Differentially Private Counting with Minimal Space: $\mathbb{F}_p$ Sketch Preserves Differential Privacy As Is

Lun Wang, Dawn Song
University of California, Berkeley
wanglun@berkeley.edu, dawsong@cs.berkeley.edu

March 27, 2021

1 Introduction

Counting is one of the most fundamental operations in almost every area of computer science. It typically refers to estimating the cardinality (the 0^{th} frequency moment) of a given set. However, counting can actually refer to the process of estimating a broader class of statistics namely p^{th} frequency moment, denoted $\mathbb{F}_p$. Frequency moments estimation is at the core of various important statistical problems. $\mathbb{F}_1$ is used for data mining [6] and hypothesis tests [15]. $\mathbb{F}_2$ has applications to calculating Gini index [21, 10] and surprise index [11], training random forests [3], numerical linear algebra [5, 24] and network anomaly detection [19, 25]. Fractional frequency moments with small p is used in Shannon entropy estimation [12, 28] and image decomposition [9].

On the other hand, the data being counted sometimes contains sensitive information. For example, to calculate Gini index, the data has to include pairs of ID and income. Thus, frequency moments of such data, if published, might also leak sensitive information. To minimize the leakage, we would like to design an estimation mechanism following the gold standard of differential privacy.

Non-private frequency moments estimation is systematically studied under the data streaming model [1, 4, 25, 8, 14, 20, 18, 22, 23, 17].

In this work, we aim to design a mechanism that maintains an estimation of $\mathbb{F}_p$ with differential privacy, while minimizing the space and update complexity. We fail to find any prior work in the rigorous sense but find that several algorithms [2, 16, 26, 13] with totally different goals (*e.g.* anonymized histogram publishing or adversarial robustness) happen to be able to tackle the same problem. Not surprisingly, since they are not specifically designed for frequency moments estimation, the space complexity is exponentially worse than the non-private version. We narrow the gap by directly applying $\mathbb{F}_p$ sketch and showing that it preserves differential privacy as is.

Problem Formulation. We use bold lowercase letters to denote vectors (*e.g.* $\mathbf{a}, \mathbf{b}, \mathbf{c}$) and bold uppercase letters to denote matrix (*e.g.* $\mathbf{A}, \mathbf{B}, \mathbf{C}$). We denote $\{1, \cdots, n\}$ by $[n]$. Let $\mathcal{S} = \{(k_1, v_1), \cdots, (k_n, v_n)\}$ be a stream of key-value pairs where $k_i \in [m], v_i \in [M]$. We would like to design a randomized mechanism $\mathcal{M}$ that correctly estimates the p^{th} frequency moment: $F_p(\mathcal{S}) = \sum_{k=1}^{m} |\sum_{k_i=k} v_i|^p$ for $p \in (0, 2]$ with distortion $(1+\gamma)$ and probability $1-\eta$.

$$(1-\gamma)F_p(\mathcal{S}) \leq \mathcal{M}(\mathcal{S}) \leq (1+\gamma)F_p(\mathcal{S}) \quad w.p. \quad 1-\eta$$

As the data stream might contain sensitive information, $\mathcal{M}$ should preserve (ϵ, δ) differential privacy. In our setting, neighboring data streams differ in one key-value pair.

Definition 1 ((ϵ, δ) Differential Privacy)**.** *A randomized algorithm $\mathcal{M}$ is said to preserve (ϵ, δ)-DP if for two neighboring datasets $\mathcal{M}, \mathcal{M}'$ and any measurable subset of the output space s,*

$$\mathbb{P}[\mathcal{M}(\mathcal{S}) \in s] \leq e^{\epsilon}\mathbb{P}[\mathcal{M}(\mathcal{S}') \in s] + \delta$$

2 Differentially Private Frequency Moments Estimation

In this section, we establish the differential privacy guarantee provided by $\mathbb{F}_p$ sketch step by step. We first revisit $\mathbb{F}_p$ sketch and then prove that $\mathbb{F}_p$ sketch preserves (ϵ, δ)-differential privacy as is. Different from most differential privacy analyses based on additive sensitivity, our proof depends on a variant of the multiplicative sensitivity [7] called *pure multiplicative sensitivity*. We give the first analysis of pure multiplicative sensitivity for p-th frequency moments. Then we motivate the differential privacy proof using a special case when $p = 2$, which is easier to deal with as 2-stable distributions have close-form density functions. Finally we proceed to the general proof that $\mathbb{F}_p$ sketch preserves differential privacy. The main challenge stems from the fact that the density functions of p-stable distributions for $1 < p < 2$ have no close-form expressions.

