Estimating Lp Norms

Piotr Indyk
MIT

Lecture 3

Recap/Today

- Two algorithms for estimating L_2 norm of a stream
  - A stream of updates $(i,1)$ interpreted as

    $$x_i = x_i + 1$$

    (fractional and negative updates also OK)
  - Algorithms maintain a linear sketch Rx, where R is a $k*m$ (pseudo)-random matrix
  - Use $||Rx||_2^2$ to estimate $||x||_2^2$
  - Polylogarithmic space
- Today:
  - Yet another algorithm for L_2 estimation
    - Generalizes to any L_p, $p \in (0,2]$
    - Polylogarithmic space
  - An algorithm for L_k estimation, $k \geq 2$
    - Works only for positive updates
    - Uses sampling, not sketches
    - Space: $O(k\, m^{1-1/k} / \varepsilon^2)$ for $(1 \pm \varepsilon)$-approximation with const. probability

Median Estimator

- Again we use a linear sketch $Rx=[Z_1 \ldots Z_k]$, where each entry of R has distribution $N(0,1)$, $k=O(1/\varepsilon^2)$
  - Therefore, each of Z_i has $N(0,1)$ distribution with variance $\sum_i x_i^2 = ||x||_2^2$
  - Alternatively, $Z_i = ||x||_2 G_i$, where G_i drawn from $N(0,1)$
- How to estimate $||x||_2$ from $Z_1 \ldots Z_k$?
- In Algorithms I, II, we used $Y=[Z_1^2 + \ldots + Z_k^2]/k$ to estimate $||x||_2^2$
- But there are many other estimators out there…
- E.g., we could instead use

$$Y = \mathrm{median}[\, |Z_1|, \ldots , |Z_k| \,] / \mathrm{median}[|G|]$$

  to estimate $||x||_2$ (G drawn from $N(0,1)$)
- The rationale:
  - $\mathrm{median}\,[\, |Z_1|, \ldots , |Z_k| \,] = ||x||_2 \,\mathrm{median}\,[\, |G_1|, \ldots , |G_k| \,]$
  - For "large enough" k, $\mathrm{median}\,[\, |G_1|, \ldots , |G_k| \,]$ is "close to" $\mathrm{median}[|G|]$ (next two slides)

* median of an <u>array</u> A of numbers is the usual number in the middle of the sorted A

** M is the median of a <u>random variable</u> U if $\Pr[U \le M] = ½$

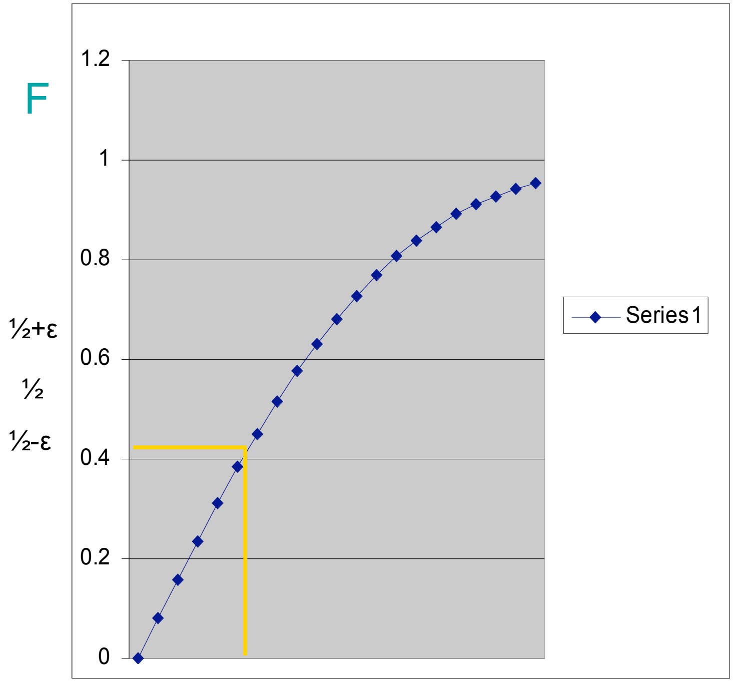
Closeness in probability

- Lemma 1: Let $U_1 \ldots U_k$ be i.i.d. real random variables chosen from any distribution having continuous c.d.f. F and median M
  - I.e., $F(t)=\Pr[U_i < t]$ and $F(M)=1/2$

