Compressed Counting

Maximally-Skewed Stable Random Projections

Ping Li

Department of Statistical Science

Faculty of Computing and Information Science

Cornell University

June, 2008

What is Counting? Why Should We Care?

Counting is just counting! Given D items, x_1, x_2, ..., x_D, we can count

- The sum $\sum_{i=1}^D x_i$. The number of non-zeros, $\sum_{i=1}^D 1_{x_i\neq 0}$
- The αth moment $F_{(\alpha)} = \sum_{i=1}^D x_i^\alpha$
  $F_{(1)}$ =the sum, $F_{(2)}$ = the power/energy, $F_{(0)}$ = number of non-zeros.
- The future fortune, $\sum_{i=1}^D x_i^{1\pm\Delta}$, Δ = interest/decay rate (usually small)
- The entropy moment $\sum_{i=1}^D x_i \log x_i$ and entropy $\sum_{i=1}^D \frac{x_i}{F_{(1)}} \log \frac{x_i}{F_{(1)}}$
- The Tsallis Entropy $\frac{1-F_{(\alpha)}/F_{(1)}^\alpha}{\alpha-1}$ The Rényi Entropy $\frac{1}{1-\alpha}\log\frac{F_{(\alpha)}}{F_{(1)}^\alpha}$

Isn't Counting a Simple (Trivial) Task?

Partially True!, if data are static. However

Real-world data are in general Massive and Dynamic —— Data Streams

- Databases in Amazon, Ebay, Walmart, and search engines
- Internet/telephone traffic, high-way traffic
- Finance (stock) data
- ...

For example, the Turnstile data stream model for an online bookstore t=0

0	0	0	0	0	0	....	0
IP 1	IP 2	IP 3	IP 4			....	IP D

t=1 arriving stream = (3, 10) user 3 ordered 10 books

0	0	10	0	0	0	....	0
IP 1	IP 2	IP 3	IP 4			....	IP D

t=2 arriving stream = (1, 5) user 1 ordered 5 books

5	0	10	0	0	0	....	0
IP 1	IP 2	IP 3	IP 4			....	IP D

t=3 arriving stream = (3, −8) user 3 cancelled 8 books

5	0	2	0	0	0	....	0
IP 1	IP 2	IP 3	IP 4			....	IP D

Turnstile Data Stream Model

At time t, an incoming element : $\boxed{a_t = (i_t, I_t)}$

$i_t \in [1, D]$ index, I_t: increment/decrement.

Updating rule : $\boxed{A_t[i_t] = A_{t-1}[i_t] + I_t}$

Goal : Count $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$

Counting: Trivial if $\alpha = 1$, but Non-trivial in General

Goal : Count $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$, where $\boxed{A_t[i_t] = A_{t-1}[i_t] + I_t}$.

When $\alpha \neq 1$, counting $F_{(\alpha)}$ exactly requires D counters. (but D can be 2^{64})

When $\alpha = 1$, however, counting the sum is trivial, using a simple counter.

$$F_{(1)} = \sum_{i=1}^{D} A_t[i] = \sum_{s=1}^{t} I_s,$$

The Intuition for $\alpha \approx 1$

There might exist an intelligent counting system which works like a simple counter when α is close 1; and its complexity is a function of how close α is to 1.

Our answer: Yes!

Two caveats:

(1) What if data are negative? Shouldn't we define $F_{(\alpha)} = \sum_{i=1}^{D} |A_t[i]|^{\alpha}$?

(2) Why the case $\alpha \approx 1$ is important ?

The Non-Negativity Constraint

"God created the natural numbers; all the rest is the work of man."

—- by German mathematician Leopold Kronecker (1823 - 1891)

Turnstile model, $a_t = (i_t, I_t), \quad A_t[i_t] = A_{t-1}[i_t] + I_t,$

$I_t > 0$: increment, insertion, eg place orders

$I_t < 0$: decrement, deletion, eg cancel orders,

This talk: Strict Turnstile model $A_t[i] \geq 0$, always.

One can only cancel an order if she/he did place the order!!

Suffices for almost all applications.

Sample Applications of αth Moments (Especially $\alpha \approx 1$)

1. $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$ itself is a useful summary statistic
   e.g., Rényi entropy, Tsallis entropy, are functions of $F_{(\alpha)}$.

2. Statistical modeling and inference of parameters using method of moments

3. $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$ is a fundamental building element for other algorithms
   Eg., estimating Shannon entropy of data streams

Estimate Shannon Entropy of Data Streams

Definition of Shannon Entropy

$$H = -\sum_{i=1}^{D} \frac{A_t[i]}{F_{(1)}} \log \frac{A_t[i]}{F_{(1)}}, \qquad F_{(1)} = \sum_{i=1}^{D} A_t[i]$$

Many papers/algorithms in theoretical CS and databases on estimating entropy.

Three Examples (all used α moments with $\alpha \to 1$)

- Difference of Two Moments (Zhao, et. al., 2007)

$$\lim_{\Delta\to 0}\frac{x^{1+\Delta}-x^{1-\Delta}}{2\Delta} = x\log(x), \qquad (\alpha = 1\pm\Delta),$$

$$\lim_{\Delta\to 0}\frac{1}{2\Delta}\left(\sum_{i=1}^{D}A_t[i]^{1+\Delta}-\sum_{i=1}^{D}A_t[i]^{1-\Delta}\right)\rightarrow\sum_{i=1}^{D}A_t[i]\log A_t[i].$$

- Rényi Entropy (Harvey, et. al., FOCS'08)

$$H_\alpha = \frac{1}{1-\alpha}\log\frac{F_{(\alpha)}}{F_{(1)}^{\alpha}} \rightarrow H, \qquad \text{as } \alpha\rightarrow 1$$

- Tsallis Entropy (Harvey, et. al., FOCS'08)

$$T_\alpha = \frac{1}{\alpha-1}\left(1-\frac{F_{(\alpha)}}{F_{(1)}^{\alpha}}\right) \rightarrow H, \qquad \text{as } \alpha\rightarrow 1$$

Rényi entropy and Tsallis entropy are themselves useful, e.g., in physics

Our Technique: Skewed Stable Random Projections

Original data stream signal: $A_t[i]$, $i = 1$ to D. eg $D = 2^{64}$

Projected signal: $X_t = A_t \times \mathbf{R} \in \mathbb{R}^k$, k is small (eg $k = 50 \sim 100$)

Projection matrix: $\mathbf{R} \in \mathbb{R}^{D\times k}$, entries are random

This talk : Skewed projections

Sample entries of $\mathbf{R}$ i.i.d. from a skewed stable distribution.

