This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Quantum Query Complexity of Entropy Estimation

Tongyang Li and Xiaodi Wu

***Abstract*—Estimation of Shannon and Rényi entropies of unknown discrete distributions is a fundamental problem in statistical property testing. In this paper, we give the first quantum algorithms for estimating α-Rényi entropies (Shannon entropy being 1-Rényi entropy). In particular, we demonstrate a quadratic quantum speedup for Shannon entropy estimation and a generic quantum speedup for α-Rényi entropy estimation for all $\alpha \geq 0$, including tight bounds for the Shannon entropy, the Hartley entropy ($\alpha = 0$), and the collision entropy ($\alpha = 2$). We also provide quantum upper bounds for estimating min-entropy ($\alpha = +\infty$) as well as the Kullback-Leibler divergence. We complement our results with quantum lower bounds on α-Rényi entropy estimation for all $\alpha \geq 0$.**

Our approach is inspired by the pioneering work of Bravyi, Harrow, and Hassidim (BHH) [1], however, with many new technical ingredients: (1) we improve the error dependence of the BHH framework by a fine-tuned error analysis together with Montanaro's approach to estimating the expected output of quantum subroutines [2] for $\alpha = 0, 1$; (2) we develop a procedure, similar to cooling schedules in simulated annealing, for general $\alpha \geq 0$; (3) in the cases of integer $\alpha \geq 2$ and $\alpha = +\infty$, we reduce the entropy estimation problem to the α-distinctness and the $\lceil \log n \rceil$-distinctness problems, respectively.

***Index Terms*—statistical property testing, sampling, entropy estimation, quantum information, query complexity.**

I. Introduction

Motivations. Property testing is a rapidly developing field in theoretical computer science (e.g. see the survey [3]). It aims to determine properties of an object with the least number of independent samples of the object. Property testing is a theoretically appealing topic with intimate connections to statistics, learning theory, and algorithm design. One important topic in property testing is to estimate statistical properties of unknown distributions (e.g., [4]), which are fundamental questions in statistics and information theory, given that much of science relies on samples furnished by nature. The Shannon [5] and Rényi [6] entropies are central information-theoretical quantities. In this paper, we focus on estimating these entropies for an unknown distribution.

Specifically, given a distribution p over a set X of size n (w.l.o.g. let $X = [n]$) where p_x denotes the probability of $x \in X$, the *Shannon entropy* $H(p)$ of this distribution p is defined by

$$H(p) := \sum_{x \in X:\, p_x > 0} p_x \log\Big(\frac{1}{p_x}\Big). \tag{I.1}$$

A natural question is to determine the *sample complexity* (i.e., the necessary number of independent samples from p) to estimate $H(p)$, with error ϵ and high probability. This problem has been intensively studied in the classical literature. For multiplicative error ϵ, Batu et al. [7, Theorem 2] provided the upper bound of $O(n^{(1+o(1))/(1+\epsilon)^2} \log n)$, while an almost matching lower bound of $\Omega(n^{(1-o(1))/(1+\epsilon)^2})$ was shown by Valiant [4, Theorem 1.3]. For additive errors, Paninski gave a nonconstructive proof of the existence of sublinear estimators in [8, 9], while an explicit construction using $\Theta(\frac{n}{\epsilon \log n})$ samples was shown by Valiant and Valiant in [10, 11] when $\epsilon = \Omega(n^{-0.03})$ and $\epsilon = O(1)$; for the case $\epsilon = O(n^{-0.03})$, Wu and Yang [12] and Jiao et al. [13] gave the optimal estimator with $\Theta(\frac{n}{\epsilon \log n} + \frac{(\log n)^2}{\epsilon^2})$ samples.

A sequence of works in information theory [12–14] studied the minimax mean-squared error, which becomes $O(1)$ also using $\Theta(n/\log n)$ samples.

One important generalization of Shannon entropy is the *Rényi entropy* of order $\alpha > 0$, denoted $H_\alpha(p)$, which is defined by

$$H_\alpha(p) := \begin{cases} \frac{1}{1-\alpha} \log \sum_{x \in X} p_x^\alpha, & \text{when } \alpha \neq 1. \\ \lim_{\alpha \to 1} H_\alpha(p), & \text{when } \alpha = 1. \end{cases} \tag{I.2}$$

The Rényi entropy of order 1 is simply the Shannon entropy, i.e., $H_1(p) = H(p)$. General Rényi entropy can be used as a bound on Shannon entropy, making it useful in many applications (e.g., [15, 16]). Rényi entropy is also of interest in its own right. One prominent example is the Rényi entropy of order 2, $H_2(p)$ (also known as the *collision entropy*), which measures the quality of random number generators (e.g., [17]) and key derivation in cryptographic applications (e.g., [18, 19]). Motivated by these and other applications, the estimation of Rényi entropy has also been actively studied [13, 14, 20]. In particular, Acharya et al. [20] have shown almost tight bounds on the classical query complexity of computing Rényi entropy. Specifically, for any non-integer $\alpha > 1$, the classical query complexity of α-Rényi entropy is $\Omega(n^{1-o(1)})$ and $O(n)$. Surprisingly, for any *integer* $\alpha > 1$, the classical query complexity is $\Theta(n^{1-1/\alpha})$, i.e., *sublinear* in n. When $0 \leq \alpha < 1$, the classical query complexity is $\Omega(n^{1/\alpha - o(1)})$ and $O(n^{1/\alpha})$, which is always superlinear.

The extreme case ($\alpha \to \infty$) is known as the *min-entropy*, denoted $H_\infty(p)$, which is defined by

$$H_\infty(p) := \lim_{\alpha \to \infty} H_\alpha(p) = -\log \max_{i \in [n]} p_i. \tag{I.3}$$

Min-entropy plays an important role in the randomness extraction (e.g., [21]) and characterizes the maximum number of uniform bits that can be extracted from a given distribution. Classically, the query complexity of min-entropy estimation is $\Theta(n/\log n)$, which follows directly from [10].

Tongyang Li and Xiaodi Wu are with Department of Computer Science, Institute for Advanced Computer Studies, and Joint Center for Quantum Information and Computer Science, University of Maryland, USA. Email: {tongyang,xwu}@cs.umd.edu.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Another extreme case ($\alpha = 0$), also known as the *Hartley entropy* [22], is the logarithm of the support size of distributions, where the *support* of any distribution p is defined by

$$\mathrm{Supp}(p) := |\{x : x \in X,\, p_x > 0\}|. \tag{I.4}$$

It is a natural and fundamental quantity of distributions with various applications (e.g., [23–29]). However, estimating the support size is impossible in general because elements with negligible but nonzero probability, which are very unlikely to be sampled, could still contribute to $\mathrm{Supp}(p)$. Two related quantities (*support coverage* and *support size*) have hence been considered as alternatives of 0-Rényi entropy with roughly $\Theta(n/\log(n))$ complexity. (See details in Section VIII.)

Besides the entropic measures of a discrete distribution, we also briefly discuss an entropic measure between two distributions, namely the *Kullback-Leibler (KL) divergence*. Given two discrete distributions p and q with cardinality n, the KL divergence is defined as

$$D_{\mathrm{KL}}(p\|q) = \sum_{i\in[n]} p_i \log \frac{p_i}{q_i}. \tag{I.5}$$

KL divergence is a key measure with many applications in information theory [30, 31], data compression [32], and learning theory [33]. Classically, under the assumption that $\frac{p_i}{q_i} \leq f(n)\ \forall\, i \in [n]$ for some $f(n)$, $D_{\mathrm{KL}}(p\|q)$ can be approximated within constant additive error with high success probability if $\Theta(\frac{n}{\log n})$ samples are taken from p and $\Theta(\frac{nf(n)}{\log n})$ samples are taken from q.

Main question. In this paper, we study the impact of quantum computation on estimation of general Rényi entropies. Specifically, we aim to characterize *quantum speed-ups for estimating Shannon and Rényi entropies*.

Our question aligns with the emerging topic called "quantum property testing" (see the survey [34]) and focuses on investigating the quantum advantage in testing classical statistical properties. To the best of our knowledge, the first research paper on distributional quantum property testing is by Bravyi, Harrow, and Hassidim (BHH) [1], where they discovered quantum speedups for testing uniformity, orthogonality, and statistical difference on unknown distributions. Some of these results were subsequently improved by Chakraborty et al. [35]. Reference [1] also claimed that Shannon entropy could be estimated with query complexity $O(\sqrt{n})$, however, without details and explicit error dependence. Indeed, our framework is inspired by [1], but with significantly new ingredients to achieve our results. There is also a related line of research on spectrum testing or tomography of quantum states [36–40]; in particular, Ref. [40] studied the quantum sample complexity of estimating von Neumann and Rényi entropies of a quantum state. However, these works aim to test properties of general quantum states, while we focus on using quantum algorithms to test properties of classical distributions (i.e., diagonal quantum states)[1].

[1]Note that one can also leverage the results of [36–40] to test properties of classical distributions. However, they are less efficient because they deal with a much harder problem involving general quantum states.

Distributions as oracles. The sampling model in the classical literature assumes that a tester is presented with independent samples from an unknown distribution. One of the contributions of BHH is an alternative model that allows coherent quantum access to unknown distributions. Specifically, BHH models a discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$ by an *oracle* $O_p\colon [S] \to [n]$ for some $S \in \mathbb{N}$. The probability p_i ($i \in [n]$) is proportional to the size of pre-image of i under O_p. Namely, an oracle $O_p\colon [S] \to [n]$ generates p if and only if for all $i \in [n]$,

$$p_i = |\{s \in [S] : O_p(s) = i\}|/S. \tag{I.6}$$

(note that we assume p_is to be rational numbers). If one samples s uniformly from $[S]$, then the output $O_p(s)$ is from distribution p. Instead of considering sample complexity—that is, the number of used samples—we consider the *query complexity* in the oracle model that counts the number of oracle uses. Note that a tester interacting with an oracle can potentially be more powerful due to the possibility of learning the internal structure of the oracle as opposed to the sampling model. However, it is shown in [1] that the query complexity of the oracle model and the sample complexity of the sampling model are in fact the same classically.

A significant advantage of the oracle model is that it naturally allows coherent access when extended to the quantum case, where we transform O_p into a unitary operator $\hat{O}_p$ acting on $\mathbb{C}^S \otimes \mathbb{C}^{n+1}$ such that

$$\hat{O}_p|s\rangle|0\rangle = |s\rangle|O_p(s)\rangle \quad \forall\, s \in [S]. \tag{I.7}$$

Moreover, this oracle model can also be readily obtained in some algorithmic settings, e.g., when distributions are generated by some classical or quantum sampling procedure. Thus, statistical property testing results in this oracle model can be potentially leveraged in algorithm design.

Our Results. Our main contribution is a systematic study of both upper and lower bounds for the *quantum query complexity* of estimation of Rényi entropies (including Shannon entropy as a special case). Specifically, we obtain the following quantum speedups for different ranges of α.

Theorem I.1. *There are quantum algorithms that approximate $H_\alpha(p)$ of distribution p on $[n]$ within an additive error $0 < \epsilon \leq O(1)$ with success probability at least 2/3 using*[2]

- $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^{1.5}}\big)$ *quantum queries when $\alpha = 0$, i.e., Hartley entropy. See Theorem VIII.2.*[3]
- $\tilde{O}\big(\frac{n^{1/\alpha-1/2}}{\epsilon^2}\big)$ *quantum queries*[4] *when $0 < \alpha < 1$. See Theorem V.2.*
- $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^2}\big)$ *quantum queries when $\alpha = 1$, i.e., Shannon entropy. See Theorem III.1.*
- $\tilde{O}\big(\frac{n^{\nu(1-1/\alpha)}}{\epsilon^2}\big)$ *quantum queries when $\alpha > 1, \alpha \in \mathbb{N}$ for some $\nu < \frac{3}{4}$. See Theorem VI.1.*

[2]It should be understood that the success probability $2/3$ can be boosted to close to 1 without much overhead, e.g., see Lemma V.5 in Section V-A5.

[3]0-Rényi entropy estimation is intractable without any assumption, both classically and quantumly. Here, the results are based on the assumption that nonzero probabilities are at least $1/n$. See Section VIII for more information.

[4]$\tilde{O}$ hides factors that are polynomial in $\log n$ and $\log 1/\epsilon$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

- $\tilde{O}\big(\frac{n^{1-1/2\alpha}}{\epsilon^2}\big)$ *quantum queries when* $\alpha > 1, \alpha \notin \mathbb{N}$. *See Theorem V.1.*
- $\tilde{O}\big(Q(\lceil\frac{16\log n}{\epsilon^2}\rceil\text{-distinctness})\big)$ *quantum queries when* $\alpha = \infty$, *where* $Q(\lceil\frac{16\log n}{\epsilon^2}\rceil\text{-distinctness})$ *is the quantum query complexity of the* $\lceil\frac{16\log n}{\epsilon^2}\rceil$*-distinctness problem. See Theorem VII.1.*

Our quantum testers demonstrate advantages over classical ones for all $0 < \alpha < \infty$; in particular, our quantum tester has a quadratic speedup in the case of Shannon entropy. When $\alpha = \infty$, our quantum upper bound depends on the quantum query complexity of the $\lceil\log n\rceil$-distinctness problem, which is open to the best of our knowledge[5] and might demonstrate a quantum advantage.

As a corollary, we also obtain quadratic quantum speedup for estimating KL divergence:

Corollary I.1 (see Theorem IV.1)**.** *Assuming* p *and* q *satisfies* $\frac{p_i}{q_i} \le f(n)\ \forall i \in [n]$ *for some function* $f : \mathbb{N} \to \mathbb{R}^+$, $D_{KL}(p\|q)$, *there is a quantum algorithm that approximates* $D_{KL}(p\|q)$ *within an additive error* $\epsilon > 0$ *with success probability at least* $\frac{2}{3}$ *using* $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^2}\big)$ *quantum queries to* p *and* $\tilde{O}\big(\frac{\sqrt{n}f(n)}{\epsilon^2}\big)$ *quantum queries to* q.

On the other hand, we obtain corresponding quantum lower bounds on entropy estimation using the polynomial method [43, 44], which are then combined with a couple of lower bounds shown in [45]. It is worth mentioning that lower bounds in [45] are established when assuming $\epsilon = O(1)$, whereas our lower bounds have precise error dependence.

We summarize both bounds in Table I and visualize them in Figure 1.