2.1 Revisiting $\mathbb{F}_p$ Sketch

For completeness, we revisit the well-celebrated $\mathbb{F}_p$ sketch [14] (also known as stable projection or compressed counting). We first introduce p-stable distribution, the basic building block used in $\mathbb{F}_p$ sketch. Then we review how to construct and query a $\mathbb{F}_p$ sketch using stable distribution. Different from most differential privacy analysis which depends on additive sensitivity, our analysis is based on multiplicative sensitivity [7].

Definition 2 (p-stable distribution)**.** *A random variable X follows a β-skewed p-stable distribution if its characteristic function is*

$$\phi_X(t) = \exp(-\zeta|t|^p(1 - \sqrt{-1}\beta\,\mathrm{sgn}(t)\tan(\frac{\pi p}{2}))$$

where $-1 \leq \beta \leq 1$ *is the skewness parameter,* $\zeta > 0$ *is the scale parameter.*

In this paper, we focus on stable distributions with $\beta = 0$, namely symmetric stable distributions. We denote symmetric p-stable distribution by $\mathcal{D}_{p,\zeta}$. We slightly abuse the notation to denote the density function as $\mathcal{D}_{p,\zeta}(x)$ for clarity. If two independent random variables $X_1, X_2 \sim \mathcal{D}_{p,1}$, then $C_1X_1 + C_2X_2 \sim \mathcal{D}_{p,C_1^p+C_2^p}$. We refer to this property as *p-stability*.

$\mathbb{F}_p$ sketch (Algorithm 1) leverages the p-stability of $\mathcal{D}_p$ to keep track of the frequency moments. As shown in line 2 of algorithm 1, for each incoming key-value pair (k_i, v_i), we multiply the random projection matrix $\mathbf{P}$ by the one-hot encoding of k_i scaled by v_i. At the end of the algorithm,

$$\mathbf{a} = \sum_{i=1}^{n} \mathbf{P} \times v_i \mathbf{e}_{k_i} = \sum_{k=1}^{m} \mathbf{P} \times (\sum_{k_i=k} v_i)\mathbf{e}_{k_i} \sim \mathcal{D}^r_{p,F_p(\mathcal{S})}$$

Thus, we can estimate the scale of the components in $\mathbf{a}$ to get $\mathbb{F}_p$.

Algorithm 1: $\mathbb{F}_p$ sketch.

Input : Data stream: $\mathcal{S} = \{(k_1, v_1), \cdots, (k_n, v_n)\}$; privacy budget: (ϵ, δ); accuracy constraint: (γ, η); p stable distribution: $\mathcal{D}_{p,1}$.

1 **Construction**

2 Initialize $\mathbf{a} = 0^r$, $P \sim \mathcal{D}_{p,1}^{r \times m}$;

3 **for** $i \in [n]$ **do** Let e_{k_i} be the one-hot encoder of k_i, $\mathbf{a} = \mathbf{a} + \mathbf{P} \times v_i \mathbf{e}_{k_i}$;

4 **Query**

5 return median($\mathbf{a}$);

2.2 Pure multiplicative sensitivity of $\mathbb{F}_p$ sketch

As we will see in the following two subsections, the differential privacy proof of $\mathbb{F}_p$ sketch depends on the pure multiplicative sensitivity of p-th frequency moments. As the first step, we give the definition of pure multiplicative differential privacy. "Pure" is to differentiate from multiplicative sensitivity as defined in [7].

Definition 3 (Pure multiplicative sensitivity)**.** *The multiplicative sensitivity of a (deterministic) mechanism* $\mathcal{M}$ *is defined as the maximum ratio between outputs on neighboring inputs* $\mathcal{S}$ *and* $\mathcal{S}'$*.*

$$\rho_{\mathcal{M}}(n) = \sup_{|\mathcal{S}|=n, |\mathcal{S}'|=n, d(\mathcal{S},\mathcal{S}')=1} |\frac{\mathcal{M}(\mathcal{S})}{\mathcal{M}(\mathcal{S}')}|$$

We might omit the subscript and argument when the mechanism is clear from the context.