——————————————————————-

Previous classical work: symmetric stable random projections (Indyk, JACM 2006)

Sample $\mathbf{R}$ from a symmetric stable distribution.

Incremental Projection

Linear Projection: $X_t = A_t \times \mathbf{R}$

$+$

Linear data model: $A_t[i_t] = A_{t-1}[i_t] + I_t$

$\Longrightarrow$

Conduct $\boxed{X_t = A_t \times \mathbf{R}}$ incrementally.

Generate entries of $\mathbf{R}$ on-demand

Recover $F_{(\alpha)}$ from Projected Data

$$X_t = (x_1, x_2, ..., x_k) = A_t \times \mathbf{R}$$

$$\mathbf{R} = \{r_{ij}\} \in \mathbb{R}^{D\times k}, \ \ r_{ij} \sim S\left(\alpha, \beta, 1\right)$$

$S\left(\alpha, \beta, \gamma\right)$: α-stable, β-skewed distribution with scale γ

Then, by stability, at any t, x_j's are i.i.d. stable samples

$$x_j \sim S\left(\alpha, \beta, F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha\right)$$

$\Longrightarrow$ A statistical estimation problem.

Review of Skewed Stable Distributions

Z follows a β-skewed α-stable distribution if Fourier transform of its density

$$\mathscr{F}_Z(t) = \text{E}\exp\left(\sqrt{-1}Zt\right) \qquad \alpha \neq 1,$$
$$= \exp\left(-F|t|^\alpha\left(1-\sqrt{-1}\beta\text{sign}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right)\right),$$

$0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$. The scale $F > 0$. $Z \sim S(\alpha, \beta, F)$

If $Z_1, Z_2 \sim S(\alpha, \beta, 1)$, independent, then for any $C_1 \geq 0$, $C_2 \geq 0$,

$$Z = C_1 Z_1 + C_2 Z_2 \sim S\left(\alpha, \beta, F = C_1^\alpha + C_2^\alpha\right).$$

If C_1 and C_2 do not have the same signs, the "stability" does not hold.

Let $Z = C_1 Z_1 - C_2 Z_2$, with $C_1 \geq 0$ and $C_2 \geq 0$.

Because $\mathscr{F}_{-Z_2}(t) = \mathscr{F}_{Z_2}(-t)$,

$$\mathscr{F}_Z(t) = \exp\left(-|C_1 t|^\alpha \left(1 - \sqrt{-1}\beta \text{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right)\right)$$
$$\times \exp\left(-|C_2 t|^\alpha \left(1 + \sqrt{-1}\beta \text{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right)\right),$$

Does NOT represent a stable law, unless $\beta = 0$ or $\alpha = 2, 0+$.

Symmetric ($\beta = 0$) projections work for any data,

but if data are non-negative, benefits of skewed projection are enormous.

The Statistical Estimation Problem

Task : Given k i.i.d. samples $x_j \sim S\left(\alpha, \beta, F_{(\alpha)}\right)$, estimate $F_{(\alpha)}$.

- No closed-form density in general, but closed-form moments exit.
- A Geometric Mean estimator based on positive moments.
- A Harmonic Mean estimator based on negative moments.
- Both estimators exhibit exponential error (tail) bounds.

The Moment Formula

Lemma 1 *If* $Z \sim S(\alpha, \beta, F_{(\alpha)})$*, then for any* $\boxed{-1 < \lambda < \alpha}$,

$$\boldsymbol{E}\left(|Z|^{\lambda}\right) = F_{(\alpha)}^{\lambda/\alpha} \cos\left(\frac{\lambda}{\alpha}\tan^{-1}\left(\beta\tan\left(\frac{\alpha\pi}{2}\right)\right)\right)$$

$$\times\left(1+\beta^2\tan^2\left(\frac{\alpha\pi}{2}\right)\right)^{\frac{\lambda}{2\alpha}}\left(\frac{2}{\pi}\sin\left(\frac{\pi}{2}\lambda\right)\Gamma\left(1-\frac{\lambda}{\alpha}\right)\Gamma\left(\lambda\right)\right),$$

Proof: ArXiv report, “Compressed Counting” Feb 2008.

Partial proof can be found at Zolotarev (1986), Hardin (1984).

$\boxed{\lambda = \frac{\alpha}{k}}$ $\Longrightarrow$ an unbiased geometric mean estimator.

Nice things happen when $\beta = 1$.

Lemma 2 *When* $\beta = 1$, *then, for* $\alpha < 1$ *and* $\boxed{-\infty < \lambda < \alpha}$,

$$\textbf{E}\left(|Z|^{\lambda}\right) = \textbf{E}\left(Z^{\lambda}\right) = F_{(\alpha)}^{\lambda/\alpha}\frac{\Gamma\left(1-\frac{\lambda}{\alpha}\right)}{\cos^{\lambda/\alpha}\left(\frac{\alpha\pi}{2}\right)\Gamma\left(1-\lambda\right)}.$$

Nice consequence :

Estimators using negative moments will have infinite moments.