Define $U=\text{median}\,[U_1,\ldots,U_k]$. Then, for some absolute const. $C>0$

$$\Pr[F(U)\in(1/2-\varepsilon,1/2+\varepsilon)]\geq 1-e^{-C\varepsilon^2 k} \quad (*)$$

- Proof:
  - Assume k odd (so that median well defined)
  - Consider events E_i: $F(U_i)<1/2-\varepsilon$
  - We have $p=\Pr[E_i]=1/2-\varepsilon$
  - $F(U)<1/2-\varepsilon$ iff at least $k/2$ of these events hold
  - By Chernoff bound, the probability that at least $k/2$ of the events hold is at most $e^{-C\varepsilon^2 k}$
  - Therefore, $\Pr[F(U)< 1/2-\varepsilon]$ is at most $e^{-C\varepsilon^2 k}$
  - The other case can be dealt with in an analogous manner



Closeness in value

- Lemma 2: Let F be c.d.f of a random variable |G|, G drawn from N(0,1).

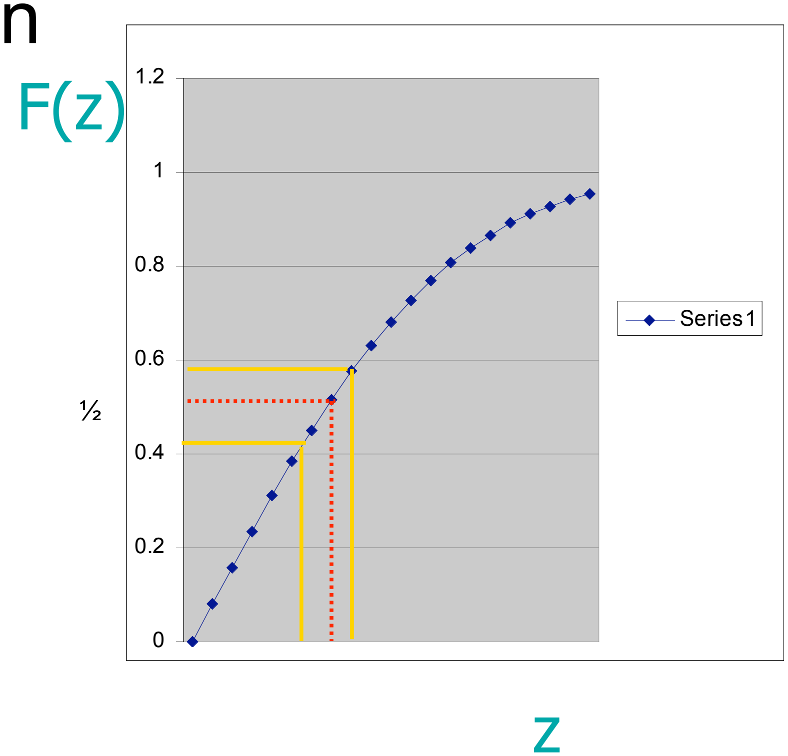
  There exists a C'>0 s.t. if for some z we have

  $$F(z) \in (1/2-\varepsilon, 1/2+\varepsilon)$$

  then

  $$z = \text{median}(g) \pm C' \varepsilon$$

- Proof: Calculus.



Altogether

- Theorem: If we use median estimator

$$Y=\text{median}[\ |Z_1|, \ldots , |Z_k|] \ / \ \text{median}[|g|]$$

(where $Z_j=\sum_i r_{ij}\, x_i$, r_{ij} chosen i.i.d. from N(0,1)), then we have

$$Y = ||x||_2 \,[\ \text{median}(g) \pm C'\, \varepsilon\ ] \ / \ \text{median}[|g|] = ||x||_2\,(1 \pm C''\, \varepsilon)$$