The pure multiplicative sensitivity of F_2 is as below. We defer the proof to Appendix A.

Theorem 1 (Multiplicative sensitivity of F_p)**.** *A mechanism $\mathcal{M}$ which accurately calculates F_p has multiplicative sensitivity upper bounded by*

$$\rho_{\mathcal{M}} = \begin{cases} 2^{2-2p}\big(\dfrac{n-1+M}{n-1+(m-1)^{\frac{p-1}{p}}}\big)^p, 0 < p < 1 \\ 2^{2p-2}\big(\dfrac{n-1+(m-1)^{\frac{p-1}{p}}M}{n}\big)^p, 1 < p \leq 2 \end{cases}$$

In a typical streaming model where $n \gg m$, $\rho_{\mathcal{M}} \approx \max\{2^{2p-2}, 2^{2-2p}\} \lessapprox 4$.

2.3 Differentially Private $\mathbb{F}_2$ Sketch

Instead of diving into the full analysis directly, we first motivate the analysis with a special case when $p = 2$. Note that the 2-stable distribution is the well-known normal distribution. As normal distribution has explicit density function, the analysis is greatly simplified.

Theorem 2. *Let ρ_2 represents the multiplicative sensitivity of the second frequency moments. $\forall \epsilon \geq \frac{1}{2}\ln\rho_2, \mathbb{F}_2$ is $(\epsilon, 2\Phi^c(\sqrt{\frac{\rho_2(\ln\rho_2+2\epsilon)}{(\rho_2-1)}}))$-differentially private where Φ^c is the complementary cumulative distribution function of the standard normal distribution.*

Proof for Theorem 2. To prove $\mathbb{F}_2$ sketch is (ϵ, δ)-differentially private, it is sufficient to prove $\mathbb{P}[e^{-\epsilon} \leq \frac{\mathcal{D}_{2,F_2}(x)}{\mathcal{D}_{2,\rho_2 F_2}(x)} \leq e^{\epsilon}] \geq 1-\delta$ as integral preserves the inequality between functions. If we set $\epsilon = \frac{1}{2}\ln\rho_2$, the right-hand side of the inequality always holds because (1) $\frac{\mathcal{D}_{2,F_2}(0)}{\mathcal{D}_{2,\rho_2 F_2}(0)} = e^{\frac{1}{2}\ln\rho_2}$; (2) the ratio between two normal density functions is decreasing in the positive half axis if the nominator has smaller scale than the denominator. Note that $\frac{\mathcal{D}_{2,F_2}(x)}{\mathcal{D}_{2,\rho_2 F_2}(x)} = \frac{2}{\ln\rho_2}$ when $x = \sqrt{\frac{2F_2\rho_2\ln\rho_2}{(\rho_2-1)}}$. Thus the left-hand side holds with probability at least $1 - \mathbb{P}_{X\sim\mathcal{D}_{2,F_2}}[|X| \geq \sqrt{\frac{2F_2\rho_2\ln\rho_2}{(\rho_2-1)}}] = 1 - 2\Phi^c(\sqrt{\frac{2\rho_2\ln\rho_2}{(\rho_2-1)}})$. Note the F_2 in the nominator disappears due to the transform between two cumulative functions with different scale. This is important as it frees the privacy parameter from the exact output. We will see later that this holds generally for all $\mathbb{F}_p$ sketches. □

Interpretation of the privacy parameters. As $\Phi^c(\cdot)$ has no close-form expression, it might be confusing how to interpret the privacy parameters. To clarify, since $\Phi^c(t) \leq \frac{1}{\sqrt{2\pi}t}e^{-t^2/2}$, the privacy parameters can be relaxed to $(\epsilon, \mathcal{O}(e^{-\frac{2\epsilon\rho}{\rho-1}}))$ for $\epsilon \geq \frac{1}{2}\ln\rho_2$.

2.4 Differentially Private $\mathbb{F}_p$ Sketch

Theorem 3. *Let ρ_p represents the multiplicative sensitivity of the p-th frequency moments. $\mathbb{F}_p$ is $\forall \epsilon \geq \frac{1}{p}\ln\rho_p, (\frac{1}{p}\ln\rho_p, 2\mathbb{P}_{X\sim\mathcal{D}_{p,1}}[X \geq \lambda])$-differentially private where λ is the smallest positive real value such that $\frac{\mathcal{D}_{p,1}(\lambda)}{\mathcal{D}_{p,\rho_p}(\lambda)} \leq e^{-\epsilon}$.*

Theorem 3 is an analogue of Theorem 2 in a general case. The proof is also similar except that $\frac{\mathcal{D}_{p,F_p}(x)}{\mathcal{D}_{p,\rho_p F_p}(x)}$ is no longer decreasing and we fail to find a close-form of λ as in $\mathbb{F}_2$ case. Thus we need to numerically find λ. To do so, we need to prove that such λ exists. Besides, we also need to prove that the ratio reaches maximum at 0.