The Geometric Mean Estimator for all β

$$X_t = (x_1, x_2, ..., x_k) = A_t \times \mathbf{R}$$

$$\hat{F}_{(\alpha),gm,\beta} = \frac{\prod_{j=1}^{k} |x_j|^{\alpha/k}}{D_{gm,\beta}},$$

$$D_{gm,\beta} = \cos^k\left(\frac{1}{k}\tan^{-1}\left(\beta\tan\left(\frac{\alpha\pi}{2}\right)\right)\right) \times$$

$$\left(1+\beta^2\tan^2\left(\frac{\alpha\pi}{2}\right)\right)^{\frac{1}{2}}\left[\frac{2}{\pi}\sin\left(\frac{\pi\alpha}{2k}\right)\Gamma\left(1-\frac{1}{k}\right)\Gamma\left(\frac{\alpha}{k}\right)\right]^k.$$

Which β ? : Variance of $\hat{F}_{(\alpha),gm,\beta}$ is decreasing in $\beta \in [0,1]$.

$$\text{Var}\left(\hat{F}_{(\alpha),gm,\beta}\right) = F_{(\alpha)}^2 V_{gm,\beta}$$

$$V_{gm,\beta} = \left[2 - \sec^2\left(\frac{1}{k}\tan^{-1}\left(\beta\tan\left(\frac{\alpha\pi}{2}\right)\right)\right)\right]^k$$
$$\times \frac{\left[\frac{2}{\pi}\sin\left(\frac{\pi\alpha}{k}\right)\Gamma\left(1-\frac{2}{k}\right)\Gamma\left(\frac{2\alpha}{k}\right)\right]^k}{\left[\frac{2}{\pi}\sin\left(\frac{\pi\alpha}{2k}\right)\Gamma\left(1-\frac{1}{k}\right)\Gamma\left(\frac{\alpha}{k}\right)\right]^{2k}} - 1,$$

A decreasing function of $\beta \in [0,1]$. $\Longrightarrow$ Use $\beta = 1$, maximally skewed

The Geometric Mean Estimator for $\beta = 1$

$$\hat{F}_{(\alpha),gm} = \frac{\prod_{j=1}^{k} |x_j|^{\alpha/k}}{D_{gm}}$$

Lemma 3

$$\text{Var}\left(\hat{F}_{(\alpha),gm}\right) = \begin{cases} \frac{F_{(\alpha)}^2}{k}\frac{\pi^2}{6}\left(1-\alpha^2\right) + O\left(\frac{1}{k^2}\right), & \text{if } \alpha < 1 \\ \frac{F_{(\alpha)}^2}{k}\frac{\pi^2}{6}\left(\alpha-1\right)\left(5-\alpha\right) + O\left(\frac{1}{k^2}\right), & \text{if } \alpha > 1 \end{cases}$$

As $\alpha \to 1$, the asymptotic variance $\to 0$.

A Geometric Mean Estimator for Symmetric Projections $\beta = 0$

(Li, SODA'08)

Symmetric projections, ie $r_{ij} \sim S(\alpha, \beta = 0, 1)$.

Projected data: $x_j \sim S\left(\alpha, \beta = 0, F_{(\alpha)}\right),\ j = 1$ to k.

Geometric mean estimator (later used by Harvey et. al. FOCS'08):

$$\hat{F}_{(\alpha),gm,sym} = \frac{\prod_{j=1}^{k} |x_j|^{\alpha/k}}{D_{gm,sym}}$$

$$\text{Var}\left(\hat{F}_{(\alpha),gm,sym}\right) = \frac{F_{(\alpha)}^2}{k}\frac{\pi^2}{12}\left(2+\alpha^2\right) + O\left(\frac{1}{k^2}\right),$$

As $\alpha \to 1$, using skewed projections achieves an "infinite improvement".

A Better Estimator Using Harmonic Mean, for $\alpha < 1$

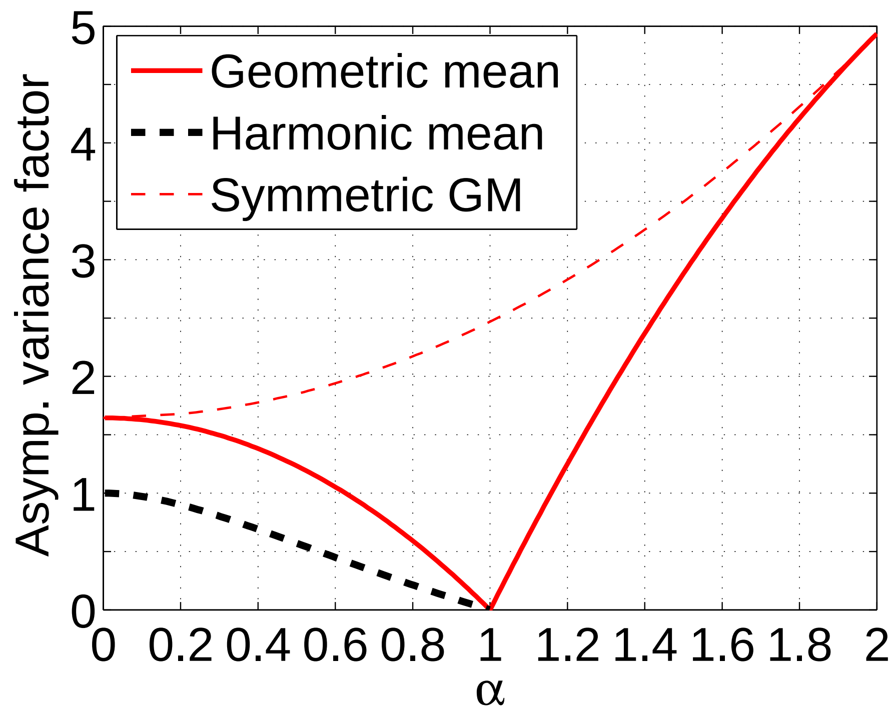
Skewed Projections $(\beta = 1)$

$$\hat{F}_{(\alpha),hm} = \frac{k\frac{\cos\left(\frac{\alpha\pi}{2}\right)}{\Gamma(1+\alpha)}}{\sum_{j=1}^{k}|x_j|^{-\alpha}}\left(1-\frac{1}{k}\left(\frac{2\Gamma^2(1+\alpha)}{\Gamma(1+2\alpha)}-1\right)\right).$$

Advantages of $\hat{F}_{(\alpha),hm}$

- Smaller variance
- Smaller tail bound constant
- Moment generating function exits.