Theorem I.2 (See Theorem IX.1)**.** *Any quantum algorithm that approximates* $H_\alpha(p)$ *of distribution* p *on* $[n]$ *within additive error* ϵ *with success probability at least 2/3 must use*

- $\Omega(\sqrt{n}+n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ *quantum queries when* $\alpha = 0$, *assuming* $1/n \le \epsilon \le 1/12$.
- $\tilde{\Omega}(n^{\frac{1}{7\alpha}-o(1)}/\epsilon^{\frac{2}{7}})$ *quantum queries when* $0 < \alpha < \frac{3}{7}$.
- $\Omega(n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ *quantum queries when* $\frac{3}{7} \le \alpha \le 3$ *and* $\alpha \ne 1$, *assuming* $1/n \le \epsilon \le 1/2$.
- $\Omega(\sqrt{n}+n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ *quantum queries when* $\alpha = 1$, *assuming* $1/n \le \epsilon \le 1/2$.
- $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}}/\epsilon)$ *quantum queries when* $3 \le \alpha < \infty$.
- $\Omega(\sqrt{n}/\epsilon)$ *quantum queries when* $\alpha = \infty$.

Techniques. At a high level, our upper bound is inspired by BHH [1], where we formulate a framework (in Section II) that generalizes the technique in BHH and makes it applicable in our case. Let $F(p) = \sum_x p_x f(p_x)$ for some function $f(\cdot)$ and distribution p. Similar to BHH, we design a master algorithm that samples x from p and then use the *quantum counting primitive* [48] to obtain an estimate $\tilde{p}_x$ of p_x and outputs $f(\tilde{p}_x)$. It is easy to see that the expectation of the output of the master algorithm is roughly[6] $F(p)$. By choosing appropriate $f(\cdot)$s, one can recover $H(p)$ or $H_\alpha(p)$ as well as the ones used in BHH. It suffices then to obtain a good estimate of the output expectation of the master algorithm, which was achieved by multiple independent runs of the master algorithm in BHH.

The performance of the above framework (and its analysis) critically depends on how close the expectation of the algorithm is to $F(p)$ and how concentrated the output distribution is around its expectation, which in turn heavily depends on the specific $f(\cdot)$ in use. Our first contribution is a fine-tuned error analysis for specific $f(\cdot)$s, such as in the case of Shannon entropy (i.e., $f(p_x) = -\log(p_x)$) whose values could be significant for boundary cases of p_x. Instead of only considering the case when $\tilde{p}_x$ is a good estimate of p_x as in BHH, we need to analyze the entire distribution of $\tilde{p}_x$ using quantum counting. We also leverage a generic quantum speedup for estimating the expectation of the output of any quantum procedure with additive errors [2], which significantly improves our error dependence as compared to BHH. These improvements already give a quadratic quantum speedup for Shannon (Section III) and 0-Rényi (Section VIII) entropy estimation. As an application, it also gives a quadratic speedup for estimating the KL-divergence between two distributions (see Section IV).

For general α-Rényi entropy $H_\alpha(p)$, we choose $f(p_x) = p_x^{\alpha-1}$ and let $P_\alpha(p) = F(p)$ so that $H_\alpha(p) \propto \log P_\alpha(p)$. Instead of estimating $F(p)$ with *additive errors* in the case of Shannon entropy, we switch to working with *multiplicative errors* which is harder since the aforementioned quantum

α	classical bounds	quantum bounds (this paper)
$\alpha = 0$	$\Theta(\frac{n}{\log n})$ [46, 47]	$\tilde{O}(\sqrt{n})$ (this paper), $\tilde{\Omega}(\sqrt{n})$ [45]
$0 < \alpha < 1$	$O(\frac{n^{\frac{1}{\alpha}}}{\log n})$, $\Omega(n^{\frac{1}{\alpha}-o(1)})$ [20]	$\tilde{O}(n^{\frac{1}{\alpha}-\frac{1}{2}})$, $\Omega(\max\{n^{\frac{1}{7\alpha}-o(1)}, n^{\frac{1}{3}}\})$
$\alpha = 1$	$\Theta(\frac{n}{\log n})$ [10, 12, 13]	$\tilde{O}(\sqrt{n})$ (this paper), $\tilde{\Omega}(\sqrt{n})$ [45]
$\alpha > 1, \alpha \notin \mathbb{N}$	$O(\frac{n}{\log n})$, $\Omega(n^{1-o(1)})$ [20]	$\tilde{O}(n^{1-\frac{1}{2\alpha}})$, $\Omega(\max\{n^{\frac{1}{3}}, n^{\frac{1}{2}-\frac{1}{2\alpha}}\})$
$\alpha = 2$	$\Theta(\sqrt{n})$ [20]	$\tilde{\Theta}(n^{\frac{1}{3}})$
$\alpha > 2, \alpha \in \mathbb{N}$	$\Theta(n^{1-1/\alpha})$ [20]	$\tilde{O}(n^{\nu(1-1/\alpha)})$, $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}})$, $\nu < 3/4$
$\alpha = \infty$	$\Theta(\frac{n}{\log n})$ [10]	$\tilde{O}(Q(\lceil\log n\rceil\text{-distinctness}))$, $\Omega(\sqrt{n})$

TABLE I: Summary of classical and quantum query complexity of $H_\alpha(p)$, assuming $\epsilon = \Theta(1)$.

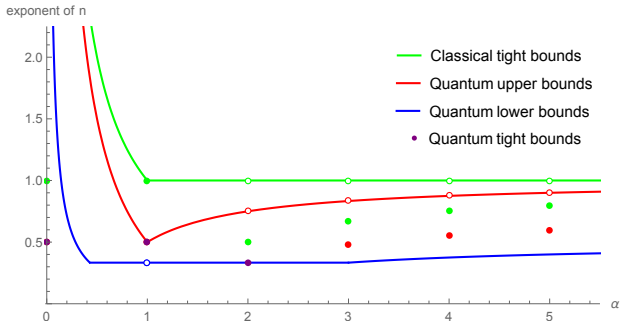

Fig. 1: Visualization of classical and quantum query complexity of $H_\alpha(p)$. The x-axis represents α and the y-axis represents the exponent of n. Red curves and points represent quantum upper bounds. Green curves and points represent classical tight bounds. Blue curve represents quantum lower bounds. Purple points represent quantum tight bounds.

[5] Existing quantum algorithms for the k-distinctness problem (e.g., [41] has query complexity $O(k^2 n^{k/k+1})$ and [42] has query complexity $O(2^{k^2} n^\nu)$ for some $\nu < 3/4$) do not behave well for super-constant ks.

[6] The accurate expectation is $\sum_x p_x \mathbb{E}[f(\tilde{p}_x)]$. Intuitively, we expect $\tilde{p}_x$ to be a good estimate of p_x.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

algorithm [2] is much weaker in this setting. Indeed, by following the same technique, we can only obtain quantum speedups for α-Rényi entropy when $1/2 < \alpha < 2$.

For general $\alpha > 0$, our first observation is that if one knew the output expectation $\mathbb{E}[X]$ is within $[a, b]$ such that $b/a = \Theta(1)$, then one can slightly modify the technique in [2] (as shown in Theorem II.2) and obtain a quadratic quantum speedup similar to the additive error setting. This approach, however, seems circular since it is unclear how to obtain such a, b in advance. Our second observation is that for any close enough α_1, α_2, $P_{\alpha_1}(p)$ can be used to bound $P_{\alpha_2}(p)$. Precisely, when $\alpha_1/\alpha_2 = 1 \pm 1/\log(n)$, we have $P_{\alpha_1}(p) = \Theta(P_{\alpha_2}(p)^{\alpha_1/\alpha_2})$ (see Lemma V.3). As a result, when estimating $P_\alpha(p)$, we can first estimate $P_{\alpha'}$ to provide a bound on P_α, where α', α differ by a $1 \pm 1/\log(n)$ factor and α' moves toward 1. We apply this strategy recursively on estimating $P_{\alpha'}$ until α' is very close to 1 from above when initial $\alpha > 1$ or from below when initial $\alpha < 1$, where a quantum speedup is already known. At a high level, we recursively estimate a sequence (of size $O(\log n)$) of such αs that eventually converges to 1, where in each iteration we establish some quantum speedup which leads to an overall quantum speedup. We remark that our approach is in spirit similar to the cooling schedules in simulated annealing (e.g. [49]). (See Section V.)

For *integer* $\alpha \geq 2$, we observe a connection between $P_\alpha(p)$ and the α-distinctness problem which leads to a more significant quantum speedup. Precisely, let $O_p : [S] \to [n]$ be the oracle in (I.7), we observe that $P_\alpha(p)$ is proportional to the α-*frequency moment* of $O_p(1), \ldots, O_p(S)$ which can be solved quantumly [50] based on any quantum algorithm for the α-distinctness problem (e.g., [42]). However, there is a catch that a direct application of [50] will lead to a dependence on S rather than n. We remedy this situation by tweaking the algorithm and its analysis in [50] to remove the dependence on S for our specific setting. (See Section VI.)

The integer α algorithm fails to extend to the *min-entropy* case (i.e., $\alpha = +\infty$) because the hidden constant in $O(\cdot)$ has a poor dependence on α (see Remark VI.1). Instead, we develop another reduction to the $\lceil \log n \rceil$-distinctness problem by exploiting the so-called "Poissonized sampling" technique [10, 13, 51]. At a high level, we construct Poisson distributions that are parameterized by p_is and leverage the "threshold" behavior of Poisson distributions (see Lemma VII.1). Roughly, if $\max_i p_i$ passes some threshold, with high probability, these parameterized Poisson distributions will lead to a collision of size $\lceil \log n \rceil$ that will be caught by the $\lceil \log n \rceil$-distinctness algorithm. Otherwise, we run again with a lower threshold until the threshold becomes trivial. (See Section VII.)

Some of our lower bounds come from reductions to existing ones in quantum query complexity, such as the quantum-classical separation of symmetric boolean functions [52], the collision problem [44, 53], and the Hamming weight problem [54], for different ranges of α. We also obtain lower bounds with a better error dependence by the polynomial method, which is inspired by the celebrated quantum lower bound for the collision problem [44, 53]. (See Section IX.)

Open questions. Our paper raises a few open questions. A natural question is to close the gaps between our quantum upper and lower bounds. Our quantum techniques on both ends are actually quite different from the state-of-the-art classical ones (e.g., [10]). It is interesting to see whether one can incorporate classical ideas to improve our quantum results. It is also possible to achieve better lower bounds by improving our application of the polynomial method or exploiting the quantum adversary method (e.g., [55, 56]). Finally, our result motivates the study of the quantum algorithm for the k-distinctness problem with super-constant k, which might also be interesting by itself.

Notations. Throughout the paper, we consider a discrete distribution $\{p_i\}_{i=1}^n$ on $[n]$, and $P_\alpha(p) := \sum_{i=1}^n p_i^\alpha$ represents the α-power sum of p. In the analyses of our algorithms, 'log' is natural logarithm; '$\approx$' omits lower order terms.

II. Master algorithm

Let $p = (p_i)_{i=1}^n$ be a discrete distribution on $[n]$ encoded by the quantum oracle $\hat{O}_p$ defined in (I.7). Inspired by [1] (BHH) and [2], we develop the following master algorithm to estimate a property F with the form $F(p) := \sum_{i \in [n]} p_i f(p_i)$ for a function $f\colon (0, 1] \to \mathbb{R}$.

Algorithm 1: Estimate $F(p) = \sum_i p_i f(p_i)$ of a discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$.

1 Set $l, M \in \mathbb{N}$;
2 **Regard the following subroutine as $\mathcal{A}$:**
3 Draw a sample $i \in [n]$ according to p ;
4 Use **EstAmp** or **EstAmp′** with M queries to obtain an estimation $\tilde{p}_i$ of p_i;
5 Output $X = f(\tilde{p}_i)$;
6 Use $\mathcal{A}$ for l executions in Theorem II.1 or Theorem II.2 and output $\tilde{F}(p)$ to estimate $F(p)$;

Here we draw the sample $i \in [n]$ following the distribution p in Line 3 by applying $\hat{O}_p$ to the uniform superposition $\frac{1}{\sqrt{S}} \sum_{s \in [S]} |s\rangle|0\rangle$ and measure the second register; see also our discussions at (I.6) and (I.7).

Comparing to BHH, we introduce a few new technical ingredients, which significantly improve the performance of Algorithm 1 especially for specific $f(\cdot)$s in our case, e.g., $f(p_x) = -\log(p_x)$ (Shannon entropy) and $f(p_x) = p_x^{\alpha-1}$ (Rényi entropy).

The **first** one is a generic quantum speedup of Monte Carlo methods [2], in particular, a quantum algorithm that approximates the output expectation of a subroutine with additive errors that has a quadratic better sample complexity than the one implied by Chebyshev's inequality.

Theorem II.1 (Additive error; Theorem 5 of [2])**.** *Let $\mathcal{A}$ be a quantum algorithm with output X such that $\mathrm{Var}[X] \leq \sigma^2$. Then for ϵ where $0 < \epsilon < 4\sigma$, by using $O((\sigma/\epsilon)\log^{3/2}(\sigma/\epsilon)\log\log(\sigma/\epsilon))$ executions of $\mathcal{A}$ and $\mathcal{A}^{-1}$,*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Algorithm 3 in [2] outputs an estimate $\tilde{\mathbb{E}}[X]$ *of* $\mathbb{E}[X]$ *such that*

$$\Pr\left[\left|\tilde{\mathbb{E}}[X]-\mathbb{E}[X]\right|\geq\epsilon\right]\leq 1/5. \tag{II.1}$$

It is worthwhile mentioning that classically one needs to use $\Omega(\sigma^2/\epsilon^2)$ executions of $\mathcal{A}$ [57] to estimate $\mathbb{E}[X]$. Theorem II.1 demonstrates a quadratic improvement on the error dependence. In the case of approximating $H_\alpha(p)$, we need to work with multiplicative errors while existing results (e.g. [2]) have a worse error dependence which is insufficient for our purposes. Instead, inspired by [2], we prove the following theorem (our **second** ingredient) that takes auxiliary information about the range of $\mathbb{E}[X]$ into consideration, which might be of independent interest.