Lemma 1. *Assume $\zeta_1 \leq \zeta_2$ without loss of generality, then $\forall x \in \mathbb{R}$, $\frac{\mathcal{D}_{p,\zeta_1}(x)}{\mathcal{D}_{p,\zeta_2}(x)} \leq \frac{\mathcal{D}_{p,\zeta_1}(0)}{\mathcal{D}_{p,\zeta_2}(0)}$.*

Proof for lemma 1 is deferred to Appendix B.

Lemma 2. *There exists $0 < \lambda < \infty$ such that $\frac{\mathcal{D}_{p,1}(\lambda)}{\mathcal{D}_{p,\rho_p}(\lambda)} = \sqrt[p]{\frac{1}{\rho_p}}$.*

Proof for Lemma 2. As discovered in [27],

$$\mathcal{D}_{p,F}(x) = 2\int_0^\infty \cos tx \cdot e^{-\zeta_1 t^p} dt = \frac{F}{x^{p+1}} \int_0^\infty e^{Ftx^{-p}} \sin t^{1/p} dt$$

Thus,

$$\lim_{x\to\infty} \frac{\mathcal{D}_{p,F}(x)}{\mathcal{D}_{p,\rho_p F}(x)} = \lim_{x->0} \frac{F\int_0^\infty \sin t^{1/p} dt}{\rho_p F \int_0^\infty \sin t^{1/p} dt} = \frac{1}{\rho_p}$$

As $\frac{\mathcal{D}_{p,F}(\lambda)}{\mathcal{D}_{p,F\rho_p}(\lambda)}$ is continuous, according to the intermediate value theorem, $\exists\lambda$, $\frac{\mathcal{D}_{p,F}(\lambda)}{\mathcal{D}_{p,F\rho_p}(\lambda)} = \sqrt[p]{\frac{1}{\rho_p}} \in [\sqrt[p]{\rho_p}, \frac{1}{\rho_p}]$. □

The next step is to prove that the probability that $X \sim \mathcal{D}_{p,F_p}$ exceeds λ does not depend on F_p.

Lemma 3. *For a fixed $x \in [\sqrt[p]{\rho_p}, \frac{1}{\rho_p}]$, let $\lambda_F \in \mathbb{R}$ be the smallest positive value such that $\frac{\mathcal{D}_{p,F}(\lambda)}{\mathcal{D}_{p,F\rho_p}(\lambda)} = x$. $\mathbb{P}_{X\sim\mathcal{D}_{p,F}}[X \leq \lambda_F] = \mathbb{P}_{X\sim\mathcal{D}_{p,F}}[X \leq \lambda_1]$.*

Proof for Lemma 3. It is easy to verify that if $\lambda_F = \sqrt[p]{F}\lambda_1$, then $\frac{\mathcal{D}_{p,F}(\lambda_F)}{\mathcal{D}_{p,F\rho_p}(\lambda_F)} = \frac{\mathcal{D}_{p,F}(\sqrt[p]{F}\lambda_1)}{\mathcal{D}_{p,F\rho_p}(\sqrt[p]{F}\lambda_1)} = \frac{\mathcal{D}_{p,1}(\lambda_1)}{\mathcal{D}_{p,\rho_p}(\lambda_1)}$. It is also easy to verify that the transformation preserves the minimality of λ using proof by contradiction. Hence,

$$\mathbb{P}_{X\sim\mathcal{D}_{p,F}}[X \leq \lambda_F] = \mathbb{P}_{X\sim\mathcal{D}_{p,F}}[X \leq \sqrt[p]{F}\lambda_F] = \mathbb{P}_{X\sim\mathcal{D}_{p,1}}[X \leq \lambda_1]$$

□

References

[1] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and system sciences*, 58(1):137–147, 1999.