Comparing Asymptotic Variances



Now What?

Question 1: Is Compressed Counting (skewed projections) practical?

Answer: Yes, it is as practical as symmetric stable random projections

Question 2: Does Compressed Counting demonstrate improvement on real data?

Answer: Yes, definitely.

Question 3: Precisely, how large k should be?

Answer: $k = O\left(1/\epsilon^2\right)$ for general α, but $k = O\left(1/\epsilon\right)$ only when $\alpha \rightarrow 1$.

The bounds are precisely specified.

Sampling From Maximally-Skewed Stable Distributions

To sample from $Z \sim S(\alpha, \beta = 1, 1)$:

$$W \sim \exp(1) \qquad U \sim \text{Uniform}\left(-\frac{\pi}{2}, \quad \frac{\pi}{2}\right)$$

$$\rho = \begin{cases} \frac{\pi}{2} & \alpha < 1 \\ \frac{\pi}{2}\frac{2-\alpha}{\alpha} & \alpha > 1 \end{cases}$$

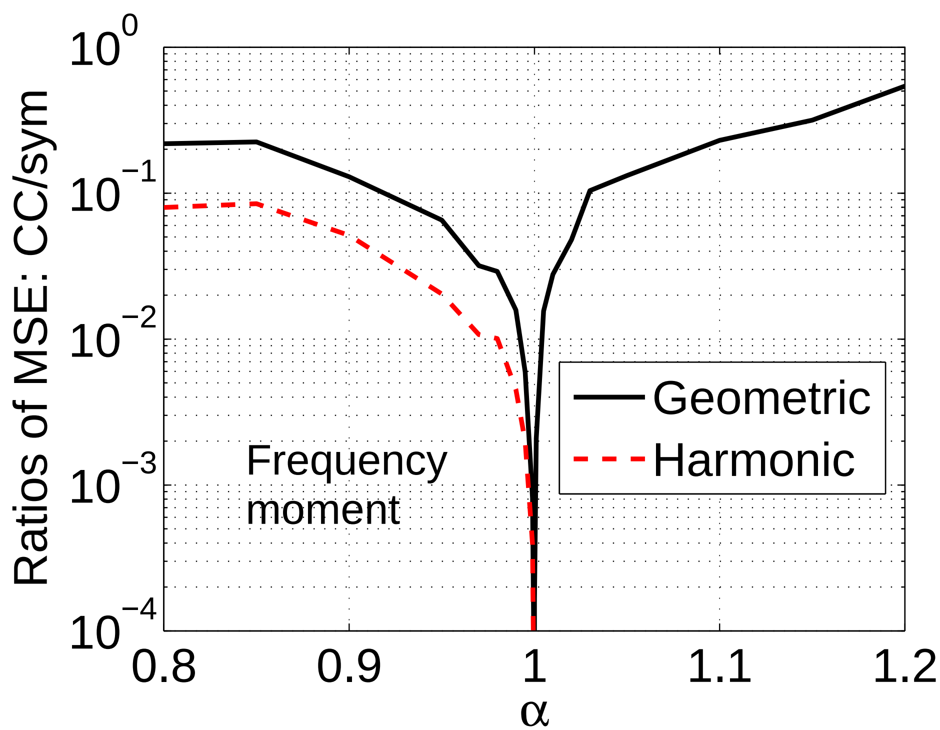
$$Z = \frac{\sin\left(\alpha(U+\rho)\right)}{\left[\cos U \cos\left(\rho\alpha\right)\right]^{1/\alpha}}\left[\frac{\cos\left(U - \alpha(U+\rho)\right)}{W}\right]^{\frac{1-\alpha}{\alpha}} \sim S(\alpha, \beta = 1, 1)$$

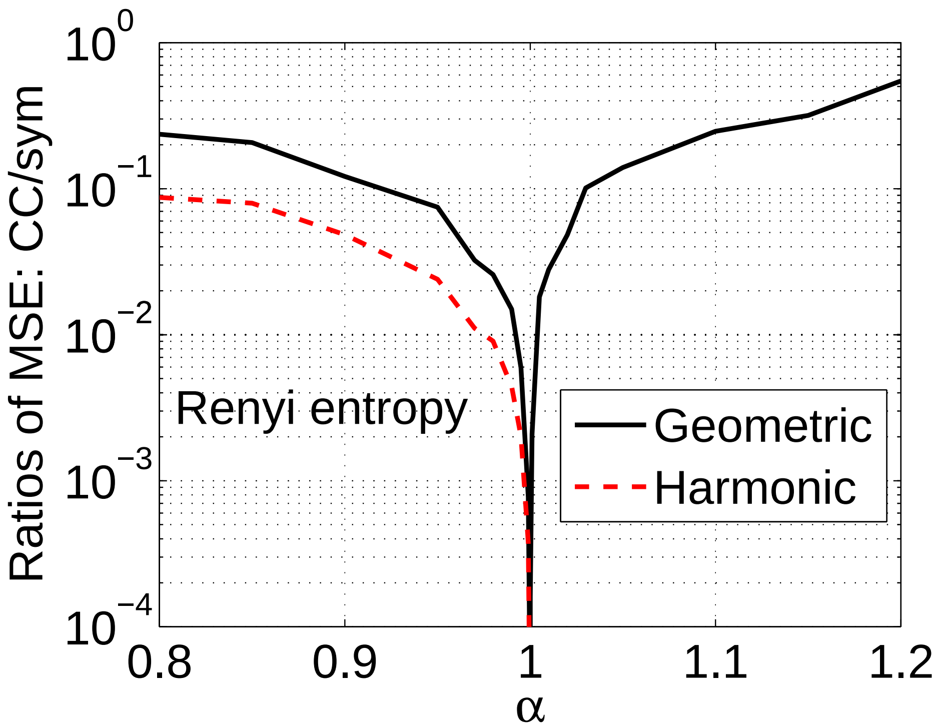
$\cos^{1/\alpha}\left(\rho\alpha\right)$ can be removed and later reflected in the estimators.