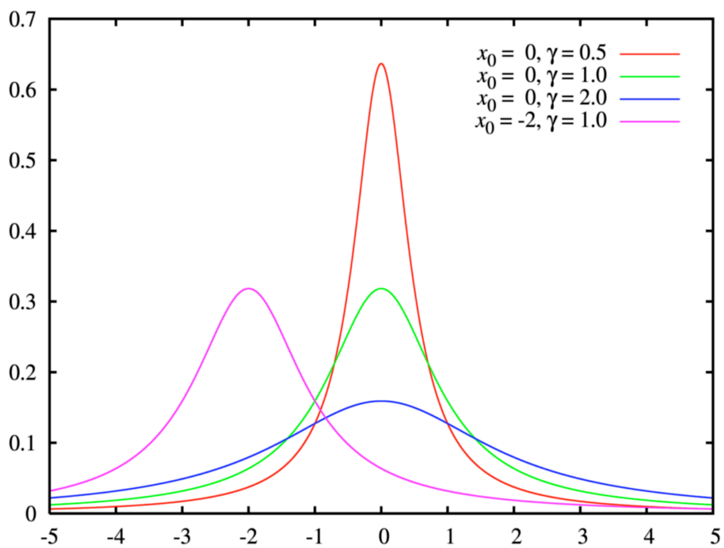
with probability $1\text{-}e^{-C\varepsilon^2 k}$

- How to extend this to $||x||_p$?

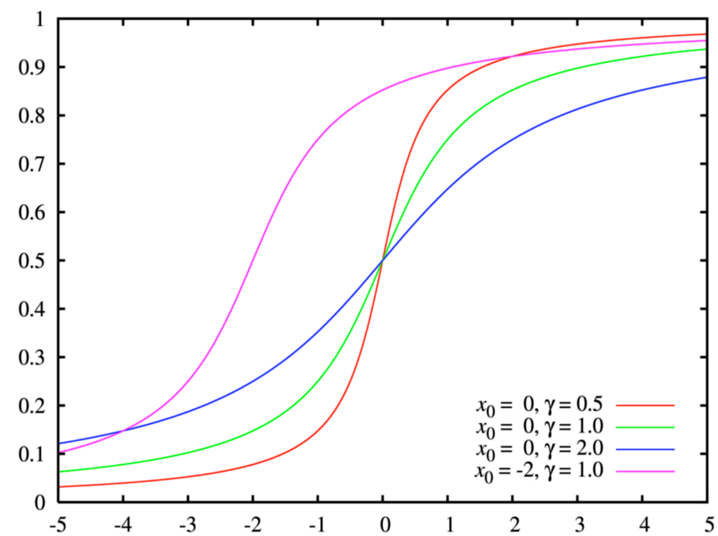
Other norms

- Key property of normal distribution:
  - If $U_1 \ldots U_k$ indep., U normal
  - Then $x_1U_1 + \ldots + x_mU_m$ is distributed as
    $(x_1^p + \ldots + x_m^p)^{1/p}U$, $p=2$
- Such distributions are called "p-stable"
- Good news: p-stable distributions exist for any $p \in (0,2]$
- For example, for $p=1$, we have Cauchy distribution:
  - Density function: $f(x)=1/[\pi(1+x^2)]$
  - C.d.f.: $F(z)=\arctan(z)/\pi+1/2$
  - 1-stability: $x_1U_1 + \ldots + x_mU_m$ is distributed as $(|x_1|+\ldots+|x_m|)U$

Cauchy (from Wiki)



Cauchy density functions



Cauchy c.d.f.'s

- The median estimator arguments go through
- Can generate random Cauchy by choosing a random $u \in [0,1]$ and computing $F^{-1}(u)$

p-stability for p≠1, 2 , 1/2

- Basically, it is a mess
  - No closed form formula for density/c.d.f.
  - Not clear where the median is
  - Not clear what the derivative of c.d.f. around the median is
- Nevertheless
  - Can generate random variables
  - Moments are known (more or less)
  - Given samples of a*|g| , g p-stable, can estimate a up to 1±ε [Indyk, JACM'06; Ping Li, SODA'08]
    (using various hacks and/or moments)
- For more info on p-stable distributions, see:

  V.V. Uchaikin, V.M. Zolotarev,
  Chance and Stability. Stable Distributions and their Applications.
  http://staff.ulsu.ru/uchaikin/uchzol.pdf

Summary

- Maintaining L_p norm of x under updates
  - Polylogarithmic space for $p \leq 2$
- Issues ignored:
  - Randomness
  - Discretization (but everything can be done using $O(\log (m+n))$ bit numbers)