[2] Raef Bassily and Adam Smith. Local, private, efficient protocols for succinct histograms. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 127–135, 2015.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. In *International Colloquium on Automata, Languages, and Programming*, pages 693–703. Springer, 2002.

[5] Kenneth L Clarkson and David P Woodruff. Numerical linear algebra in the streaming model. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 205–214, 2009.

[6] Graham Cormode, S Muthukrishnan, and Irina Rozenbaum. Summarizing and mining inverse distributions on data streams via dynamic inverse sampling. In *VLDB*, volume 5, pages 25–36, 2005.

[7] Cynthia Dwork, Weijie Su, and Li Zhang. Private false discovery rate control. *arXiv preprint arXiv:1511.03803*, 2015.

[8] Joan Feigenbaum, Sampath Kannan, Martin J Strauss, and Mahesh Viswanathan. An approximate l 1-difference algorithm for massive data streams. *SIAM Journal on Computing*, 32(1):131–151, 2002.

[9] Davi Geiger, Tyng-Luh Liu, and Michael J Donahue. Sparse representations for image decompositions. *International Journal of Computer Vision*, 33(2):139–156, 1999.

[10] Corrado Gini. Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E*, 1912.

[11] IJ Good. C332. surprise indexes and p-values. 1989.

[12] Nicholas JA Harvey, Jelani Nelson, and Krzysztof Onak. Sketching and streaming entropy via approximation theory. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 489–498. IEEE, 2008.

[13] Avinatan Hassidim, Haim Kaplan, Yishay Mansour, Yossi Matias, and Uri Stemmer. Adversarially robust streaming algorithms via differential privacy. *arXiv preprint arXiv:2004.05975*, 2020.

[14] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM (JACM)*, 53(3):307–323, 2006.

[15] Piotr Indyk and Andrew McGregor. Declaring independence via the sketching of sketches. In *SODA*, volume 8, pages 737–745, 2008.

[16] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*, pages 2436–2444. PMLR, 2016.

[17] Daniel M Kane, Jelani Nelson, Ely Porat, and David P Woodruff. Fast moment estimation in data streams in optimal space. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 745–754, 2011.

[18] Daniel M Kane, Jelani Nelson, and David P Woodruff. On the exact space complexity of sketching and streaming small norms. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1161–1178. SIAM, 2010.

[19] Balachander Krishnamurthy, Subhabrata Sen, Yin Zhang, and Yan Chen. Sketch-based change detection: Methods, evaluation, and applications. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 234–247, 2003.

[20] Ping Li. Estimators and tail bounds for dimension reduction in l α ($0 < \alpha \leq 2$) using stable random projections. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 10–19, 2008.

[21] Max O Lorenz. Methods of measuring the concentration of wealth. *Publications of the American statistical association*, 9(70):209–219, 1905.

[22] Jelani Nelson and David P Woodruff. A near-optimal algorithm for l1-difference. *arXiv preprint arXiv:0904.2027*, 2009.

[23] Jelani Nelson and David P Woodruff. Fast manhattan sketches in data streams. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 99–110, 2010.

[24] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 143–152. IEEE, 2006.

[25] Mikkel Thorup and Yin Zhang. Tabulation based 4-universal hashing with applications to second moment estimation. In *SODA*, volume 4, pages 615–624, 2004.

[26] Tianhao Wang, Milan Lopuhaä-Zwakenberg, Zitao Li, Boris Skoric, and Ninghui Li. Locally differentially private frequency estimation with consistency. *arXiv preprint arXiv:1905.08320*, 2019.

[27] Aurel Wintner. The singularities of Cauchy's distributions. *Duke Mathematical Journal*, 8(4):678 – 681, 1941.

[28] Haiquan Zhao, Ashwin Lall, Mitsunori Ogihara, Oliver Spatscheck, Jia Wang, and Jun Xu. A data streaming algorithm for estimating entropies of od flows. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 279–290, 2007.

A Proof for Theorem 1

Theorem 1 gives an upper bound on the multiplicative change when two input datasets with the same size m differ in one entry. To prove, we first consider a slightly different setting when the second dataset is generated by adding an entry to the first dataset.