Sampling from Skewed distributions is as easy as from symmetric distributions.

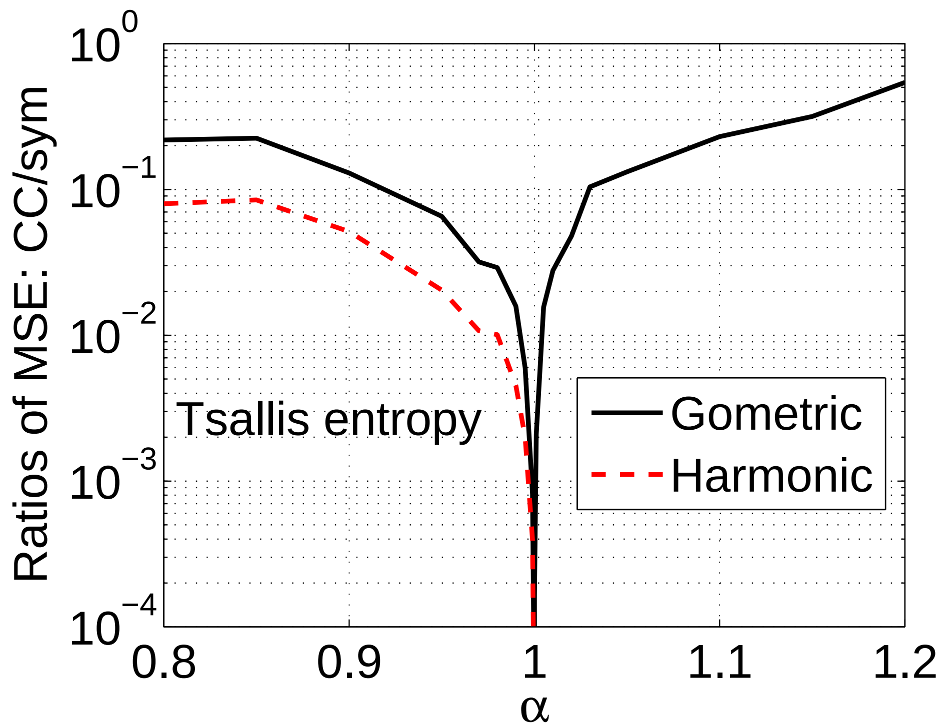
Experiments on Real Data (Word "A")

Comparing mean square errors (MSE): $\frac{\text{Compressed Counting}}{\text{symmetric projection}}$

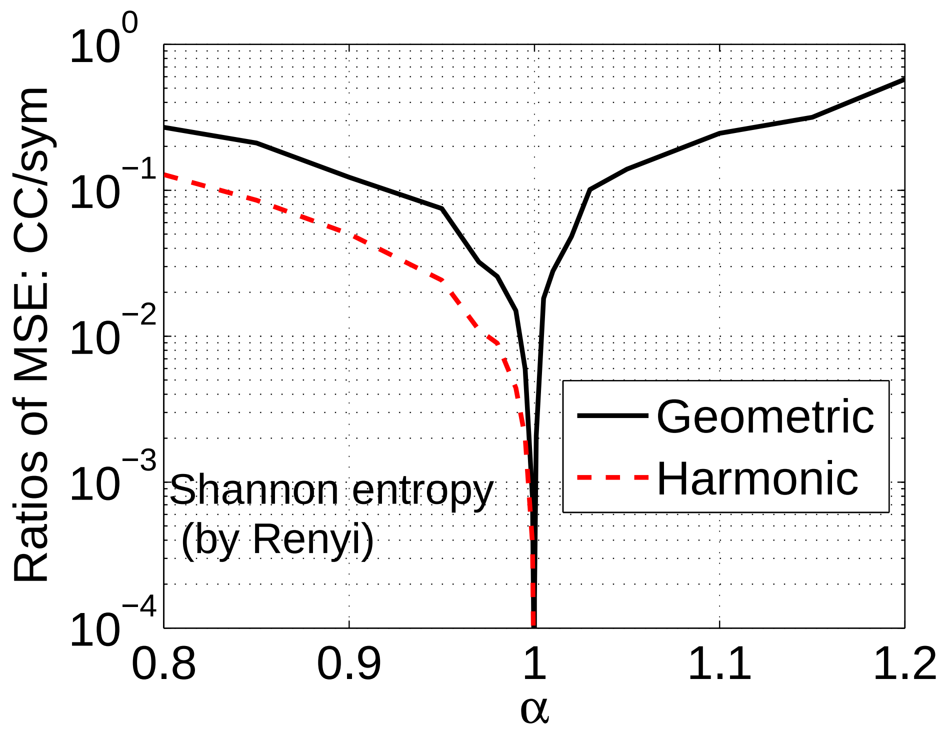




Ratios of MSE: CC/sym
10^0
10^−1
10^−2
10^−3
10^−4
0.8
0.9
1
1.1
1.2
α
Renyi entropy
Geometric
Harmonic


Ratios of MSE: CC/sym
10^0
10^−1
10^−2
10^−3
10^−4
0.8
0.9
1
1.1
1.2
α
Tsallis entropy
Gometric
Harmonic

Estimate Shannon entropy using Rényi entropy



Compressed Counting is practical and highly effective!

