ_2
- If f(0) is bounded then any symmetric f-divergence can be approximated to $\pm\varepsilon$ using $\sqrt{n}\log^{O(1)} n$ space
- If one distribution is known $\Rightarrow$ Polylog.

Takeaway …

- Order of the input is important…
- Hellinger: easy in aggregated model
  can embed in small space

  hard in update models
- It's the update which is the problem.

Changing gears…

- Analysis of streaming model is typically worst case.

- What if we consider average case?
- Average over what?
- The order. ⇒ Exchangeability …

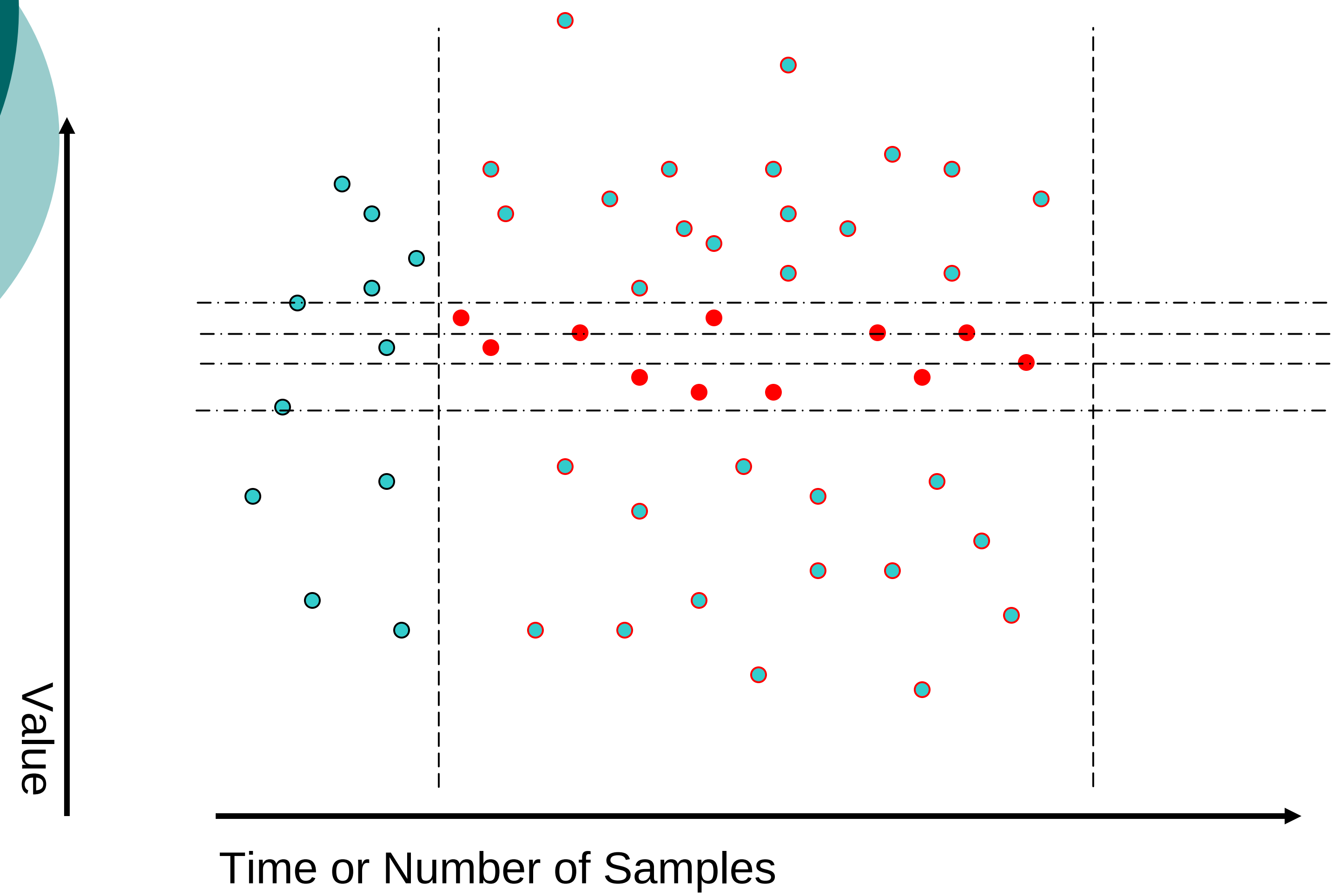
Order Statistic – Median finding

- Given a sequence of n numbers, find the median. Space is restricted.
- Munro Paterson 1978
  - For p passes $n^{1/p}$ space suffices
  - Mention that for random order log log n pass and polylog space appears feasible, but known techniques do not seem to work.
- Manku, Rajagopalan, Lindsay; Greenwald Khanna,
  - error $\pm\ \varepsilon$ n using $O((1/\varepsilon) \log n)$ space

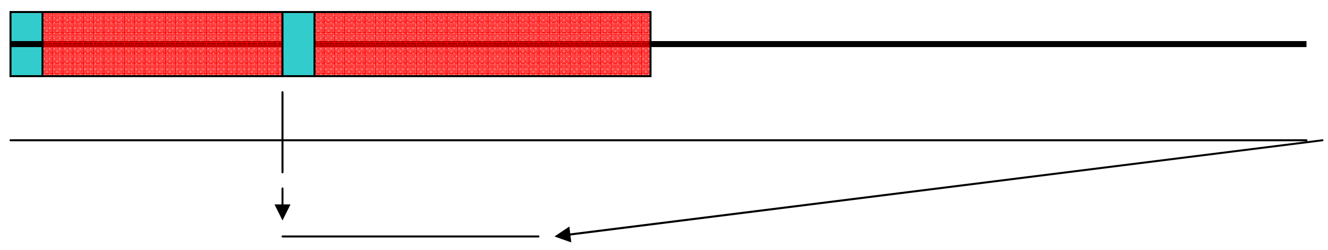
Exact Median Finding

- For p passes $n^{1/p}$ space suffices
  - This is best possible $\Rightarrow \Omega(\log n)$ passes.
- error $\pm\, \varepsilon$ n using $O((1/\varepsilon) \log n)$ space
  - 1 pass adversarial order $\pm\, n^{\delta}$ error $\Rightarrow \Omega(n^{1-\delta})$ space
  - 1 pass random order $\pm\, n^{1/2+\varepsilon}$ error in polylog space
  - Multipass extention not automatic …

A new hope in thousand words



A proof is not an idea …



- Now we do not know the length of the stream anymore – it is $\zeta N \pm O(\sqrt{N})$

- Ignore and repeat. The constant in O() increases but N is already ζN.

The takeaway …

- Random order gives an exponential speedup in passes.
- Permuting your data might give you a faster algorithm. The question is of course to analyze the benefit.

The road ahead

- A promising idea
  - Assume random order
  - Prove your claim
  - Go back and "fix/simulate" the randomness

  - Clustering data streams
    - G., Motwani, Mishra & O'callaghan – $n^{1/p}$ & $2^{O(p)}$
    - Meyerson : Random order $\log^{O(1)} n$ & $O(1)$
    - Charikar, Panigrahy, O'callaghan: $\log^{O(1)} n$ & $O(1)$
  - More examples

- What about SVD ?
  - Assume any random order you see fit
  - Can you analyze passes/runtime/space better?

That's all Folks