Theorem II.2 (Multiplicative error; Appendix A)**.** *Let* $\mathcal{A}$ *be a quantum algorithm with output* X *such that* $\mathrm{Var}[X]\leq\sigma^2\mathbb{E}[X]^2$ *for a known* σ*. Assume that* $\mathbb{E}[X]\in[a,b]$*. Then for* ϵ *where* $0<\epsilon<24\sigma$*, by using* $\mathcal{A}$ *and* $\mathcal{A}^{-1}$ *for* $O((\sigma b/\epsilon a)\log^{3/2}(\sigma b/\epsilon a)\log\log(\sigma b/\epsilon a))$ *executions, Algorithm 10 (given in Appendix A) outputs an estimate* $\tilde{\mathbb{E}}[X]$ *of* $\mathbb{E}[X]$ *such that*

$$\Pr\left[\left|\tilde{\mathbb{E}}[X]-\mathbb{E}[X]\right|\geq\epsilon\mathbb{E}[X]\right]\leq 1/10. \tag{II.2}$$

The **third** ingredient is a fine-tuned error analysis due to the specific $f(\cdot)$s. Similar to BHH, we rely on quantum counting (named **EstAmp**) [48] to estimate the pre-image size of a Boolean function, which provides another source of quantum speedup. In particular, we approximate any probability p_x in the query model ((I.7)) by $\tilde{p}_x$ by estimating the size of the pre-image of a Boolean function $\chi\colon[S]\to\{0,1\}$ with $\chi(s)=1$ if $O(s)=i$ and $\chi(s)=0$ otherwise. However, for cases in BHH, it suffices to only consider the probability when p_x and $\tilde{p}_x$ are close, while in our case, we need to analyze the whole output distribution of quantum counting. Specifically, letting $t=\left|\chi^{-1}(1)\right|$ and $a=t/S=\sin^2(\omega\pi)$ for some ω, we have

Theorem II.3 ([48])**.** *For any* $k,M\in\mathbb{N}$*, there is a quantum algorithm (named* ***EstAmp****) with* M *quantum queries to* χ *that outputs* $\tilde{a}=\sin^2\left(\frac{l\pi}{M}\right)$ *for some* $l\in\{0,\ldots,M-1\}$ *such that*

$$\Pr\left[\tilde{a}=\sin^2\left(\frac{l\pi}{M}\right)\right]=\frac{\sin^2(M\Delta\pi)}{M^2\sin^2(\Delta\pi)}\leq\frac{1}{(2M\Delta)^2}, \tag{II.3}$$

where $\Delta=|\omega-\frac{l}{M}|$*. This promises* $|\tilde{a}-a|\leq 2\pi k\frac{\sqrt{a(1-a)}}{M}+k^2\frac{\pi^2}{M^2}$ *with probability at least* $\frac{8}{\pi^2}$ *for* $k=1$ *and with probability greater than* $1-\frac{1}{2(k-1)}$ *for* $k\geq 2$*. If* $a=0$ *then* $\tilde{a}=0$ *with certainty.*

Moreover, we also need to slightly modify **EstAmp** to avoid outputting $\tilde{p}_x=0$ in estimating Shannon entropy. This is because $f(\tilde{p}_x)=\log(\tilde{p}_x)$ is not well-defined at $\tilde{p}_x=0$. Let **EstAmp′** be the modified algorithm. It is required that **EstAmp′** outputs $\sin^2(\frac{\pi}{2M})$ when **EstAmp** outputs 0 and outputs **EstAmp**'s output otherwise.

By leveraging Theorem II.1 and Theorem II.2, and carefully setting parameters in Algorithm 1 according to Theorem II.3, we have the following corollaries that describe the complexity of estimating any $F(p)$.

Corollary II.1 (additive error)**.** *Given* $\epsilon>0$*. If* $l=\Theta\big(\big(\frac{\sigma}{\epsilon}\big)\log^{3/2}\big(\frac{\sigma}{\epsilon}\big)\log\log\big(\frac{\sigma}{\epsilon}\big)\big)$ *where* $\mathrm{Var}[X]\leq\sigma^2$ *and* M *is large enough such that* $\left|\mathbb{E}[X]-F(p)\right|\leq\epsilon$*, then Algorithm 1 approximates* $F(p)$ *with an additive error* ϵ *and success probability* $2/3$ *using* $O\big(M\cdot l\big)$ *quantum queries to* p*.*

Corollary II.2 (multiplicative error)**.** *Assume a procedure using* $C_{a,b}$ *quantum queries that returns an estimated range* $[a,b]$*, and that* $\mathbb{E}[X]\in[a,b]$ *with probability at least 0.9. Let* $l=\Theta\big(\big(\frac{\sigma b}{\epsilon a}\big)\log^{3/2}\big(\frac{\sigma b}{\epsilon a}\big)\log\log\big(\frac{\sigma b}{\epsilon a}\big)\big)$ *where* $\mathrm{Var}[X]/(\mathbb{E}[X])^2\leq\sigma^2$ *and* $\epsilon>0$*. For large enough* M *such that* $\left|\mathbb{E}[X]-F(p)\right|\leq\epsilon\mathbb{E}[X]$*, Algorithm 1 estimates* $F(p)$ *with a multiplicative error* ϵ *and success probability* $2/3$ *with* $O\big(M\cdot l+C_{a,b}\big)$ *queries.*

III. Shannon entropy estimation

We develop Algorithm 2 for Shannon entropy estimation with **EstAmp′** in Line 4, which provides quadratic quantum speedup in n.

Algorithm 2: Estimate the Shannon entropy of $p=(p_i)_{i=1}^n$ on $[n]$.

1 Set $l=\Theta\big(\frac{\log(n/\epsilon^2)}{\epsilon}\log^{3/2}\big(\frac{\log(n/\epsilon^2)}{\epsilon}\big)\log\log\big(\frac{\log(n/\epsilon^2)}{\epsilon}\big)\big)$;
2 **Regard the following subroutine as** $\mathcal{A}$**:**
3 Draw a sample $i\in[n]$ according to p;
4 Use **EstAmp′** with $M=2^{\lceil\log_2(\sqrt{n}/\epsilon)\rceil}$ queries to obtain an estimation $\tilde{p}_i$ of p_i;
5 Output $\tilde{x}_i=\log(1/\tilde{p}_i)$;
6 Use $\mathcal{A}$ for l executions in Theorem II.1 and output an estimation $\tilde{H}(p)$ of $H(p)$;

Theorem III.1. *Algorithm 2 approximates* $H(p)$ *within an additive error* $0<\epsilon\leq O(1)$ *with success probability at least* $\frac{2}{3}$ *using* $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^2}\big)$ *quantum queries to* p*.*

Proof. We prove this theorem in two steps. The **first** step is to show that the expectation of the subroutine $\mathcal{A}$'s output (denoted $\tilde{E}:=\sum_{i\in[n]}p_i\cdot\log(1/\tilde{p}_i)$) is close to $E:=\sum_{i\in[n]}p_i\cdot\log(1/p_i)=H(p)$. To that end, we divide $[n]$ into partitions based on the corresponding probabilities. Let $m=\lceil\log_2(\sqrt{n}/\epsilon)\rceil$ and $S_0=\{i:p_i\leq\sin^2(\pi/2^{m+1})\}$, $S_1=\{i:\sin^2(\pi/2^{m+1})<p_i\leq\sin^2(\pi/2^m)\}$, $S_2=\{i:\sin^2(\pi/2^m)<p_i\leq\sin^2(\pi/2^{m-1})\},\ldots,S_m=\{i:\sin^2(\pi/4)<p_i\leq\sin^2(\pi/2)\}$. For convenience, denote $s_0=|S_0|,s_1=|S_1|,\ldots,s_m=|S_m|$. Then

$$\sum_{j=0}^{m}s_j=n,\qquad\sum_{j=0}^{m}\frac{2^{2j}}{2^{2m}}s_j=\Theta(1). \tag{III.1}$$

Our main technical contribution is the following upper bound on the expected difference between $\log\tilde{p}_i$ and $\log p_i$ in terms of the partition S_i, $i=1,\cdots,n$:

$$\sum_{i\in S_j}p_i\mathbb{E}\big[\big|\log\tilde{p}_i-\log p_i\big|\big]\leq 7\cdot\frac{2^j s_j}{2^{2m}}\quad\forall\, j\in[m]. \tag{III.2}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

By linearity of expectation, we have

$$|\tilde{E}-E| \le \sum_{i\in[n]} p_i \mathbb{E}\big[\big|\log\tilde{p}_i - \log p_i\big|\big] \tag{III.3}$$

$$= \sum_{j=0}^{m}\sum_{i\in S_j} p_i \mathbb{E}\big[\big|\log\tilde{p}_i - \log p_i\big|\big] \tag{III.4}$$

$$\le 7\sum_{j=0}^{m}\frac{2^j s_j}{2^{2m}}. \tag{III.5}$$

As a result, by applying (III.1) and Cauchy-Schwartz inequality to (III.5), we have

$$|\tilde{E}-E| \le 7\sum_{j=0}^{m}\frac{2^j s_j}{2^{2m}} \tag{III.6}$$

$$\le 7\sqrt{\Big(\sum_{j=0}^{m}\frac{1}{2^{2m}}s_j\Big)\Big(\sum_{j=0}^{m}\frac{2^{2j}}{2^{2m}}s_j\Big)} \tag{III.7}$$

$$= O(\epsilon). \tag{III.8}$$

Because a constant overhead does not influence the query complexity, we may rescale Algorithm 2 by a large enough constant so that $|\tilde{E}-E|\le\epsilon/2$.

The **second** step is to bound the variance of the random variable, which is

$$\sum_{i\in[n]} p_i(\log\tilde{p}_i)^2 - \Big(\sum_{i\in[n]} p_i\log\tilde{p}_i\Big)^2 \le \sum_{i\in[n]} p_i(\log\tilde{p}_i)^2. \tag{III.9}$$

Since for any i, **EstAmp′** outputs $\tilde{p}_i$ such that $\tilde{p}_i \ge \sin^2(\frac{\pi}{2M}) \ge \frac{1}{M^2} \ge \frac{\epsilon^2}{4n}$, we have $\sum_{i\in[n]} p_i(\log\tilde{p}_i)^2 \le \sum_{i\in[n]} p_i\cdot\big(\log\frac{4n}{\epsilon^2}\big)^2 = \big(\log\frac{4n}{\epsilon^2}\big)^2$. As a result, by Corollary II.1 we can approximate $\tilde{E}$ up to additive error $\epsilon/2$ with failure probability at most $1/3$ using

$$O\Big(\frac{\log(n/\epsilon^2)}{\epsilon}\log^{3/2}\Big(\frac{\log(n/\epsilon^2)}{\epsilon}\Big)\log\log\Big(\frac{\log(n/\epsilon^2)}{\epsilon}\Big)\Big) \cdot 2^{\lceil\log_2(\sqrt{n}/\epsilon)\rceil} = \tilde{O}\Big(\frac{\sqrt{n}}{\epsilon^2}\Big) \tag{III.10}$$

quantum queries. Together with $|\tilde{E}-E|\le\epsilon/2$, Algorithm 2 approximates $E=H(p)$ up to additive error ϵ with failure probability at most $1/3$. □

It remains to prove (III.2). We prove:

Lemma III.1.

$$\sum_{i\in S_0} p_i\mathbb{E}\big[\big|\log\tilde{p}_i-\log p_i\big|\big] \le \frac{7s_0}{2^{2m}}. \tag{III.11}$$

For $j\in\{1,2,\dots,m\}$ in (III.2), the proof is similar because the dominating term has the angles of $\tilde{p}_i$ and p_i fall into the same interval of length $\frac{1}{2^m}$, and as a result $|\log\tilde{p}_i-\log p_i| = O(\frac{1}{2^j})$.

Proof of Lemma III.1. For convenience, denote $h(x) := x(\log t-\log x)$ where $0<t\le 1$ and $x\in(0,t]$. We notice that $h(x)\le t/e$: because $h'(x)=\log t-\log x-1$, when $x\in(0,t/e)$, $h'(x)>0$ hence $h(x)$ is an increasing function; when $x\in(t/e,t)$, $h'(x)<0$ hence $h(x)$ is a decreasing function; when $x=t/e$, $h'(x)=0$ and h reaches its maximum t/e.

Since $i\in S_0$, we can write $p_i=\sin^2(\theta_i\pi)$ where $0<\theta_i\le 1/2^{m+1}$. By Theorem II.3, for any $l\in\{1,\dots,2^{m-1}\}$, the output of **EstAmp′** when taking 2^m queries satisfies

$$\Pr\Big[\tilde{p}_i=\sin^2\Big(\frac{\pi}{2^{m+1}}\Big)\Big] = \frac{\sin^2(2^m\theta_i\pi)}{2^{2m}\sin^2(\theta_i\pi)} \le 1; \tag{III.12}$$

$$\Pr\Big[\tilde{p}_i=\sin^2\Big(\frac{l\pi}{2^m}\Big)\Big] = \frac{\sin^2(2^m(\frac{l}{2^m}-\theta_i)\pi)}{2^{2m}\sin^2((\frac{l}{2^m}-\theta_i)\pi)} \le \frac{1}{(2^{m+1}(\frac{l}{2^m}-\theta_i))^2}. \tag{III.13}$$

Combining (III.12), (III.13), and the property of function h discussed above, for any $i\in S_0$ we have

$$p_i\mathbb{E}\big[\big|\log\tilde{p}_i-\log p_i\big|\big] \le 1\cdot p_i\Big(\log\sin^2\Big(\frac{\pi}{2^{m+1}}\Big)-\log p_i\Big) + \sum_{l=1}^{2^{m-1}}\frac{p_i\big(\log\sin^2(\frac{l\pi}{2^m})-\log p_i\big)}{(2^{m+1}(\frac{l}{2^m}-\theta_i))^2} \tag{III.14}$$

$$\le \frac{\sin^2(\frac{\pi}{2^{m+1}})}{e} + \sum_{l=1}^{2^{m-1}}\frac{1}{(2l-1)^2}\cdot\sin^2\Big(\frac{\pi}{2^{m+1}}\Big)\cdot\Big(\log\sin^2\Big(\frac{l\pi}{2^m}\Big)-\log\sin^2\Big(\frac{\pi}{2^{m+1}}\Big)\Big) \tag{III.15}$$

$$\le \frac{\pi^2}{4e}\frac{1}{2^{2m}} + \frac{1}{2^{2m}}\cdot\frac{\pi^2}{4}\sum_{l=1}^{2^{m-1}}\frac{1}{(2l-1)^2}\log\Big(\frac{\sin(\frac{l\pi}{2^m})}{\sin(\frac{\pi}{2^{m+1}})}\Big)^2 \tag{III.16}$$

$$\le \frac{\pi^2}{4e}\frac{1}{2^{2m}} + \frac{1}{2^{2m}}\cdot\frac{\pi^2}{2}\sum_{l=1}^{2^{m-1}}\frac{\log 2l}{(2l-1)^2} \tag{III.17}$$

$$\le \frac{7}{2^{2m}}, \tag{III.18}$$

where (III.14) comes from (III.12) and (III.13), (III.15) comes from the properties of h (first term by $h(x)\le t/e$ and second term by the monotonicity of h on $(0,t/e)$), (III.16) holds because $\sin^2(\frac{\pi}{2^{m+1}})\le\frac{\pi^2}{2^{2m+2}}$, (III.17) holds because $\sin^2(\frac{l\pi}{2^m})\le 4l^2\sin^2(\frac{\pi}{2^{m+1}})$, and (III.18) holds because $\sum_{l=1}^{\infty}\frac{\log 2l}{(2l-1)^2}<1.2$ and $\frac{\pi^2}{4e}+\frac{1.2\pi^2}{2}<7$. Consequently,

$$\sum_{i\in S_0} p_i\mathbb{E}\big[\big|\log\tilde{p}_i-\log p_i\big|\big] \le \frac{7}{2^{2m}}\cdot s_0 = \frac{7s_0}{2^{2m}}. \tag{III.19}$$

□

IV. Application: KL divergence estimation

Classically, there does not exist any consistent estimator that guarantees asymptotically small error over the set of all pairs of distributions [58, 59]. These two papers then consider pairs of distributions with bounded probability ratios specified by a function $f:\mathbb{N}\to\mathbb{R}^+$, namely all pairs of distributions in the set as follows:

$$\mathcal{U}_{n,f(n)} := \big\{(p,q) : |p|=|q|=n, \frac{p_i}{q_i}\le f(n)\ \forall\, i\in[n]\big\}. \tag{IV.1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Denote the number of samples from p and q to be M_p and M_q, respectively. References [58, 59] shows that classically, $D_{\mathrm{KL}}(p\|q)$ can be approximated within constant additive error with high success probability if and only if $M_p = \Omega(\frac{n}{\log n})$ and $M_q = \Omega(\frac{nf(n)}{\log n})$.