Tail Bounds of the Geometric Mean Estimator

Lemma 4

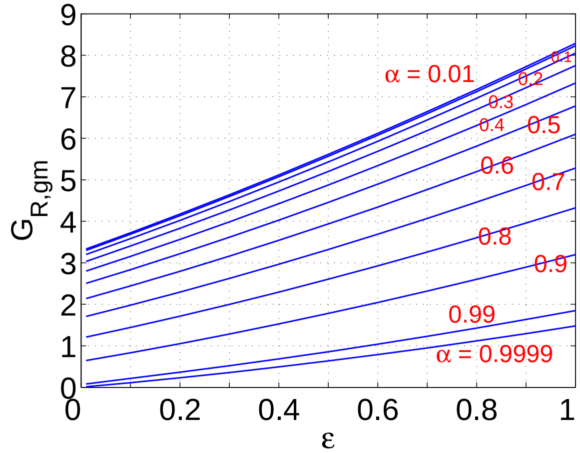
$$\mathbf{Pr}\left(\hat{F}_{(\alpha),gm} - F_{(\alpha)} \geq \epsilon F_{(\alpha)}\right) \leq \exp\left(-k\frac{\epsilon^2}{G_{R,gm}}\right), \ \ \epsilon > 0,$$

$$\mathbf{Pr}\left(\hat{F}_{(\alpha),gm} - F_{(\alpha)} \leq -\epsilon F_{(\alpha)}\right) \leq \exp\left(-k\frac{\epsilon^2}{G_{L,gm}}\right), \ 0 < \epsilon < 1,$$

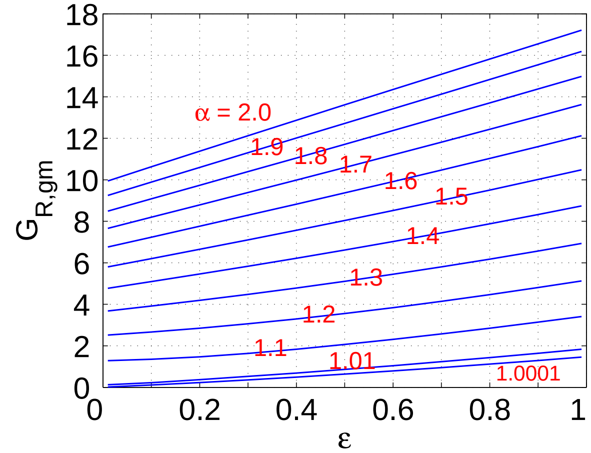
$$\frac{\epsilon^2}{G_{R,gm}} = C_R \log(1+\epsilon) - C_R\gamma_e(\alpha-1)$$
$$-\log\left(\cos\left(\frac{\kappa(\alpha)\pi C_R}{2}\right)\frac{2}{\pi}\Gamma\left(\alpha C_R\right)\Gamma\left(1-C_R\right)\sin\left(\frac{\pi\alpha C_R}{2}\right)\right)$$

C_R *is the solution to to*

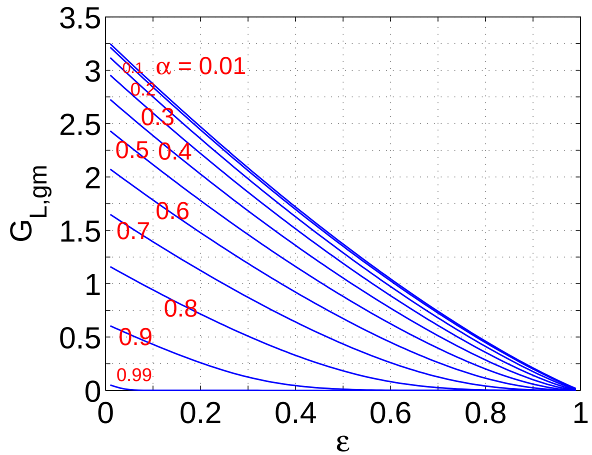
$$-\gamma_e(\alpha-1) + \log(1+\epsilon) + \frac{\kappa(\alpha)\pi}{2}\tan\left(\frac{\kappa(\alpha)\pi}{2}C_R\right)$$
$$-\frac{\alpha\pi/2}{\tan\left(\frac{\alpha\pi}{2}C_R\right)} - \frac{\Gamma'\left(\alpha C_R\right)}{\Gamma\left(\alpha C_R\right)}\alpha + \frac{\Gamma'\left(1-C_R\right)}{\Gamma\left(1-C_R\right)} = 0,$$



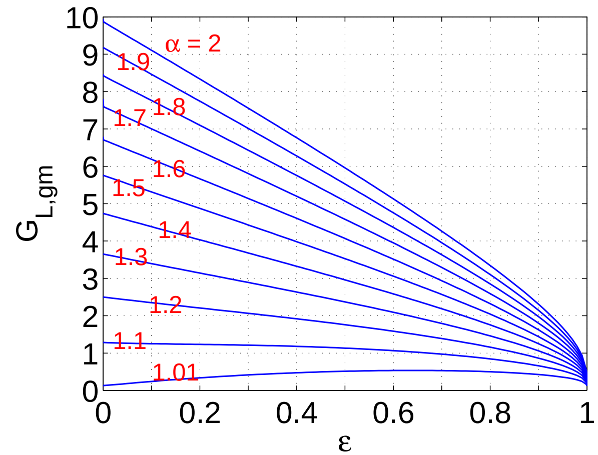
(a) Right bound, $\alpha < 1$



(b) Right bound, $\alpha > 1$



(c) Left bound, $\alpha < 1$



(d) Left bound, $\alpha > 1$

The Sample Complexity Bound

Let $G = \max\{G_{L,gm}, G_{R,gm}\}$.