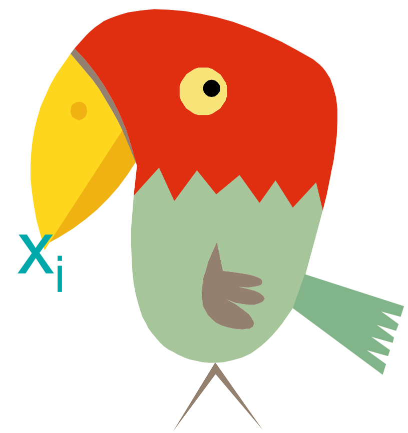
L_k norm, $k \geq 2$

L_k norm

- Algorithm for estimating L_k norm of a stream
  - A stream of elements $i_1 \ldots i_n$
  - Each i can be interpreted as $x_i = x_i + 1$ (only positive updates)
  - Space: $O(m^{1-1/k} / \varepsilon^2)$ for $(1 \pm \varepsilon)$-approximation with const. probability
  - Sampling, not sketching

L_k Norm Estimation: AMS'96

- Useful notion: $F_k = \sum_{i=1}^m x_i^k = ||x||_k^k$
(frequency moment of the stream $i_1 \ldots i_n$)
- Algorithm A: two passes
  - Pass 1: Pick a stream element $i=i_j$ uniformly at random
  - Pass 2: Compute x_i
  - Return $Y=n\ x_i^{k-1}$
- Alternative view:
  - Little birdy that samples i and returns x_i (Sublinear-Time Algorithms class)



Analysis

- Estimator $Y=n\, x_i^{k-1}$
- Expectation

$$E[Y]= \sum_i x_i/n * n x_i^{k-1} = \sum_i x_i^k \;=F_k$$

- Second moment (≥variance)

$$E[Y^2]= \sum_i x_i/n * n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = n\, F_{2k-1}$$

- Claim:

$$n\, F_{2k-1} \le m^{1-1/k}\, (F_k)^2$$

- Therefore, averaging over $O(m^{1-1/k}/\varepsilon^2)$ samples + Chebyshev does the job (Lecture 2)

Claim

- Claim: $n F_{2k-1} \le m^{1-1/k} (F_k)^2$
- Proof:

$$
\begin{aligned}
& n F_{2k-1} \\
= \; & n \|x\|_{2k-1}^{2k-1} \\
\le \; & n \|x\|_k^{2k-1} \\
= \; & \|x\|_1 \|x\|_k^{2k-1} \\
\le \; & m^{1-1/k} \|x\|_k \|x\|_k^{2k-1} \\
= \; & m^{1-1/k} \|x\|_k^{2k} \\
= \; & m^{1-1/k} F_k^2
\end{aligned}
$$

One Pass

- Cannot compute x_i exactly
- Instead:
  - Pick $i=i_j$ uniformly at random from the stream
  - Compute r=#occurrences of i in $i_j \ldots i_n$
  - Use r instead of x_i
  - Clearly $r \le x_i$
  - ..but $E[r]=(x_i+1)/2$, so things should work out up to constant factor (depending on k)
- Even better idea: use estimator

$$Y' = n\,(r^k - (r-1)^k)$$

Analysis

- Expectation:
  $E[Y'] = n\, E[(r^k - (r-1)^k)]$
  $= n * 1/n \sum_i \sum_{j=1}^{x_i} [j^k - (j-1)^k]$
  $= \sum_i x_i^k$
- Second moment:
  - Observe that $Y' = n\,(r^k - (r-1)^k) \le n\,k\,r^{k-1} \le k\,Y$
  - Therefore $Var[Y'] \le E[Y']^2 \le k^2\, E[Y]^2 \le k^2\, m^{1-1/k}\, F_k^2$ (can improve to $k\, m^{1-1/k}\, F_k^2$ for integer k)
- Altogether:
  - One pass algorithm for F_k (positive updates)
  - Space: $O(km^{1-1/k}/\varepsilon^2)$ for $(1\pm\varepsilon)$-approximation

Notes

- The analysis in AMS'96, as is, works only for integer k

  (but is easy to adapt to any $k>1$)
- The analysis* in these notes is somewhat simpler (but yields $k^2\ m^{1-1/k}$ space)

* Contributed by David Woodruff

Summary

- Can $(1\pm\varepsilon)$-approximate L_k norm of a stream (insertions-only) in $O(m^{1-1/k}/\varepsilon^2)$ space
- Sampling - quite general
  - Entropy, i.e., $\sum_i x_i/n \log(x_i/n)$ in polylog n space
  - Other stuff