Quantumly, we are given unitary oracles $\hat{O}_p$ and $\hat{O}_q$ defined by (I.7). Algorithm 3 below estimates the KL-divergence between p and q, which is similar to Algorithm 2 that uses **EstAmp′**, while adapts f to be mutually defined by p and q.

Algorithm 3: Estimate the KL divergence of $p = (p_i)_{i=1}^n$ and $q = (q_i)_{i=1}^n$ on $[n]$.

Set $l = \Theta\big(\frac{\log^2(\frac{nf(n)}{\epsilon^2})}{\epsilon}\log^{3/2}\big(\frac{\log^2(\frac{nf(n)}{\epsilon^2})}{\epsilon}\big)\log\log\big(\frac{\log^2(\frac{nf(n)}{\epsilon^2})}{\epsilon}\big)\big)$;

Regard the following subroutine as $\mathcal{A}$:

  Draw a sample $i \in [n]$ according to p;

  Use the modified amplitude estimation procedure **EstAmp′** with $2^{\lceil\log_2(\sqrt{n}/\epsilon)\rceil}$ and $2^{\lceil\log_2(\sqrt{n}f(n)/\epsilon)\rceil}$ quantum queries to p and q to obtain estimates $\tilde{p}_i$ and $\tilde{q}_i$, respectively;

  Output $\tilde{x}_i = \log\tilde{p}_i - \log\tilde{q}_i$;

Use $\mathcal{A}$ for l times in Theorem II.1 and outputs an estimation $\tilde{D}_{\mathrm{KL}}(p\|q)$ of $D_{\mathrm{KL}}(p\|q)$;

Theorem IV.1. *For $(p,q) \in \mathcal{U}_{n,f(n)}$, Algorithm 3 approximates $D_{KL}(p\|q)$ within an additive error $\epsilon > 0$ with success probability at least $\frac{2}{3}$ using $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^2}\big)$ quantum queries to p and $\tilde{O}\big(\frac{\sqrt{n}f(n)}{\epsilon^2}\big)$ quantum queries to q, where $\tilde{O}$ hides polynomials terms of $\log n$, $\log 1/\epsilon$, and $\log f(n)$.*

Proof. If the estimates $\tilde{p}_i$ and $\tilde{q}_i$ were precisely accurate, the expectation of the subroutine's output would be $E := \sum_{i\in[n]} p_i \cdot (\log p_i - \log q_i) = D_{\mathrm{KL}}(p\|q)$. On the one hand, we bound how far the actual expectation of the subroutine's output $\tilde{E}$ is from its exact value E. By linearity of expectation,

$$|\tilde{E} - E| \le \sum_{i\in[n]} p_i\mathbb{E}\big[\big|(\log\tilde{p}_i - \log p_i) + (\log\tilde{q}_i - \log q_i)\big|\big] \tag{IV.2}$$

$$\le \sum_{i\in[n]} p_i\mathbb{E}\big[\big|\log\tilde{p}_i - \log p_i\big|\big] + \sum_{i\in[n]} p_i\mathbb{E}\big[\big|\log\tilde{q}_i - \log q_i\big|\big] \tag{IV.3}$$

$$\le \sum_{i\in[n]} p_i\mathbb{E}\big[\big|\log\tilde{p}_i - \log p_i\big|\big] + f(n)\sum_{i\in[n]} q_i\mathbb{E}\big[\big|\log\tilde{q}_i - \log q_i\big|\big], \tag{IV.4}$$

where (IV.4) comes from the definition of $\mathcal{U}_{n,f(n)}$ in (IV.1). By the proof of Theorem III.1, in particular equation (III.8), $2^{\lceil\log_2(\sqrt{n}/\epsilon)\rceil}$ and $2^{\lceil\log_2(\sqrt{n}f(n)/\epsilon)\rceil}$ quantum queries to p and q give

$$\sum_{i\in[n]} p_i\mathbb{E}\big[\big|\log\tilde{p}_i - \log p_i\big|\big] = O(\epsilon) \tag{IV.5}$$

$$\sum_{i\in[n]} q_i\mathbb{E}\big[\big|\log\tilde{q}_i - \log q_i\big|\big] = O\Big(\frac{\epsilon}{f(n)}\Big), \tag{IV.6}$$

respectively. Plugging them into (IV.4) and rescaling Algorithm 3 by a large enough constant, we get $|\tilde{E} - E| \le \frac{\epsilon}{2}$.

On the other hand, the variance of the random variable is at most

$$\sum_{i\in[n]} p_i(\log\tilde{p}_i - \log\tilde{q}_i)^2 = \sum_{i:\,\tilde{p}_i<\tilde{q}_i} p_i(\log\tilde{q}_i - \log\tilde{p}_i)^2 + \sum_{i:\,\tilde{p}_i\ge\tilde{q}_i} p_i(\log\tilde{p}_i - \log\tilde{q}_i)^2. \tag{IV.7}$$

For the first term in (IV.7), because **EstAmp′** outputs $\tilde{p}_i$ such that $\tilde{p}_i \ge \sin^2(\frac{\pi}{2^{\lceil\log_2(\sqrt{n}/\epsilon)\rceil+1}}) \ge \frac{\epsilon^2}{4n}$ for any i, we have

$$\sum_{i:\,\tilde{p}_i<\tilde{q}_i} p_i(\log\tilde{q}_i - \log\tilde{p}_i)^2 \le \sum_{i:\,\tilde{p}_i<\tilde{q}_i} p_i\Big(\log 1 - \log\frac{\epsilon^2}{4n}\Big)^2 \le \Big(\log\frac{4n}{\epsilon^2}\Big)^2. \tag{IV.8}$$

For the second term in (IV.7), we have

$$\sum_{i:\,\tilde{p}_i\ge\tilde{q}_i} p_i(\log\tilde{p}_i - \log\tilde{q}_i)^2 \le \sum_{i:\,\tilde{p}_i\ge\tilde{q}_i} p_i(\log f(n))^2 \le (\log f(n))^2. \tag{IV.9}$$

Plugging (IV.8) and (IV.9) into (IV.7), the variance of the random variable is at most

$$\Big(\log\frac{4n}{\epsilon^2}\Big)^2 + (\log f(n))^2 = O\Big(\Big(\log\frac{nf(n)}{\epsilon^2}\Big)^2\Big). \tag{IV.10}$$

As a result, by Corollary II.1 we can approximate $\tilde{E}$ up to additive error $\epsilon/2$ with success probability at least $2/3$ using $\tilde{O}(\frac{1}{\epsilon})\cdot 2^{\lceil\log_2(\sqrt{n}/\epsilon)\rceil} = \tilde{O}\big(\frac{\sqrt{n}}{\epsilon^2}\big)$ quantum queries to p and $\tilde{O}(\frac{1}{\epsilon})\cdot 2^{\lceil\log_2(\sqrt{n}f(n)/\epsilon)\rceil} = \tilde{O}\big(\frac{\sqrt{n}f(n)}{\epsilon^2}\big)$ quantum queries to q, respectively. Together with $|\tilde{E} - E| \le \epsilon/2$, Algorithm 3 approximates $E = D_{\mathrm{KL}}(p\|q)$ up to additive error ϵ with success probability at least $2/3$. □

V. Non-integer Rényi entropy estimation

Recall the classical query complexity of non-integer and integer Rényi entropy estimations are different [20]. Quantumly, we also consider them separately; in this section, we consider α-Rényi entropy estimation for general non-integer $\alpha > 0$.

Let $P_\alpha(p) := \sum_{i=1}^n p_i^\alpha$. Since $H_\alpha(p) = \frac{1}{1-\alpha}\log P_\alpha(p)$, to approximate $H_\alpha(p)$ within an additive error $\epsilon > 0$ it suffices to approximate $P_\alpha(p)$ within a multiplicative error $e^{(\alpha-1)\epsilon} - 1 = \Theta(\epsilon)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Algorithm 4: Estimate the α-power sum $P_\alpha(p)$ of $p = (p_i)_{i=1}^n$ on $[n]$, $\alpha > 1, \alpha \notin \mathbb{N}$.

Regard the following subroutine as $\mathcal{A}$:
- Draw a sample $i \in [n]$ according to p;
- Use the amplitude estimation procedure **EstAmp** with $M = 2^{\lceil \log_2(\frac{\sqrt{n}}{\epsilon}\log(\frac{\sqrt{n}}{\epsilon}))\rceil+1}$ queries to obtain an estimate $\tilde{p}_i$ of p_i;
- Output $\tilde{x}_i = \tilde{p}_i^{\alpha-1}$;

1 Input parameters $(\alpha, \epsilon, \delta)$, where ϵ is the multiplicative error and δ is the failure probability;
2 **if** $\alpha < 1 + \frac{1}{\log n}$ **then**
3 Take $a = \frac{1}{e}$ and $b = 1$ as lower and upper bounds on $P_\alpha(p)$, respectively;
4 **else**
5 Recursively call Algorithm 4 with $\alpha' = \alpha(1 + \frac{1}{\log n})^{-1}$, $\epsilon = 1/4$, and $\delta = \frac{1}{12\log n \log \alpha}$ therein to give an estimate $\tilde{P}_{\alpha'}(p)$ of $P_{\alpha'}(p)$. For simplicity, denote $P := \tilde{P}_{\alpha'}(p)$. Take $a = \frac{(3P/4)^{1+\frac{1}{\log n}}}{e}$ and $b = \big(\frac{5P}{4}\big)^{1+\frac{1}{\log n}}$ as lower and upper bounds on $P_\alpha(p)$, respectively;
6 Set $l = \Theta\big(\frac{n^{\frac{1}{2}-\frac{1}{2\alpha}}}{\epsilon}\log^{3/2}(\frac{n^{\frac{1}{2}-\frac{1}{2\alpha}}}{\epsilon})\log\log(\frac{n^{\frac{1}{2}-\frac{1}{2\alpha}}}{\epsilon})\big)$;
7 Use $\mathcal{A}$ for l executions in Theorem II.2 using a and b as auxiliary information and output an estimation of $P_\alpha(p)$;
8 Run Line 1 to Line 7 for $\lceil 48\log\frac{1}{\delta}\rceil$ executions and take the median of all outputs in Line 7, denoted as $\tilde{P}_\alpha(p)$. Output $\tilde{P}_\alpha(p)$;

A. Case 1: $\alpha > 1, \alpha \notin \mathbb{N}$

We develop Algorithm 4 to approximate $P_\alpha(p)$ with a multiplicative error ϵ.

Theorem V.1. *The output of Algorithm 4 approximates $P_\alpha(p)$ within a multiplicative error $0 < \epsilon \le 1/4$ with success probability at least $1-\delta$ for some $\delta > 0$ using $\tilde{O}\big(\frac{n^{1-\frac{1}{2\alpha}}}{\epsilon^2}\big)$ quantum queries to p, where $\tilde{O}$ hides polynomials terms of $\log n$, $\log 1/\epsilon$, and $\log 1/\delta$.*

Proof of Theorem V.1. First, we design a subroutine $\mathcal{A}$ in Algorithm 4 to approximate $P_\alpha(p)$ following the same principle as in Algorithm 2. If the estimate $\tilde{p}_i$ in $\mathcal{A}$ were precisely accurate, its expectation would be $E := \sum_{i\in[n]} p_i \cdot p_i^{\alpha-1} = P_\alpha(p)$. To be precise, we bound how far the actual expectation of the subroutine's output $\tilde{E}$ is from the exact value $P_\alpha(p)$. In Lemma V.1, we show that when taking $M = 2^{\lceil \log_2(\frac{\sqrt{n}}{\epsilon}\log(\frac{\sqrt{n}}{\epsilon}))\rceil+1}$ queries in **EstAmp**, we have $|\tilde{E} - E| = O(\epsilon E)$.

As a result, to approximate $P_\alpha(p)$ within multiplicative error $\Theta(\epsilon)$, it is equivalent to approximate $\tilde{E}$ within multiplicative error $\Theta(\epsilon)$. Recall Theorem II.2 showed that if the variance of the random variable output by $\mathcal{A}$ is at most $\sigma^2\tilde{E}^2$ for a known σ, and if we can obtain two values a, b such that $\tilde{E} \in [a, b]$, then $\tilde{O}(\sigma b/\epsilon a)$ executions of $\mathcal{A}$ suffice to approximate $\tilde{E}$ within multiplicative error ϵ with success probability at least $9/10$. In the main body of the algorithm (Line 1 to Line 8), we use Theorem II.2 to approximate $\tilde{E}$.

On the one hand, in Lemma V.2, we show that for $\alpha > 1$ and large enough n, the variance is at most $5n^{1-1/\alpha}\tilde{E}^2$ with probability at least $\frac{8}{\pi^2}$. This gives $\sigma = \sqrt{5n^{1-1/\alpha}} = O(n^{1/2-1/2\alpha})$.

On the other hand, we need to compute the lower bound a and upper bound b. A key observation (Lemma V.3) is that for any $0 < \alpha_1 < \alpha_2$, we have

$$\Big(\sum_{i\in[n]} p_i^{\alpha_2}\Big)^{\frac{\alpha_1}{\alpha_2}} \le \sum_{i\in[n]} p_i^{\alpha_1} \le n^{1-\frac{\alpha_1}{\alpha_2}}\Big(\sum_{i\in[n]} p_i^{\alpha_2}\Big)^{\frac{\alpha_1}{\alpha_2}}. \quad \text{(V.1)}$$

Because $n^{1/\log n} = e$, if $\frac{\alpha_2}{\alpha_1} = 1 + O(\frac{1}{\log n})$, then

$$\sum_{i\in[n]} p_i^{\alpha_1} = \Theta\Big(\Big(\sum_{i\in[n]} p_i^{\alpha_2}\Big)^{\frac{\alpha_1}{\alpha_2}}\Big). \quad \text{(V.2)}$$

As a result, we compute a and b by recursively calling Algorithm 4 to estimate $P_{\alpha'}(p)$ for $\alpha' = \alpha/(1 + 1/\log n)$, which is used to compute the lower bound a and upper bound b in Line 5; the recursive call keeps until $\alpha < 1 + \frac{1}{\log n}$, when $a = \frac{1}{e}$ and $b = 1$ (as in Line 3) are simply lower and upper bounds on $P_\alpha(p)$ by (V.1).