Bound the error (tail) probability by δ, the level of significance (eg 0.05)

$$\mathbf{Pr}\left(|\hat{F}_{(\alpha),gm} - F_{(\alpha)}| \geq \epsilon F_{(\alpha)}\right) \leq 2\exp\left(-k\frac{\epsilon^2}{G}\right) \leq \delta$$

$$\Longrightarrow k \geq \frac{G}{\epsilon^2}\log\frac{2}{\delta}$$

Sample Complexity Bound (large-deviation bound):

If $k \geq \frac{G}{\epsilon^2}\log\frac{2}{\delta}$, then with probability at least $1-\delta$, $F_{(\alpha)}$ can be approximated within a factor of $1 \pm \epsilon$.

The $O\left(1/\epsilon^2\right)$ bound in general can not be improved — Central Limit Theorem

The Sample Complexity for $\alpha = 1 \pm \Delta$

Lemma 5 *For fixed* ϵ*, as* $\alpha \rightarrow 1$ *(i.e.,* $\Delta \rightarrow 0$*),*

$$G_{R,gm} = \frac{\epsilon^2}{\log(1+\epsilon) - 2\sqrt{\Delta \log(1+\epsilon)} + o\left(\sqrt{\Delta}\right)} = O(\epsilon)$$

If $\alpha > 1,$ *then*

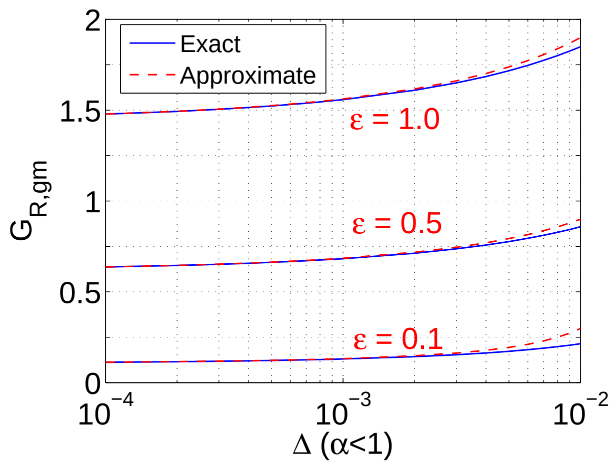
$$G_{L,gm} = \frac{\epsilon^2}{-\log(1-\epsilon) - 2\sqrt{-2\Delta \log(1-\epsilon)} + o\left(\sqrt{\Delta}\right)} = O(\epsilon)$$

If $\alpha < 1,$ *then*

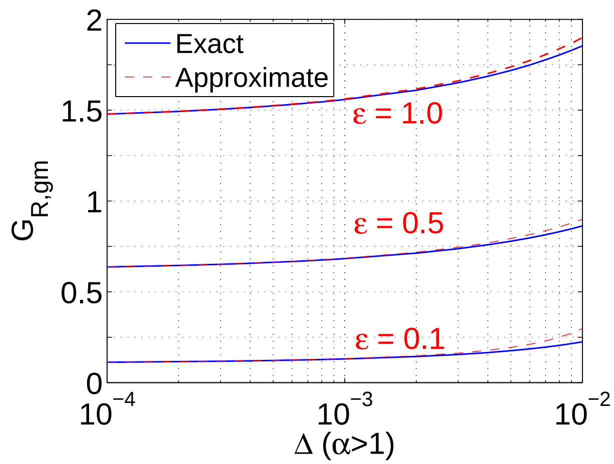
$$G_{L,gm} = \frac{\epsilon^2}{\Delta\left(\exp\left(\frac{-\log(1-\epsilon)}{\Delta} - 1 - \gamma_e\right)\right) + o\left(\Delta \exp\left(\frac{1}{\Delta}\right)\right)} = O\left(\epsilon \exp\left(-\frac{\epsilon}{\Delta}\right)\right)$$

For α close to 1, sample complexity is $O(1/\epsilon)$ not $O\left(1/\epsilon^2\right)$.

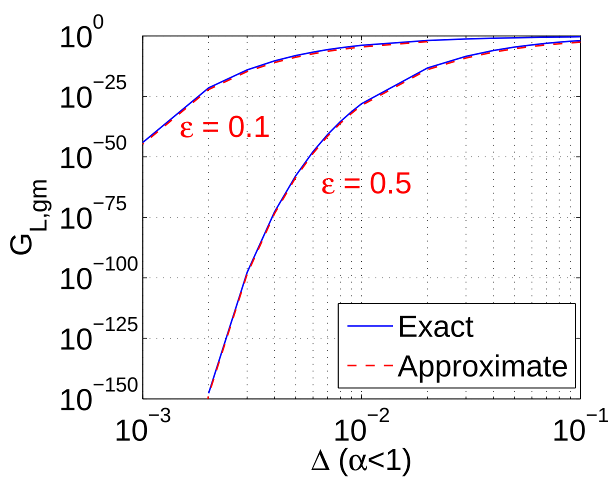
Not violating fundamental principles.