Lemma 4. *Let* $\boldsymbol{u} = \{u_1, \cdots, u_m\}$ *where* $u_i \geq 0, \sum_{i=1}^m u_i = s$*, then,*

$$\forall 0 < p < 1, \sup_{\boldsymbol{u}} \frac{\sum_{i=2}^m u_i^p + (u_1 + \Delta)^p}{\sum_{i=2}^m u_i^p + u_1^p} \leq 2^{1-p}(1 + \frac{\Delta}{s})^p$$

$$\forall 0 < p < 1, \inf_{\boldsymbol{u}} \frac{\sum_{i=2}^m u_i^p + (u_1 + \Delta)^p}{\sum_{i=2}^m u_i^p + u_1^p} \geq 2^{p-1}(1 + \frac{(m-1)^{\frac{p-1}{p}}\Delta}{s})^p$$

$$\forall 1 < p \leq 2, \sup_{\boldsymbol{u}} \frac{\sum_{i=2}^m u_i^p + (u_1 + \Delta)^p}{\sum_{i=2}^m u_i^p + u_1^p} \leq 2^{p-1}(1 + \frac{(m-1)^{\frac{p-1}{p}}\Delta}{s})^p$$

$$\forall 1 < p \leq 2, \inf_{\boldsymbol{u}} \frac{\sum_{i=2}^m u_i^p + (u_1 + \Delta)^p}{\sum_{i=2}^m u_i^p + u_1^p} \geq 2^{1-p}(1 + \frac{\Delta}{s})^p$$

To prove Lemma 4, we need the following two lemmas.

Lemma 5. *Given* $a, b, c, d \geq 0, a > b, c > d$*, then* $\frac{a+c}{a+d} \leq \frac{b+c}{b+d}$.

Lemma 5 seems to be a standard result but we fail to find a formal theorem for it. For completeness, we provide the proof here.

Proof for Lemma 5.

$$a(c-d) \geq b(c-d) \Rightarrow ad + bc \leq ac + bd \Rightarrow ab + cd + ad + bc \leq ab + cd + ac + bd$$

$$\Rightarrow (a+c)(b+d) \leq (a+d)(b+c) \Rightarrow \frac{a+c}{a+d} \leq \frac{b+c}{b+d}$$

□

Lemma 6. *Let* $\boldsymbol{u} = \{u_1, \cdots, u_m\}$ *where* $u_i \geq 0$*, then,*

$$\forall 0 < p < 1, (\sum_{i=1}^m u_i)^p \leq \sum_{i=1}^m u_i^p \leq m^{1-p}(\sum_{i=1}^m u_i)^p$$

$$\forall 1 < p \leq 2, m^{1-p}(\sum_{i=1}^{m} u_i)^p \leq \sum_{i=1}^{m} u_i^p \leq (\sum_{i=1}^{m} u_i)^p$$

Proof for Lemma 6. When $0 < p < 1$, we first prove the inequality on the right. According to Hólder's inequality, if $r > 1$, then

$$\sum_{i=1}^{m} |a_i||b_i| \leq (\sum_{i=1}^{m} |a_i|^r)^{\frac{1}{r}} (\sum_{i=1}^{m} |b_i|^{\frac{r}{r-1}})^{\frac{r-1}{r}} \tag{1}$$

Let $a_i = u_i^p, b_i = 1, r = \frac{1}{p}$, we have

$$\sum_{i=1}^{m} u_i^p \leq (\sum_{i=1}^{m} (u_i^p)^{\frac{1}{p}})^p (\sum_{i=1}^{m} 1)^{1-p} = m^{1-p}(\sum_{i=1}^{m} u_i)^p$$

Then we prove the left-hand-side inequality by induction.

Base case: When $m = 2$, without loss of generality, we assume $u_2 \leq u_1$

$$(u_1 + u_2)^p = (1 + u_1/u_2)^p u_2^p \leq (1 + u_1/u_2) u_2^p \leq u_2^p + u_1 u_2^{p-1} \leq u_1^p + u_2^p$$

Hypothesis: When $m = k$,

$$(\sum_{i=1}^{k} u_i)^p \leq \sum_{i=1}^{k} u_i^p$$

Inductive Step: When $m = k + 1$

$$\sum_{i=1}^{k+1} u_i^p = \sum_{i=1}^{k} u_i^p + u_{k+1}^p \geq (\sum_{i=1}^{k} u_i)^p + u_{k+1}^p \geq (\sum_{i=1}^{k+1} u_i)^p$$

When $1 < p \leq 2$, we first prove the inequality on the left. Let $a_i = u_i, b_i = 1, r = p$ in (1), we have

$$\sum_{i=1}^{m} u_i \leq (\sum_{i=1}^{m} u_i^p)^{\frac{1}{p}} (\sum_{i=1}^{m} 1)^{\frac{p-1}{p}} = m^{\frac{p-1}{p}} (\sum_{i=1}^{m} |u_i|^p)^{\frac{1}{p}} \Rightarrow m^{1-p}(\sum_{i=1}^{m} u_i)^p \leq \sum_{i=1}^{m} u_i^p$$

Then we prove the right-hand-side inequality by induction.