To be precise, in Lemma V.4, we prove that $b/a < 4e = O(1)$, and with probability at least $1/e^{1/12} > 0.92$, a and b are indeed lower and upper bounds on $P_\alpha(p)$, respectively; furthermore, in Line 5, Algorithm 4 is recursively called by at most $\log n \log\alpha$ times, and each recursive call takes at most $\tilde{O}(n^{1-\frac{1}{2\alpha}})$ queries. This promises that when we apply Corollary II.2, the cost $C_{a,b}$ is dominated by the query cost from Algorithm 10.

Combining all points above, Corollary II.2 approximates $\tilde{E}$ up to multiplicative error $\Theta(\epsilon)$ with success probability at least $\frac{8}{\pi^2}\cdot 0.92 \cdot 9/10 > 2/3$ using

$$\log n\log\alpha\cdot\tilde{O}\Big(\frac{4e\cdot\sqrt{5n^{1-1/\alpha}}}{\epsilon}\Big)\cdot 2^{\lceil\log_2(\frac{\sqrt{n}}{\epsilon}\log(\frac{\sqrt{n}}{\epsilon}))\rceil+1} = \tilde{O}\Big(\frac{n^{1-1/2\alpha}}{\epsilon^2}\Big) \quad \text{(V.3)}$$

quantum queries. Together with $|\tilde{E} - E| = O(\epsilon E)$ and rescale l, M by a large enough constant, Line 1 to Line 7 in Algorithm 4 approximates $E = P_\alpha(p)$ up to multiplicative error ϵ with success probability at least $2/3$.

Finally, in Lemma V.5, we show that after repeating the procedure for $\lceil 48\log\frac{1}{\delta}\rceil$ executions and taking the median $\tilde{P}_\alpha(p)$ (as in Line 8), the success probability that $\tilde{P}_\alpha(p)$ approximates $P_\alpha(p)$ within multiplicative error ϵ is boosted to $1-\delta$. □

It remains to prove the lemmas mentioned above.

1) Expectation of $\mathcal{A}$ is ϵ-close to $P_\alpha(p)$:

Lemma V.1. $|\tilde{E} - E| = O(\epsilon E)$.

Proof of Lemma V.1. For convenience, denote $m = \lceil\log_2(\sqrt{n}/\epsilon\log(\sqrt{n}/\epsilon))\rceil + 1$, and $S_0, S_1, \ldots, S_m$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

the same as in Section III. We still have (III.1). By linearity of expectation,

$$|\tilde{E}-E| \le \sum_{i\in[n]} p_i \mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1}-p_i^{\alpha-1}\big|\big] \tag{V.4}$$

$$= \sum_{j=0}^{m}\sum_{i\in S_j} p_i \mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1}-p_i^{\alpha-1}\big|\big]. \tag{V.5}$$

Therefore, to prove $|\tilde{E}-E| = O(\epsilon E)$ it suffices to show

$$\sum_{j=0}^{m}\sum_{i\in S_j} p_i \mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1}-p_i^{\alpha-1}\big|\big] = O\Big(\epsilon \sum_{i\in[n]} p_i^{\alpha}\Big). \tag{V.6}$$

For each $i\in[n]$ we write $p_i = \sin^2(\theta_i\pi)$. Assume $k\in\mathbb{Z}$ such that $k \le 2^m\theta_i < k+1$. By Theorem II.3, for any $l \in \{1,2,\dots,\max\{k-1, 2^m-k-1\}\}$ the output of **EstAmp** taking 2^m queries satisfies

$$\Pr\Big[\tilde{p}_i = \sin^2\big(\frac{(k\pm(l+1))\pi}{2^m}\big)\Big] \le \frac{1}{4l^2}. \tag{V.7}$$

Furthermore, because $\sin\theta_i = \theta_i - O(\theta_i^3)$, $\cos\theta_i = 1-O(\theta_i^2)$, and $(1+\theta_i)^{2\alpha-1} = 1+(2\alpha-1)\theta_i + o(\theta_i)$,

$$\big(\sin((\theta_i+\frac{l}{2^m})\pi)\big)^{2(\alpha-1)} - \big(\sin(\theta_i\pi)\big)^{2(\alpha-1)} = O\Big(\frac{l}{2^m}(\theta_i\pi)^{2\alpha-3}\Big). \tag{V.8}$$

Combining (V.7), (V.8), and the fact that $\sum_{l=1}^{2^m}\frac{1}{l} = \Theta(m)$, we have

$$\sum_{j=0}^{m}\sum_{i\in S_j} p_i \mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1}-p_i^{\alpha-1}\big|\big] = O\Big(\sum_{j=0}^{m} s_j\cdot\big(\frac{2^j}{2^m}\pi\big)^2\cdot 2\sum_{l=1}^{2^m}\frac{1}{4l^2}\frac{l}{2^m}\big(\frac{2^j}{2^m}\pi\big)^{2\alpha-3}\Big) \tag{V.9}$$

$$= O\Big(\frac{\pi^{2\alpha-1}m}{2^{2\alpha m}}\cdot\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}\Big). \tag{V.10}$$

On the other side,

$$\epsilon\sum_{i\in[n]} p_i^{\alpha} = \Theta\Big(\epsilon\sum_{j=0}^{m} s_j\cdot\big(\frac{2^j}{2^m}\pi\big)^{2\alpha}\Big) = \Theta\Big(\frac{\epsilon\cdot\pi^{2\alpha}}{2^{2\alpha m}}\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big). \tag{V.11}$$

Therefore, to prove equation (V.6), by (V.10) and (V.11) it suffices to prove

$$\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j} = O\Big(\frac{\epsilon}{m}\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big). \tag{V.12}$$

Since $m = \lceil \log_2(\frac{\sqrt{n}}{\epsilon}\log(\frac{\sqrt{n}}{\epsilon}))\rceil + 1$, we have $\frac{2^m}{m} \ge \frac{\sqrt{n}}{\epsilon}$, thus $\frac{\epsilon}{m} \ge \frac{\sqrt{n}}{2^m}$. Therefore, it suffices to show

$$\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j} = O\Big(\frac{\sqrt{n}}{2^m}\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big). \tag{V.13}$$

If $\alpha \ge 3/2$, by Hölder's inequality we have

$$\Big(\sum_{j=0}^{m} s_j\Big)^{\frac{1}{2\alpha}}\Big(\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big)^{\frac{2\alpha-1}{2\alpha}} \ge \sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}. \tag{V.14}$$

By equation (III.1), this gives

$$n^{\frac{1}{2\alpha-1}}\Big(\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big) \ge \Big(\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}\Big)\Big(\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}\Big)^{\frac{1}{2\alpha-1}}. \tag{V.15}$$

By Hölder's inequality and also equation (III.1), we have

$$\Big(\sum_{j=0}^{m} s_j\Big)^{\frac{2\alpha-3}{2\alpha-1}}\Big(\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}\Big)^{\frac{2}{2\alpha-1}} \ge \sum_{j=0}^{m} s_j 2^{2j} = \Theta(2^{2m}). \tag{V.16}$$

This is equivalent to

$$n^{\frac{2\alpha-3}{2(2\alpha-1)}}\Big(\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}\Big)^{\frac{1}{2\alpha-1}} \ge \Theta(2^m). \tag{V.17}$$

Combining (V.15) and (V.17), we get exactly (V.13).

If $1<\alpha<3/2$, by Hölder's inequality we have

$$\Big(\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big)^{\frac{1}{\alpha}}\Big(\sum_{j=0}^{m} s_j\Big)^{\frac{\alpha-1}{\alpha}} \ge \sum_{j=0}^{m} s_j 2^{2j}; \tag{V.18}$$

$$\Big(\sum_{j=0}^{m} s_j 2^{2j}\Big)^{\frac{2\alpha-1}{2}}\Big(\sum_{j=0}^{m} s_j\Big)^{\frac{3-2\alpha}{2}} \ge \sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}. \tag{V.19}$$

By equation (III.1), the two inequalities above give

$$\sum_{j=0}^{m} s_j 2^{2\alpha j} \ge n^{1-\alpha}2^{2\alpha m} \tag{V.20}$$

$$\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j} \le n^{1.5-\alpha}2^{(2\alpha-1)m}, \tag{V.21}$$

which give (V.13). □

2) Bound the variance of $\mathcal{A}$ by the square of its expectation:

Lemma V.2. *With probability at least $\frac{8}{\pi^2}$, the variance of the random variable output by $\mathcal{A}$ is at most $5n^{1-1/\alpha}\tilde{E}^2$.*

Proof of Lemma V.2. The expectation and variance of the output by $\mathcal{A}$ are $\tilde{E} = \sum_{i=1}^{n} p_i\cdot\tilde{p}_i^{\alpha-1}$ and $\sum_{i\in[n]} p_i\cdot(\tilde{p}_i^{\alpha-1})^2 - \big(\sum_{i\in[n]} p_i\cdot\tilde{p}_i^{\alpha-1}\big)^2$, respectively. Therefore, it suffices to show that with probability at least $\frac{8}{\pi^2}$,

$$\sum_{i\in[n]} p_i\cdot(\tilde{p}_i^{\alpha-1})^2 \le 5n^{1-1/\alpha}\Big(\sum_{i=1}^{n} p_i\cdot\tilde{p}_i^{\alpha-1}\Big)^2. \tag{V.22}$$

By Theorem II.3, with probability at least $\frac{8}{\pi^2}$, we have

$$|\tilde{p}_i - p_i| \le \frac{2\pi\sqrt{p_i}}{2^m} \le \frac{\epsilon\pi\sqrt{p_i}}{\sqrt{n}} \qquad i\in[n]. \tag{V.23}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

For convenience, denote $p := p_{i^*}$ to be the maximal one among $p_1, \ldots, p_n$, i.e., $p = \max_{i\in\{1,\ldots,n\}} p_i$. We also denote $\tilde{p} := \tilde{p}_{i^*}$. Then we have

$$\frac{\left(\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}\right)^2}{\sum_{i\in[n]} p_i \cdot (\tilde{p}_i^{\alpha-1})^2} \geq \frac{\left(\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}\right)^2}{\tilde{p}^{\alpha-1} \cdot \sum_{i\in[n]} p_i \cdot \tilde{p}_i^{\alpha-1}} = \frac{\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}}{\tilde{p}^{\alpha-1}}. \tag{V.24}$$

Furthermore, because x^α is a convex function in $[0,1]$, by (V.23) and Jensen's inequality we have

$$\frac{\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}}{\tilde{p}^{\alpha-1}} = \frac{p \cdot \tilde{p}^{\alpha-1} + \sum_{i\neq i^*} p_i \cdot \tilde{p}_i^{\alpha-1}}{\tilde{p}^{\alpha-1}} \tag{V.25}$$

$$\gtrsim p + (1-p)\Big(\frac{1-p-\epsilon\pi\sqrt{1-p}}{np+\epsilon\pi\sqrt{np}}\Big)^{\alpha-1}. \tag{V.26}$$

Therefore, it suffices to show that for large enough n,

$$p + (1-p)\Big(\frac{1-p-\epsilon\pi\sqrt{1-p}}{np+\epsilon\pi\sqrt{np}}\Big)^{\alpha-1} \geq 0.2n^{-(1-1/\alpha)}. \tag{V.27}$$

If $p \geq 0.2n^{-(1-1/\alpha)}$, equation (V.27) directly follows. If $p < 0.2n^{-(1-1/\alpha)}$,

$$\begin{aligned}
&\lim_{n\to\infty} n^{1-1/\alpha} \cdot (1-p)\Big(\frac{1-p-\epsilon\pi\sqrt{1-p}}{np+\epsilon\pi\sqrt{np}}\Big)^{\alpha-1} \\
&= \lim_{n\to\infty} (1 - 0.2n^{-(1-1/\alpha)}) \\
&\quad \cdot \Big(\frac{n^{1/\alpha}(1-0.2n^{-(1-1/\alpha)}-\epsilon\pi)}{0.2n^{1/\alpha}+\sqrt{0.2}\epsilon\pi n^{1/2\alpha}}\Big)^{\alpha-1} &\text{(V.28)}\\
&= 1 \cdot \Big(\frac{1-\epsilon\pi}{0.2}\Big)^{\alpha-1} > 1 > 0.2, &\text{(V.29)}
\end{aligned}$$

where (V.29) is true because $\frac{1-\epsilon\pi}{0.2} > \frac{1-3.2/4}{0.2} = 1$. Because (V.26) only omits lower order terms and the limit in (V.29) is a constant larger than 0.2, Lemma V.2 follows. □

3) Give tight bounds on $P_\alpha(p)$ by $P_{\alpha'}(p)$:

Lemma V.3. *For any distribution $(p_i)_{i=1}^n$ and $0 < \alpha_1 < \alpha_2$, we have*

$$\Big(\sum_{i\in[n]} p_i^{\alpha_2}\Big)^{\frac{\alpha_1}{\alpha_2}} \leq \sum_{i\in[n]} p_i^{\alpha_1} \leq n^{1-\frac{\alpha_1}{\alpha_2}}\Big(\sum_{i\in[n]} p_i^{\alpha_2}\Big)^{\frac{\alpha_1}{\alpha_2}}. \tag{V.30}$$

Proof of Lemma V.3. On the one hand, by the generalized mean inequality, we have

$$\left(\frac{\sum_{i\in[n]} p_i^{\alpha_2}}{n}\right)^{\frac{1}{\alpha_2}} \geq \left(\frac{\sum_{i\in[n]} p_i^{\alpha_1}}{n}\right)^{\frac{1}{\alpha_1}}, \tag{V.31}$$

which gives the second inequality in (V.30).