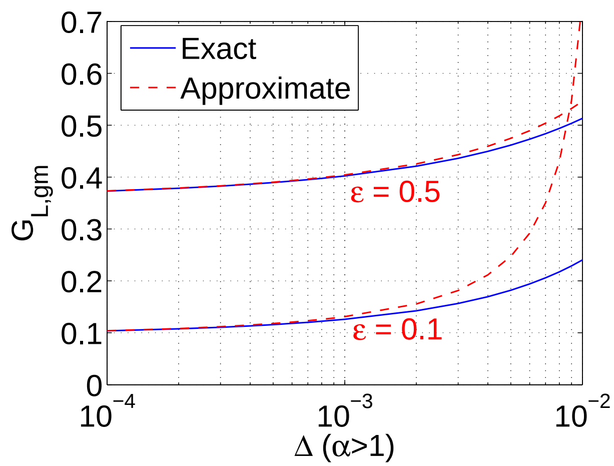
(e) Right bound, $\alpha < 1$



(f) Right bound, $\alpha > 1$



(g) Left bound, $\alpha < 1$



(h) Left bound, $\alpha > 1$

Applications in Method of Moments

For example, z_i, $i = 1$ to D are collected from data streams. z_i's follow a generalized gamma distribution $z_i \sim GG(\theta_1, \theta_2, \theta_3)$:

$$\mathsf{E}(z_i) = \theta_1\theta_2, \qquad \mathsf{Var}(z) = \theta_1\theta_2^2, \qquad \mathsf{E}\left(z - \mathsf{E}(z)\right)^3 = (\theta_3 + 1)\theta_1\theta_2^3$$

Estimate θ_1, θ_2, θ_3 using

- First three moments ($\alpha = 1, 2, 3$) $\Longrightarrow$ Computationally very expensive
- Fractional moments (eg. $\alpha = 0.95, 1.05, 1$) $\Longrightarrow$ Computationally cheap

Will this affect estimation accuracy? Not really, because D is large!

A Simple Example with One Parameter

Suppose $z_i \sim Gamma(\theta, 1)$. The data z_i's are collected from data streams.

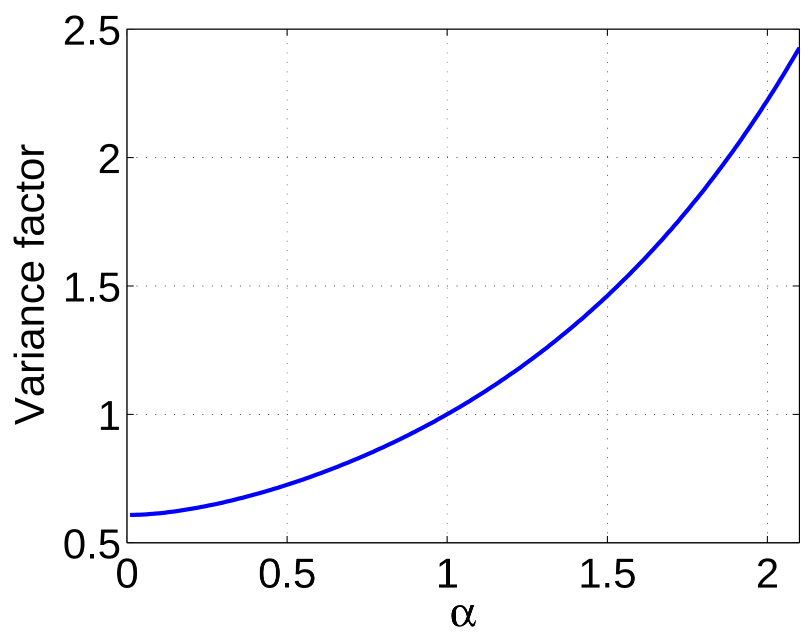
Estimate θ by αth moment: $\boxed{\text{E}(z_i^\alpha) = \Gamma(\alpha+\theta)/\Gamma(\theta)}$.

Solve for $\hat{\theta}$ from the moment equation:

$$\frac{\Gamma(\alpha+\hat{\theta})}{\Gamma(\hat{\theta})} = \frac{1}{D}\sum_{i=1}^D z_i^\alpha$$

$$\text{Var}\left(\hat{\theta}\right) \approx \frac{1}{D}\left(\frac{\Gamma(2\alpha+\theta)\Gamma(\theta)}{\Gamma^2(\alpha+\theta)} - 1\right)\frac{1}{\left(\frac{\Gamma'(\alpha+\theta)}{\Gamma(\alpha+\theta)} - \frac{\Gamma'(\theta)}{\Gamma(\theta)}\right)^2}$$

$$\text{Var}(\hat{\theta})|_{\alpha=0} \approx \frac{0.608}{D}, \qquad \text{Var}(\hat{\theta})|_{\alpha=1} \approx \frac{1}{D},$$



Trade-off:

$\alpha = 1$, higher variance, fewer counters

$\alpha = 0$, smaller variance, more counters

Since D is very large, the difference between $\frac{0.608}{D}$ and $\frac{1}{D}$ may not matter.

Summary

Goal: Efficiently count the αth moment $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$.
Since A_t is dynamic, an exact answer requires D counters

Intuition: An intelligent counting system should resemble a simple counter for α close 1, with complexity varying continuously as a function of how close α is to 1.

Compressed Counting (CC) is such an intelligent counting system, based on maximally-skewed α-stable random projections. $\Longrightarrow$ a statistical estimation task.

Estimators: The geometric mean and harmonic mean estimators. Sample complexity = $O\left(1/\epsilon\right)$ for α close to 1, instead of the usual $O\left(1/\epsilon^2\right)$ bound.

Applications:

1. $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$ itself is a useful summary statistic,
   e.g., the sum in the future (interest/decay), Rényi entropy, Tsallis entropy.

2. Statistical modeling and inference of parameters using method of moments

3. $F_{(\alpha)} = \sum_{i=1}^{D} A_t[i]^\alpha$ is a fundamental building block for other algorithms
   e.g., estimating entropy of data streams

Limitation: CC can not be used for estimating pairwise distances!!

Thank you!