Base case: When $m = 2$, without loss of generality, we assume $u_2 \leq u_1$

$$(u_1 + u_2)^p = (1 + u_2/u_1)^p u_1^p \geq (1 + u_2/u_1) u_1^p \geq u_1^p + u_2 u_1^{p-1} \geq u_1^p + u_2^p$$

Hypothesis: When $m = k$,

$$\sum_{i=1}^{k} u_i^p \leq (\sum_{i=1}^{k} u_i)^p$$

Inductive Step: When $m = k + 1$

$$\sum_{i=1}^{k+1} u_i^p = \sum_{i=1}^{k} u_i^p + u_{k+1}^p \leq (\sum_{i=1}^{k} u_i)^p + u_{k+1}^p \leq (\sum_{i=1}^{k+1} u_i)^p$$

□

Proof for Lemma 4. When $0 < p < 1$,we prove the supremum first. According to Lemma 5 and 6,

$$\frac{\sum_{i=2}^{m} u_i^p + (u_1+\Delta)^p}{\sum_{i=2}^{m} u_i^p + u_1^p} \leq \frac{(\sum_{i=2}^{m} u_i)^p + (u_1+\Delta)^p}{(\sum_{i=2}^{m} u_i)^p + u_1^p} = \frac{(s-u_1)^p + (u_1+\Delta)^p}{(s-u_1)^p + u_1^p} \leq 2^{1-p}(1+\frac{\Delta}{s})^p$$

Similarly, for the infimum,

$$\begin{aligned}
\frac{\sum_{i=2}^{m} u_i^p + (u_1+\Delta)^p}{\sum_{i=2}^{m} u_i^p + u_1^p} &\geq \frac{(m-1)^{1-p}(\sum_{i=2}^{m} u_i)^p + (u_1+\Delta)^p}{(m-1)^{1-p}(\sum_{i=2}^{m} u_i)^p + u_1^p} \\
&= \frac{(s-u_1)^p + ((m-1)^{\frac{p-1}{p}}(u_1+\Delta))^p}{(s-u_1)^p + ((m-1)^{\frac{p-1}{p}}u_1)^p} \\
&\geq 2^{p-1}\big(\frac{((m-1)^{\frac{p-1}{p}}-1)u_i + s + (m-1)^{\frac{p-1}{p}}\Delta}{((m-1)^{\frac{p-1}{p}}-1)u_i + s}\big)^p \\
&= 2^{p-1}(1+\frac{(m-1)^{\frac{p-1}{p}}\Delta}{s})^p
\end{aligned}$$

When $1 < p \leq 2$, we will prove the infimum first. According to Lemma 5 and 6,

$$\frac{\sum_{i=2}^{m} u_i^p + (u_1+\Delta)^p}{\sum_{i=2}^{m} u_i^p + u_1^p} \geq \frac{(\sum_{i=2}^{m} u_i)^p + (u_1+\Delta)^p}{(\sum_{i=2}^{m} u_i)^p + u_1^p} = \frac{(s-u_1)^p + (u_1+\Delta)^p}{(s-u_1)^p + u_1^p} \geq 2^{1-p}(1+\frac{\Delta}{s})^p$$

Similarly, for the supremum,

$$\begin{aligned}
\frac{\sum_{i=2}^{m} u_i^p + (u_1+\Delta)^p}{\sum_{i=2}^{m} u_i^p + u_1^p} &\leq \frac{(m-1)^{1-p}(\sum_{i=2}^{m} u_i)^p + (u_1+\Delta)^p}{(m-1)^{1-p}(\sum_{i=2}^{m} u_i)^p + u_1^p} \\
&= \frac{(s-u_1)^p + ((m-1)^{\frac{p-1}{p}}(u_1+\Delta))^p}{(s-u_1)^p + ((m-1)^{\frac{p-1}{p}}u_1)^p} \\
&\leq 2^{p-1}\big(\frac{((m-1)^{\frac{p-1}{p}}-1)u_i + s + (m-1)^{\frac{p-1}{p}}\Delta}{((m-1)^{\frac{p-1}{p}}-1)u_i + s}\big)^p \\
&= 2^{p-1}(1+\frac{(m-1)^{\frac{p-1}{p}}\Delta}{s})^p
\end{aligned}$$

□

B Proof for Lemma 1

Although Lemma 1 seems natural, it is not trivial to rigorously prove since general p-stable distributions do not have explicit probability density functions. To prove Lemma 1, we need the following lemma.