On the other hand, since $\frac{\alpha_1}{\alpha_2} \leq 1$ and

$$0 \leq \frac{p_i^{\alpha_2}}{\sum_{j\in[n]} p_j^{\alpha_2}} \leq 1 \qquad \forall\, i \in [n], \tag{V.32}$$

we have

$$\frac{\sum_{i\in[n]} p_i^{\alpha_1}}{\left(\sum_{j\in[n]} p_j^{\alpha_2}\right)^{\frac{\alpha_1}{\alpha_2}}} = \sum_{i\in[n]} \left(\frac{p_i^{\alpha_2}}{\sum_{j\in[n]} p_j^{\alpha_2}}\right)^{\frac{\alpha_1}{\alpha_2}} \tag{V.33}$$

$$\geq \sum_{i\in[n]} \frac{p_i^{\alpha_2}}{\sum_{j\in[n]} p_j^{\alpha_2}} = 1, \tag{V.34}$$

which is equivalent to the first inequality in (V.30). □

4) Analyze the recursive calls:

Lemma V.4. *With probability at least 0.92, the a and b in Line 3 or Line 5 of Algorithm 4 are indeed lower and upper bounds on $P_\alpha(p)$, respectively, and $b/a = O(1)$; furthermore, in Line 5, Algorithm 4 is recursively called for at most $\log n \log\alpha$ executions, and each recursive call takes at most $\tilde{O}(n^{1-\frac{1}{2\alpha}})$ queries.*

Proof of Lemma V.4. We decompose the proof into two parts:

- **In Line 5, Algorithm 4 is recursively called for at most $\log n \log\alpha$ executions, and each recursive call takes at most $\tilde{O}(n^{1-\frac{1}{2\alpha}})$ queries:**
  Because each recursive call of Algorithm 4 reduces α by multiplying $(1+\frac{1}{\log n})^{-1}$ and the recursion ends when $\alpha < 1 + \frac{1}{\log n}$, the total number of recursive calls is at most $\frac{\log\alpha}{\log(1+\frac{1}{\log n})} \leq \log n \log\alpha$.

  When $\alpha < 1 + \frac{1}{\log n}$, a and b are set in Line 3 and no extra queries are needed; when Line 5 calls $\alpha(1+\frac{1}{\log n})^{-k}$-power sum estimation for some $k \in \mathbb{N}$, by induction on k, we see that this call takes at most $\tilde{O}\big(n^{1-\frac{(1+1/\log n)^k}{2\alpha}}\big) \leq \tilde{O}(n^{1-\frac{1}{2\alpha}})$ queries. As a result, when we apply Corollary II.2, the cost $C_{a,b}$ is dominated by the query cost from Algorithm 10.

- **With probability at least 0.92, a and b are lower and upper bounds on $P_\alpha(p)$ respectively, and $b/a = O(1)$:**
  When $1 < \alpha < 1 + \frac{1}{\log n}$, on the one hand we have $\sum_{i=1}^n p_i^\alpha \leq \sum_{i=1}^n p_i = 1$; on the other hand, because $n^{\frac{1}{\log n}} = e$, by Lemma V.3 we have

  $$\sum_{i=1}^n p_i^\alpha \geq \frac{\left(\sum_{i=1}^n p_i\right)^\alpha}{n^{\alpha-1}} \geq \frac{1}{e}. \tag{V.35}$$

  Therefore, $a = 1/e$ and $b = 1$ in Line 3 are lower and upper bounds on $P_\alpha(p)$ respectively, and $b/a = e = O(1)$.

  When $\alpha > 1 + \frac{1}{\log n}$, for convenience denote $\alpha' = \alpha(1+\frac{1}{\log n})^{-1}$. As justified above, the total number of recursive calls in Line 5 is at most $\log n \log\alpha$. Because we take $\delta = \frac{1}{12\log n \log\alpha}$ in Line 5, with probability at least

  $$\Big(1 - \frac{1}{12\log n\log\alpha}\Big)^{\log n\log\alpha} \geq \frac{1}{e^{1/12}} > 0.92, \tag{V.36}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

the output of every recursive call is within $1/4$-multiplicative error. As a result, the P in Line 5 satisfies $3P/4 \leq \sum_{i=1}^{n} p_i^{\alpha'} \leq 5P/4$. Combining this with Lemma V.3 and using $n^{\frac{1}{\log n}} = e$, we have

$$\frac{(3P/4)^{1+\frac{1}{\log n}}}{e} \leq \sum_{i=1}^{n} p_i^{\alpha} \leq \Big(\frac{5P}{4}\Big)^{1+\frac{1}{\log n}}. \tag{V.37}$$

In other words, a and b are indeed lower and upper bounds on $P_\alpha(p)$, respectively. Furthermore, $b/a = O(1)$ because

$$\frac{b}{a} = e \cdot \Big(\frac{5}{3}\Big)^{1+\frac{1}{\log n}} < 4e = O(1). \tag{V.38}$$

□

5) Boost the success probability:

Lemma V.5. *By repeating Line 1 to Line 7 in Algorithm 4 for $\lceil 48 \log \frac{1}{\delta} \rceil$ executions and taking the median $\tilde{P}_\alpha(p)$, the success probability is boosted to $1-\delta$.*

Proof of Lemma V.5. Denote the outputs after running Line 1 to Line 7 for $\lceil 48 \log \frac{1}{\delta} \rceil$ executions as $\tilde{P}_\alpha(p)^{(1)}, \ldots, \tilde{P}_\alpha(p)^{(\lceil 48 \log \frac{1}{\delta} \rceil)}$, respectively. Based on the correctness of Lemma V.1, Lemma V.2, and Lemma V.4, for each $i \in \{1, \ldots, \lceil 48 \log \frac{1}{\delta} \rceil\}$, with probability at least $2/3$ we have

$$|\tilde{P}_\alpha(p)^{(i)} - P_\alpha(p)| \leq \epsilon P_\alpha(p). \tag{V.39}$$

For each $i \in \{1, \ldots, \lceil 48 \log \frac{1}{\delta} \rceil\}$, denote X_i to be a Boolean random variable such that $X_i = 1$ if (V.39) holds, and $X_i = 0$ otherwise. Then $\Pr[X_i = 1] \geq 2/3$. Because in Line 8 the output $\tilde{P}_\alpha(p)$ is the median of all $\tilde{P}_\alpha(p)^{(1)}, \ldots, \tilde{P}_\alpha(p)^{(\lceil 48 \log \frac{1}{\delta} \rceil)}$, $|\tilde{P}_\alpha(p) - P_\alpha(p)| > \epsilon P_\alpha(p)$ leads to $\sum_{i=1}^{\lceil 48 \log \frac{1}{\delta} \rceil} X_i < \lceil 48 \log \frac{1}{\delta} \rceil / 2$. On the other hand, by Chernoff bound we have

$$\begin{aligned} &\Pr\Big[\sum_{i=1}^{\lceil 48 \log \frac{1}{\delta} \rceil} X_i < \frac{\lceil 48 \log \frac{1}{\delta} \rceil}{2} \Big] \\ &\quad \leq \exp\Big(- \frac{2/3 \lceil 48 \log \frac{1}{\delta} \rceil \cdot (1/4)^2}{2} \Big) \leq \delta. \end{aligned} \tag{V.40}$$

Therefore, with probability at least $1-\delta$, we have $|\tilde{P}_\alpha(p) - P_\alpha(p)| \leq \epsilon P_\alpha(p)$. □

B. Case 2: $0 < \alpha < 1$

When $0 < \alpha < 1$, our quantum algorithm follows the same structure as Algorithm 4:
The main difference is that, in the case $\alpha > 1$, Algorithm 4 makes α' smaller and smaller by multiplying $(1 + \frac{1}{\log n})^{-1}$ each time, whereas in the case $0 < \alpha < 1$, Algorithm 5 makes α' larger and larger by multiplying $(1 - \frac{1}{\log n})^{-1}$ each time; nevertheless, both recursions end when α' is close enough to 1. On the more technical level, they have different M in $\mathcal{A}$, different upper bounds on the variance of $\mathcal{A}$, and different expressions for a and b in Line 3 and Line 5.

Theorem V.2. *The output of Algorithm 5 approximates $P_\alpha(p)$ within a multiplicative error $0 < \epsilon \leq O(1)$ with success probability at least $1-\delta$ for some $\delta > 0$ using $\tilde{O}\big(\frac{n^{\frac{1}{\alpha}-\frac{1}{2}}}{\epsilon^2}\big)$ quantum queries to p, where $\tilde{O}$ hides polynomials terms of $\log n$, $\log 1/\epsilon$, and $\log 1/\delta$.*

Algorithm 5: Estimate the α-power sum $P_\alpha(p)$ of $p = (p_i)_{i=1}^n$ on $[n]$, $0 < \alpha < 1$.

Regard the following subroutine as $\mathcal{A}$:
- Draw a sample $i \in [n]$ according to p;
- Use the amplitude estimation procedure **EstAmp** with $M = 2^{\lceil \log_2(\frac{n^{1/2\alpha}}{\epsilon} \log(\frac{n^{1/2\alpha}}{\epsilon}))\rceil + 1}$ queries to obtain an estimate $\tilde{p}_i$ of p_i;
- Output $\tilde{x}_i = \tilde{p}_i^{\alpha-1}$;

1 Input parameters $(\alpha, \epsilon, \delta)$, where ϵ is the multiplicative error and δ is the failure probability;
2 **if** $\alpha > 1 - \frac{1}{\log n}$ **then**
3 Take $a = 1$ and $b = e$ as lower and upper bounds on $P_\alpha(p)$, respectively;
4 **else**
5 Recursively call Algorithm 5 with $\alpha' = \alpha(1 - \frac{1}{\log n})^{-1}$, $\epsilon = 1/2$, and $\delta = \frac{1}{12 \log n \log 1/\alpha}$ therein to give an estimate $\tilde{P}_{\alpha'}(p)$ of $P_{\alpha'}(p)$. For simplicity, denote $P := \tilde{P}_{\alpha'}(p)$. Take $a = (P/2)^{1-\frac{1}{\log n}}$ and $b = e(2P)^{1-\frac{1}{\log n}}$ as lower and upper bounds on $P_\alpha(p)$, respectively;
6 Set $l = \Theta\big(\frac{n^{\frac{1}{2\alpha}-\frac{1}{2}}}{\epsilon} \log^{3/2}(\frac{n^{\frac{1}{2\alpha}-\frac{1}{2}}}{\epsilon}) \log\log(\frac{n^{\frac{1}{2\alpha}-\frac{1}{2}}}{\epsilon})\big)$;
7 Use $\mathcal{A}$ for l executions in Theorem II.2 using a and b as auxiliary information and output an estimation of $P_\alpha(p)$;
8 Run Line 1 to Line 7 for $\lceil 48 \log \frac{1}{\delta} \rceil$ executions and take the median of all outputs in Line 7, denoted as $\tilde{P}_\alpha(p)$. Output $\tilde{P}_\alpha(p)$;

Before we give the formal proof of Theorem V.2, we compare the similarities and differences between Algorithm 4 and Algorithm 5, listed below:

- In both algorithms, the subroutine $\mathcal{A}$ has the same structure, and is designed to estimate $P_\alpha(p)$. However, to make the expectation of $\mathcal{A}$ ϵ-close to $P_\alpha(p)$, the **EstAmp** in Algorithm 4 suffices to take $M = 2^{\lceil \log_2(\frac{\sqrt{n}}{\epsilon} \log(\frac{\sqrt{n}}{\epsilon}))\rceil + 1}$ queries (see Lemma V.1), whereas the **EstAmp** in Algorithm 5 needs to take $M = 2^{\lceil \log_2(\frac{n^{1/2\alpha}}{\epsilon} \log(\frac{n^{1/2\alpha}}{\epsilon}))\rceil + 1}$ queries (see Lemma V.6);
- In both algorithms, we use Theorem II.2 to approximate the expectation of $\mathcal{A}$ (denoted $\tilde{E}$), hence they both need to upper-bound the variance of $\mathcal{A}$ by a multiple of $\tilde{E}^2$. However, technically the proofs are different, and we obtain different upper bounds in Lemma V.2 and Lemma V.7, respectively;
- Since both algorithms use Theorem II.2, they both need to compute a lower bound a and upper bound b on $P_\alpha(p)$. Both algorithms achieve this by observing Lemma V.3, and they both compute a and b by recursively call the estimation of $P_{\alpha'}(p)$ for some α' closer to 1. However, in the case $\alpha > 1$, Algorithm 4 makes α' smaller and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

smaller by multiplying $(1+\frac{1}{\log n})^{-1}$ each time, and ends the recursion when $\alpha' < 1+\frac{1}{\log n}$; in the case $0<\alpha<1$, Algorithm 5 makes α' larger and larger by multiplying $(1-\frac{1}{\log n})^{-1}$ each time, and ends the recursion when $\alpha' > 1-\frac{1}{\log n}$. This leads to different expressions for a and b in Line 3 and Line 5 of both algorithms, and technically the proofs for Lemma V.4 and Lemma V.8 is different;
- Both algorithms boost the success probability to $1-\delta$ by repeating the algorithm for $\lceil 48\log\frac{1}{\delta}\rceil$ executions and taking the median, and their correctness is both promised by Lemma V.5.

Proof of Theorem V.2. First, if the estimate $\tilde{p}_i$ in the subroutine $\mathcal{A}$ of Algorithm 5 were precisely accurate, the expectation of the subroutine's output would be $E := \sum_{i\in[n]} p_i \cdot p_i^{\alpha-1} = P_\alpha(p)$. To be precise, we bound how far the actual expectation of the subroutine's output $\tilde{E}$ is from the exact value $P_\alpha(p)$. In Lemma V.6, we show that when taking $M = 2^{\lceil \log_2(\frac{n^{1/2\alpha}}{\epsilon}\log(\frac{n^{1/2\alpha}}{\epsilon}))\rceil+1}$ queries in **EstAmp**, we have $|\tilde{E}-E| = O(\epsilon E)$.

As a result, to approximate $P_\alpha(p)$ within multiplicative error $\Theta(\epsilon)$, it is equivalent to approximate $\tilde{E}$ within multiplicative error $\Theta(\epsilon)$. Recall Theorem II.2 showed that if the variance of the random variable output by $\mathcal{A}$ is at most $\sigma^2\tilde{E}^2$ for a known σ, and if we can obtain two values a,b such that $\tilde{E}\in[a,b]$, then $\tilde{O}(\sigma b/\epsilon a)$ executions of $\mathcal{A}$ suffice to approximate $\tilde{E}$ within multiplicative error ϵ with success probability at least $9/10$. In the main body of the algorithm (Line 1 to Line 8), we use Theorem II.2 to approximate $\tilde{E}$.

On the one hand, in Lemma V.7, we show that for any $0<\alpha<1$, the variance is at most $2n^{1/\alpha-1}\tilde{E}^2$ with probability at least $\frac{8}{\pi^2}$. This gives $\sigma = \sqrt{2n^{1/\alpha-1}} = O(n^{1/2\alpha-1/2})$.

On the other hand, we need to compute the lower bound a and upper bound b. As stated in the proof of Theorem V.1, for any $0<\alpha_1<\alpha_2$ with $\frac{\alpha_2}{\alpha_1} = 1+O(\frac{1}{\log n})$,

$$\sum_{i\in[n]} p_i^{\alpha_1} = \Theta\Big(\Big(\sum_{i\in[n]} p_i^{\alpha_2}\Big)^{\frac{\alpha_1}{\alpha_2}}\Big). \tag{V.41}$$

As a result, we compute a and b by recursively calling Algorithm 5 to estimate $P_{\alpha'}(p)$ for $\alpha' = \alpha/(1-1/\log n)$, which is used to compute the lower bound a and upper bound b in Line 5; the recursive call keeps until $\alpha > 1-\frac{1}{\log n}$, when $a=1$ and $b=e$ (as in Line 3) are simply lower and upper bounds on $P_\alpha(p)$.