Lemma 7. *Given two decreasing function $f(t)$ and $g(t)$ integrable on $[0,\infty)$ where $f(t) \geq g(t) > 0$, if $\frac{f(t)}{g(t)}$ is an increasing function and x is a non-negative real number, then we have*

$$\frac{\int_0^\infty \cos(tx)f(t)dt}{\int_0^\infty \cos(tx)g(t)dt} \leq \frac{\int_0^\infty f(t)dt}{\int_0^\infty g(t)dt}$$

Proof for Lemma 7. We first give the following claim, which can be verified within a few lines of algebra.

Claim 1. *Given positive real numbers a, b, c, d, $a > b, c > d, \frac{a}{c} \leq \frac{b}{d}$. If $-1 \leq s \leq 1$, then*

$$\frac{a+sb}{c+sd} \leq \frac{a+b}{c+d}$$

We now prove Lemma 7 with induction.

Base case: If we set $a = f(0), b = f(\Delta), c = g(0), d = g(\Delta), s = \cos(x\Delta), \Delta > 0$, then we have

$$\frac{f(0)+\cos(x\Delta)f(\Delta)}{g(0)+\cos(x\Delta)g(\Delta)} \leq \frac{f(0)+f(\Delta)}{g(0)+g(\Delta)}$$

Hypothesis:

$$\frac{\sum_{i=0}^k \cos(ix\Delta)f(i\Delta)}{\sum_{i=0}^k \cos(ix\Delta)g(i\Delta)} \leq \frac{\sum_{i=0}^k f(i\Delta)}{\sum_{i=0}^k g(i\Delta)}$$

Inductive step: According to the hypothesis and Claim 1,

$$\frac{\sum_{i=0}^{k+1} \cos(ix\Delta)f(i\Delta)}{\sum_{i=0}^{k+1} \cos(ix\Delta)g(i\Delta)} \leq \frac{\sum_{i=0}^{k+1} f(i\Delta)}{\sum_{i=0}^{k+1} g(i\Delta)}$$

Thus,

$$\frac{\int_0^\infty \cos(tx)f(t)dt}{\int_0^\infty \cos(tx)g(t)dt} = \frac{\lim_{\Delta\to 0}\sum_{i=0}^\infty \cos(ix\Delta)f(i\Delta)\Delta}{\lim_{\Delta\to 0}\sum_{i=0}^\infty \cos(ix\Delta)g(i\Delta)\Delta} \leq \frac{\lim_{\Delta\to 0}\sum_{i=0}^\infty f(i\Delta)\Delta}{\lim_{\Delta\to 0}\sum_{i=0}^\infty g(i\Delta)\Delta} \leq \frac{\int_0^\infty f(t)dt}{\int_0^\infty g(t)dt}$$

□

Proof for Lemma 1. The pdf of $\mathcal{D}_{p,\zeta}$ is the inverse Fourier transform of the characteristic function $\phi(t)$: $\mathcal{D}_{p,\zeta}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi(t)dt = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\cdot e^{-\zeta|t|^p}dt = \frac{1}{\pi}\int_0^\infty \cos tx \cdot e^{-\zeta|t|^p}dt = f_\zeta(x)$.

$$\frac{\mathcal{D}_{p,\zeta_1}(x)}{\mathcal{D}_{p,\zeta_2}(x)} = \frac{\int_0^\infty \cos tx \cdot e^{-\zeta_1|t|^p}dt}{\int_0^\infty \cos tx \cdot e^{-\zeta_2|t|^p}dt}$$

If we set $f(t) = e^{-\zeta_1 t^p}, g(t) = e^{-\zeta_2 t^p}$, it is easy to verify they satisfy all the requirements in Lemma 7. Thus Lemma 1 is a simple application of Lemma 7. □