To be precise, in Lemma V.8, we prove that $b/a \le 4e = O(1)$, and with probability at least $1/e^{1/12} > 0.92$, a and b are indeed lower and upper bounds on $P_\alpha(p)$, respectively; furthermore, in Line 5, Algorithm 5 is recursively called by at most $\log n\log\frac{1}{\alpha}$ times, and each recursive call takes at most $\tilde{O}(n^{\frac{1}{\alpha}-\frac{1}{2}})$ queries. This promises that when we apply Corollary II.2, the cost $C_{a,b}$ is dominated by the query cost from Algorithm 10.

Combining all points above, Corollary II.2 approximates $\tilde{E}$ up to multiplicative error $\Theta(\epsilon)$ with success probability at least $\frac{8}{\pi^2}\cdot 0.92\cdot 9/10 > 2/3$ using

$$\log n\log\frac{1}{\alpha}\cdot\tilde{O}\Big(\frac{4e\cdot\sqrt{2n^{1/\alpha-1}}}{\epsilon}\Big) \cdot 2^{\lceil\log_2(\frac{n^{1/2\alpha}}{\epsilon}\log(\frac{n^{1/2\alpha}}{\epsilon}))\rceil+1} = \tilde{O}\Big(\frac{n^{\frac{1}{\alpha}-\frac{1}{2}}}{\epsilon^2}\Big) \tag{V.42}$$

quantum queries. Together with $|\tilde{E}-E| = O(\epsilon E)$ and rescale l, M by a large enough constant, Line 1 to Line 7 in Algorithm 5 approximates $E = P_\alpha(p)$ up to multiplicative error ϵ with success probability at least $2/3$.

Finally, following from Lemma V.5, after repeating the procedure for $\lceil 48\log\frac{1}{\delta}\rceil$ executions and taking the median $\tilde{P}_\alpha(p)$ (as in Line 8), the success probability that $\tilde{P}_\alpha(p)$ approximates $P_\alpha(p)$ within multiplicative error ϵ is boosted to $1-\delta$. □

It remains to prove the lemmas mentioned above.

1) Expectation of $\mathcal{A}$ is ϵ-close to $P_\alpha(p)$:

Lemma V.6. $|\tilde{E}-E| = O(\epsilon E)$.

Proof of Lemma V.6. For convenience, we denote $m = \lceil\log_2(\frac{n^{1/2\alpha}}{\epsilon}\log(\frac{n^{1/2\alpha}}{\epsilon}))\rceil+1$, and $S_0, S_1, \ldots, S_m$ the same as previous definitions. We still have (III.1). By linearity of expectation,

$$|\tilde{E}-E| \le \sum_{i\in[n]} p_i\mathbb{E}\Big[\Big|\frac{1}{\tilde{p}_i^{1-\alpha}} - \frac{1}{p_i^{1-\alpha}}\Big|\Big] \tag{V.43}$$

$$= \sum_{j=0}^{m}\sum_{i\in S_j} p_i\mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1} - p_i^{\alpha-1}\big|\big]. \tag{V.44}$$

Therefore, to prove $|\tilde{E}-E| = O(\epsilon E)$ it suffices to show

$$\sum_{j=0}^{m}\sum_{i\in S_j} p_i\mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1} - p_i^{\alpha-1}\big|\big] = O\Big(\epsilon\sum_{i\in[n]} p_i^\alpha\Big). \tag{V.45}$$

Similar to the proof of Lemma V.1, we have

$$\sum_{j=0}^{m}\sum_{i\in S_j} p_i\mathbb{E}\big[\big|\tilde{p}_i^{\alpha-1} - p_i^{\alpha-1}\big|\big] = O\Big(\frac{\pi^{2\alpha-1}m}{2^{2\alpha m}}\cdot\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j}\Big). \tag{V.46}$$

On the other side,

$$\epsilon\sum_{i\in[n]} p_i^\alpha = \Theta\Big(\epsilon\sum_{j=0}^{m} s_j\cdot\big(\frac{2^j}{2^m}\pi\big)^{2\alpha}\Big) = \Theta\Big(\frac{\epsilon\cdot\pi^{2\alpha}}{2^{2\alpha m}}\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big). \tag{V.47}$$

Therefore, to prove Equation (V.45), by (V.46) and (V.47) it suffices to prove

$$\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j} = O\Big(\frac{\epsilon}{m}\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big). \tag{V.48}$$

Since $m = \lceil\log_2(\frac{n^{1/2\alpha}}{\epsilon}\log(\frac{n^{1/2\alpha}}{\epsilon}))\rceil+1$, we have $\frac{2^m}{m} \ge \frac{n^{1/2\alpha}}{\epsilon}$, thus $\frac{\epsilon}{m} \ge \frac{n^{1/2\alpha}}{2^m}$. Therefore, it suffices to prove

$$\sum_{j=0}^{m} s_j 2^{(2\alpha-1)j} = O\Big(\frac{n^{1/2\alpha}}{2^m}\sum_{j=0}^{m} s_j 2^{2\alpha j}\Big). \tag{V.49}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Since $s_j \in \mathbb{N}$, $s_j \le s_j^{1/\alpha}$; as a result,

$$\frac{\sum_{j=0}^m s_j 2^{2j}}{\left(\sum_{j=0}^m s_j 2^{2\alpha j}\right)^{1/\alpha}} \le \frac{\sum_{j=0}^m (s_j 2^{2\alpha j})^{1/\alpha}}{\left(\sum_{j=0}^m s_j 2^{2\alpha j}\right)^{1/\alpha}} \tag{V.50}$$

$$= \sum_{j=0}^m \left(\frac{s_j 2^{2\alpha j}}{\sum_{k=0}^m s_k 2^{2\alpha k}}\right)^{1/\alpha} \tag{V.51}$$

$$\le \sum_{j=0}^m \frac{s_j 2^{2\alpha j}}{\sum_{k=0}^m s_k 2^{2\alpha k}} = 1. \tag{V.52}$$

Plugging (III.1) into the inequality above, we have

$$\left(\sum_{j=0}^m s_j 2^{2\alpha j}\right)^{\frac{1}{2\alpha}} = \Omega(2^m). \tag{V.53}$$

On the other side, by Hölder's inequality we have

$$\left(\sum_{j=0}^m s_j\right)^{\frac{1}{2\alpha}} \left(\sum_{j=0}^m s_j 2^{2\alpha j}\right)^{\frac{2\alpha-1}{2\alpha}} \ge \sum_{j=0}^m s_j 2^{(2\alpha-1)j}. \tag{V.54}$$

Combining (III.1), (V.53), and (V.54), we get exactly (V.49). □

2) Bound the variance of $\mathcal{A}$ by the square of its expectation:

Lemma V.7. *With probability at least $\frac{8}{\pi^2}$, the variance of the random variable output by $\mathcal{A}$ is at most $2n^{1/\alpha-1}\tilde{E}^2$.*

Proof of Lemma V.7. Because $\tilde{E} = \sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}$ and the variance is $\sum_{i=1}^n p_i \cdot (\tilde{p}_i^{\alpha-1})^2 - \left(\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}\right)^2 \le \sum_{i=1}^n p_i \cdot (\tilde{p}_i^{\alpha-1})^2$, it suffices to show that

$$\sum_{i=1}^n p_i \cdot (\tilde{p}_i^{\alpha-1})^2 \le 2n^{1/\alpha-1}\Big(\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}\Big)^2. \tag{V.55}$$

By Theorem II.3, with probability at least $\frac{8}{\pi^2}$, we have

$$|\tilde{p}_i - p_i| \le \frac{2\pi\sqrt{p_i}}{2^m} \le \frac{\epsilon\pi\sqrt{p_i}}{n^{1/2\alpha}} \qquad i \in [n]. \tag{V.56}$$

As a result,

$$\sum_{i=1}^n p_i(\tilde{p}_i^{\alpha-1})^2 \le \sum_{i=1}^n p_i\Big(p_i - \frac{\epsilon\pi\sqrt{p_i}}{n^{1/2\alpha}}\Big)^{-2(1-\alpha)} \tag{V.57}$$

$$= \sum_{i=1}^n p_i^{2\alpha-1}\Big(1 - \frac{\epsilon\pi}{n^{1/2\alpha}\sqrt{p_i}}\Big)^{-2(1-\alpha)} \tag{V.58}$$

$$\approx \sum_{i=1}^n p_i^{2\alpha-1}\Big(1 + 2(1-\alpha)\frac{\epsilon\pi}{n^{1/2\alpha}\sqrt{p_i}}\Big) \tag{V.59}$$

$$= \sum_{i=1}^n p_i^{2\alpha-1} + \frac{2(1-\alpha)\epsilon\pi}{n^{1/2\alpha}}\sum_{i=1}^n p_i^{2\alpha-0.5}. \tag{V.60}$$

Furthermore,

$$\sqrt{n}\Big(\sum_{i=1}^n p_i^{2\alpha-1}\Big) \ge \Big(\sum_{j=1}^n \sqrt{p_j}\Big)\Big(\sum_{i=1}^n p_i^{2\alpha-1}\Big) \tag{V.61}$$

$$\ge \sum_{i=j=1}^n \sqrt{p_j} p_i^{2\alpha-1} = \sum_{i=1}^n p_i^{2\alpha-0.5}. \tag{V.62}$$

Plugging this into (V.60), we have

$$\sum_{i=1}^n p_i(\tilde{p}_i^{\alpha-1})^2 \le \Big(1 + \frac{2(1-\alpha)\epsilon\pi}{n^{1/2\alpha-1/2}}\Big)\sum_{i=1}^n p_i^{2\alpha-1}. \tag{V.63}$$

Using similar techniques, we can show

$$\Big(\sum_{i=1}^n p_i \cdot \tilde{p}_i^{\alpha-1}\Big)^2 \ge \Big(1 - \frac{2(1-\alpha)\epsilon\pi}{n^{1/\alpha-1}}\Big)\Big(\sum_{i=1}^n p_i^\alpha\Big)^2. \tag{V.64}$$

Since $0 < \alpha < 1$,

$$\lim_{n\to\infty} 1 + \frac{2(1-\alpha)\epsilon\pi}{n^{1/2\alpha-1/2}} = 1, \quad \lim_{n\to\infty} 1 - \frac{2(1-\alpha)\epsilon\pi}{n^{1/\alpha-1}} = 1. \tag{V.65}$$

Because (V.59) only omits lower order terms and the limits in (V.65) are both 1, to prove (V.55) it suffices to prove that for large enough n,

$$\sum_{i=1}^n p_i^{2\alpha-1} \le n^{1/\alpha-1}\Big(\sum_{i=1}^n p_i^\alpha\Big)^2. \tag{V.66}$$

By generalized mean inequality, we have

$$\Big(\frac{1}{n}\sum_{i=1}^n p_i^{2\alpha-1}\Big)^{\frac{1}{2\alpha-1}} \le \Big(\frac{1}{n}\sum_{i=1}^n p_i^\alpha\Big)^{\frac{1}{\alpha}}. \tag{V.67}$$

Therefore,

$$\sum_{i=1}^n p_i^{2\alpha-1} \le n^{1-\frac{2\alpha-1}{\alpha}}\Big(\sum_{i=1}^n p_i^\alpha\Big)^{\frac{2\alpha-1}{\alpha}} \tag{V.68}$$

$$= n^{1/\alpha-1}\Big(\sum_{i=1}^n p_i^\alpha\Big)^{2-1/\alpha} \tag{V.69}$$

$$\le n^{1/\alpha-1}\Big(\sum_{i=1}^n p_i^\alpha\Big)^2. \tag{V.70}$$

Hence the result follows. □

3) Analyze the recursive calls:

Lemma V.8. *With probability at least 0.92, the a and b in Line 3 or Line 5 of Algorithm 5 are indeed lower and upper bounds on $P_\alpha(p)$, respectively, and $b/a = O(1)$; furthermore, in Line 5, Algorithm 5 is recursively called for at most $\log n \log \frac{1}{\alpha}$ executions, and each recursive call takes at most $\tilde{O}(n^{\frac{1}{\alpha}-\frac{1}{2}})$ queries.*

Proof of Lemma V.8. Similar to Lemma V.4, we decompose the proof into two parts:

- **In Line 5, Algorithm 5 is recursively called for at most $\log n \log \frac{1}{\alpha}$ executions, and each recursive call takes at most $\tilde{O}(n^{\frac{1}{\alpha}-\frac{1}{2}})$ queries:**
  Because each recursive call of Algorithm 5 increases α by multiplying $(1 - \frac{1}{\log n})^{-1}$ and the recursion ends when $\alpha > 1 - \frac{1}{\log n}$, the total number of recursive calls is at most $\frac{\log \alpha}{\log(1-\frac{1}{\log n})} \le \log n \log \frac{1}{\alpha}$.

  When $\alpha > 1 - \frac{1}{\log n}$, a and b are set in Line 3 and no extra queries are needed; when Line 5 calls $\alpha(1 - \frac{1}{\log n})^{-k}$-power sum estimation for some $k \in \mathbb{N}$, by induction on k, we see that this call takes at most

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

$\tilde{O}\big(n^{\frac{(1-1/\log n)^k}{\alpha}-\frac{1}{2}}\big) \le \tilde{O}(n^{\frac{1}{\alpha}-\frac{1}{2}})$ queries. As a result, when we apply Corollary II.2, the cost $C_{a,b}$ is dominated by the query cost from Algorithm 10.

- **With probability at least 0.92, a and b are lower and upper bounds on $P_\alpha(p)$ respectively, and $b/a = O(1)$:** When $1 - \frac{1}{\log n} < \alpha < 1$, on the one hand we have $\sum_{i=1}^n p_i^\alpha \ge \sum_{i=1}^n p_i = 1$; on the other hand, because $n^{\frac{1}{\log n}} = e$, by Lemma V.3 we have

$$\sum_{i=1}^n p_i^\alpha \le n^{1-\alpha}\Big(\sum_{i=1}^n p_i\Big)^\alpha \le e. \tag{V.71}$$

Therefore, $a = 1$ and $b = e$ in Line 3 are lower and upper bounds on $P_\alpha(p)$ respectively, and $b/a = e = O(1)$.

When $\alpha < 1 - \frac{1}{\log n}$, for convenience denote $\alpha' = \alpha(1-\frac{1}{\log n})^{-1}$. As justified above, the total number of recursive calls in Line 5 is at most $\log n \log \frac{1}{\alpha}$. Because we take $\delta = \frac{1}{12 \log n \log 1/\alpha}$ in Line 5, with probability at least

$$\Big(1 - \frac{1}{12\log n \log 1/\alpha}\Big)^{\log n \log 1/\alpha} \ge \frac{1}{e^{1/12}} > 0.92, \tag{V.72}$$

the output of every recursive call is within $1/2$-multiplicative error. As a result, the P in Line 5 satisfies $P/2 \le \sum_{i=1}^n p_i^{\alpha'} \le 2P$. Combining this with Lemma V.3 and using $n^{\frac{1}{\log n}} = e$, we have

$$(P/2)^{1-\frac{1}{\log n}} \le \sum_{i=1}^n p_i^\alpha \le e(2P)^{1-\frac{1}{\log n}}. \tag{V.73}$$

In other words, a and b are indeed lower and upper bounds on $P_\alpha(p)$, respectively. Furthermore, $b/a = O(1)$ because

$$\frac{b}{a} = e \cdot 4^{1-\frac{1}{\log n}} \le 4e = O(1). \tag{V.74}$$

□

VI. Integer Rényi entropy estimation

Recall the classical query complexity of α-Rényi entropy estimation for $\alpha \in \mathbb{N}, \alpha \ge 2$ is $\Theta(n^{1-1/\alpha})$ [20], which is smaller than non-integer cases. Quantumly, we also provide a more significant speedup.

Given the oracle $O_p\colon [S] \to [n]$ in (I.7), we denote the occurrences of $1, 2, \dots, n$ among $O_p(1), \dots, O_p(S)$ as $m_1, \dots, m_n$, respectively. A key observation is that by (I.6), we have

$$P_\alpha(p) = \sum_{i=1}^n (m_i/S)^\alpha = S^{-\alpha}\sum_{i=1}^n m_i^\alpha. \tag{VI.1}$$

Therefore, it suffices to approximate $\sum_{i\in[n]} m_i^\alpha$, which is known as the *α-frequency moment* of $O_p(1), \dots, O_p(S)$. Based on the quantum algorithm for α-distinctness [42], Montanaro [50] proved:

Fact VI.1 ([50], Step 3b-step 3e in Algorithm 2; Lemma 4)**.** *Fix l where $l \in \{1, \dots, n\}$. Let $s_1, \dots, s_l \in [S]$ be picked uniformly at random, and denote the number of α-wise collisions in $\{O_p(s_1), \dots, O_p(s_l)\}$ as $C(s_1, \dots, s_l)$. Then:*

- *$C(s_1, \dots, s_l)$ can be computed using $O(l^\nu \log(l/\epsilon^2))$ queries to $\hat{O}_p$ with failure probability at most $O(\epsilon^2/l)$, where $\nu := 1 - 2^{\alpha-2}/(2^\alpha - 1) < \frac{3}{4}$;*
- *$\mathbb{E}[C(s_1, \dots, s_l)] = \binom{l}{\alpha} P_\alpha(p)$ and $\mathrm{Var}[C(s_1, \dots, s_l)] = O(1)$.*

However, a direct application of [50] will lead to a complexity depending on S (in particular, l in Fact VI.1 can be as large as S) rather than n. Our solution is Algorithm 6 that is almost the same as Algorithm 2 in [50] except Line 1 and Line 2, where we set $2^{\lceil \log_2 \alpha n \rceil}$ as an upper bound on l. We claim that such choice of l is valid because by the pigeonhole principle, αn elements $O_p(s_1), \dots, O_p(s_{\alpha n})$ in $[n]$ must have an α-collision, so the first for-loop must terminate at some $i \le \lceil \log_2 \alpha n \rceil$. With this modification, we have Theorem VI.1 for integer Rényi entropy estimation.

Algorithm 6: Estimate the α-power sum $P_\alpha(p)$ of $p = (p_i)_{i=1}^n$ on $[n]$, $\alpha > 1, \alpha \in \mathbb{N}$.

1 Set $l = 2^{\lceil \log_2 \alpha n \rceil}$;
2 **for** $i = 0, \dots, \lceil \log_2 \alpha n \rceil$ **do**
3 Pick $s_1, \dots, s_{2^i} \in [S]$ uniformly at random and let $\mathcal{S}$ be the sequence $O_p(s_1), \dots, O_p(s_{2^i})$;
4 Apply the α-distinctness algorithm in [42] to $\mathcal{S}$ with failure probability $\frac{1}{10\lceil \log_2 \alpha n \rceil}$;
5 If it returns a set of α equal elements, set $l = 2^i$ and terminate the loop;
6 Set $M = \lceil K/\epsilon^2 \rceil$ for some $K = \Theta(1)$;
7 **for** $r = 1, \dots, M$ **do**
8 Pick $s_1, \dots, s_l \in [S]$ uniformly at random;
9 Apply the first bullet in Fact VI.1 to give an estimate $C^{(r)}$ of the number of α-wise collisions in $\{O_p(s_1), \dots, O_p(s_l)\}$;
10 Output $\tilde{P}_\alpha(p) = \frac{1}{M\binom{l}{\alpha}} \sum_{r=1}^M C^{(r)}$;

Theorem VI.1. *Assume $\alpha > 1, \alpha \in \mathbb{N}$. Algorithm 6 approximates $P_\alpha(p)$ within a multiplicative error $0 < \epsilon \le O(1)$ with success probability at least $\frac{2}{3}$ using $\tilde{O}\big(\frac{n^{\nu(1-1/\alpha)}}{\epsilon^2}\big) = o\big(\frac{n^{\frac{3}{4}(1-1/\alpha)}}{\epsilon^2}\big)$ quantum queries to p, where $\nu := 1 - 2^{\alpha-2}/(2^\alpha - 1) < \frac{3}{4}$.*

Our proof of Theorem VI.1 is inspired by the proof of Theorem 5 in [50].

Proof. Because O_p takes values in $[n]$, by pigeonhole principle, for any $s_1, \dots, s_{\alpha n} \in [S]$ there exists a α-wise collision among $O_p(s_1), \dots, O_p(s_{\alpha n})$. Therefore, Line 5 terminates the first loop with some $l \le 2^{\lceil \log_2 \alpha n \rceil}$ with probability at least $(1 - 1/10\lceil \log_2 \alpha n \rceil)^{\lceil \log_2 \alpha n \rceil} \ge e^{-1/10} > 0.9$.

Moreover, tighter bounds on l are established next. On the one hand, by Chebyshev's inequality and Fact VI.1, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

probability that the first for-loop fails to terminate when $l \le \frac{B}{P_\alpha(p)^{1/\alpha}}$ for some constant $B > 0$ is at most

$$\Pr\big[C(s_1,\dots,s_l)=0\big] \le \frac{\mathrm{Var}[C(s_1,\dots,s_l)]}{\mathbb{E}[C(s_1,\dots,s_l)]^2} \tag{VI.2}$$

$$= O\Big(\frac{1}{l^{2\alpha}P_\alpha(p)^2}\Big) \tag{VI.3}$$

$$= O\Big(\frac{1}{B^{2\alpha}}\Big). \tag{VI.4}$$

Therefore, taking a large enough B ensures that $l = O\big(\frac{1}{P_\alpha(p)^{1/\alpha}}\big)$ with failure probability at most $1/20$. On the other hand, by Markov's inequality and Fact VI.1, we have

$$\Pr\big[C(s_1,\dots,s_l)\ge 1\big] \le \mathbb{E}[C(s_1,\dots,s_l)] = O(l^\alpha P_\alpha(p)). \tag{VI.5}$$

As a result, the probability that the first for-loop terminates when $l \le \frac{A}{P_\alpha(p)^{1/\alpha}}$ for some constant $A > 0$ is at most

$$O(P_\alpha(p)) \cdot \sum_{i=0}^{\lfloor \log_2(\frac{A}{P_\alpha(p)^{1/\alpha}})\rfloor} 2^{i\alpha} = O(A^\alpha). \tag{VI.6}$$

Therefore, taking a small enough $A > 0$ ensures that $l = \Omega\big(\frac{1}{P_\alpha(p)^{1/\alpha}}\big)$ with failure probability at most $1/20$. In all, we have $l = \Theta\big(\frac{1}{P_\alpha(p)^{1/\alpha}}\big)$ with probability at least 0.9.

By Fact VI.1, the output $\mathbb{E}[\tilde{P}_\alpha(p)]$ in Line 10 of Algorithm 6 satisfies

$$\mathbb{E}[\tilde{P}_\alpha(p)] = \frac{1}{M\binom{l}{\alpha}}\sum_{r=1}^{M}\mathbb{E}[C^{(r)}] = P_\alpha(p), \tag{VI.7}$$

$$\mathrm{Var}[\tilde{P}_\alpha(p)] = \frac{1}{(M\binom{l}{\alpha})^2}\sum_{r=1}^{M}\mathrm{Var}[C^{(r)}] = O\Big(\frac{1}{Ml^{2\alpha}}\Big). \tag{VI.8}$$

Therefore, by Chebyshev's inequality and recall $l = \Theta\big(\frac{1}{P_\alpha(p)^{1/\alpha}}\big)$, we have

$$\Pr\big[|\tilde{P}_\alpha(p)-P_\alpha(p)|\ge \epsilon P_\alpha(p)\big] \le O\Big(\frac{1}{Ml^{2\alpha}\epsilon^2 P_\alpha(p)^2}\Big) = O\Big(\frac{1}{K}\Big). \tag{VI.9}$$

Taking a large enough constant K in Line 6 of Algorithm 6, we have $\Pr\big[|\tilde{P}_\alpha(p)-P_\alpha(p)|\le\epsilon P_\alpha(p)\big]\ge 0.9$. In all, with probability at least $0.9\times0.9\times0.9 > 2/3$, $\tilde{P}_\alpha(p)$ approximates $P_\alpha(p)$ within multiplicative error ϵ.

For the rest of the proof, it suffices to compute the quantum query complexity of Algorithm 6. Because the α-distinctness algorithm on m elements in [42] takes $O(m^\nu\log(1/\delta))$ quantum queries when the success probability is $1-\delta$, the first for-loop in Algorithm 6 takes $\sum_{i=0}^{\log_2 l} O(2^{\nu i}\log\lceil\log_2 \alpha n\rceil) = \tilde{O}(l^\nu) = \tilde{O}(n^{\nu(1-1/\alpha)})$ quantum queries because

$$l = \Theta\Big(\frac{1}{P_\alpha(p)^{1/\alpha}}\Big) = O(n^{1-1/\alpha}), \tag{VI.10}$$

following from $P_\alpha(p)\ge n^{1-\alpha}$. The second for-loop takes $\lceil K/\epsilon^2\rceil\cdot O(l^\nu\log(l/\epsilon^2)) = \tilde{O}(\frac{n^{\nu(1-1/\alpha)}}{\epsilon^2})$ quantum queries by Fact VI.1 and (VI.10). In total, the number of quantum queries is $\tilde{O}(\frac{n^{\nu(1-1/\alpha)}}{\epsilon^2})$. □

Remark VI.1. *In Theorem VI.1, we regard α as a constant, i.e., the query complexity $\tilde{O}(\frac{n^{\nu(1-1/\alpha)}}{\epsilon^2})$ hides the multiple in α. In fact, by analyzing the dependence on α carefully in the above proof, the query complexity of Algorithm 6 is actually*

$$\tilde{O}\Big(\alpha^{8\alpha^2}\cdot\frac{n^{\nu(1-1/\alpha)}}{\epsilon^2}\Big). \tag{VI.11}$$

The dependence on α is super-exponential; therefore, Algorithm 6 is not good enough to approximate min-entropy (i.e., $\alpha=\infty$). As a result, we give the quantum algorithm for estimating min-entropy separately (see Section VII).

VII. Min-entropy estimation

Since the min-entropy of p is $H_\infty(p) = -\log\max_{i\in[n]}p_i$ by (I.3), it is equivalent to approximate $\max_{i\in[n]}p_i$ within multiplicative error ϵ. We propose Algorithm 7 below to achieve this task.

Algorithm 7: Estimate $\max_{i\in[n]}p_i$ of a discrete distribution $p=(p_i)_{i=1}^n$ on $[n]$.

1 Set $\lambda = 1$;
2 **while** $\lambda \le n$ **do**
3 Take $M\sim\mathrm{Poi}(\frac{16\lambda\log n}{\epsilon^2})$. Pick $s_1,\dots,s_M\in[S]$ uniformly at random and let $\mathcal{S}$ be the sequence $O_p(s_1),\dots,O_p(s_M)$;
4 Apply a $\lceil\frac{16\log n}{\epsilon^2}\rceil$-distinctness quantum algorithm to $\mathcal{S}$ with failure probability at most $\frac{\epsilon}{2\log n}$;
5 If Line 4 outputs a $\lceil\frac{16\log n}{\epsilon^2}\rceil$-collision of elements $i^*\in[n]$, apply Theorem II.3 to approximate p_{i^*} with multiplicative error ϵ and output its result; if not, set $\lambda\leftarrow\lambda\cdot\sqrt{1+\epsilon}$ and jump to the start of the loop;
6 If $\lambda > n$ and no output has been given, output $1/n$;

A key property of the Poisson distribution is that if we take $M\sim\mathrm{Poi}(\nu)$ samples from p (as in Line 3), then for each $j\in[n]$, the number of occurrences of j in $O_p(s_1),\dots,O_p(s_M)$ follows the Poisson distribution $M_j\sim\mathrm{Poi}(\nu p_j)$, and $M_j, M_{j'}$ are independent for all $j\ne j'$. Furthermore:

Lemma VII.1. *Let $X\sim\mathrm{Poi}(\mu)$. Then, if $\mu < \frac{1}{\sqrt{1+\epsilon}}\cdot\frac{16\log n}{\epsilon^2}$, we have*

$$\Pr\Big[X\ge\frac{16\log n}{\epsilon^2}\Big]\le\frac{1}{n^2}; \tag{VII.1}$$

If $\mu\ge\frac{16\log n}{\epsilon^2}$, we have

$$\Pr\Big[X\ge\frac{16\log n}{\epsilon^2}\Big] > 0.15. \tag{VII.2}$$

Based on Lemma VII.1, our strategy is to set $\frac{16\log n}{\epsilon^2}$ as a threshold, take $\nu = \frac{16\lambda\log n}{\epsilon^2}$ as in Line 3, and gradually increase the parameter λ. For convenience, denote $p_{i^*} = \max_i p_i$. As long as $\nu\cdot p_{i^*} < \frac{16\log n}{\epsilon^2}$, with high probability there is no $\lceil\frac{16\log n}{\epsilon^2}\rceil$-collision in $\mathcal{S}$, the distinctness quantum algorithm in Line 4 rejects, and λ increases by multiplying $\sqrt{1+\epsilon}$ in Line 5; right after the first time when $\nu\cdot p_{i^*}\ge$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

$\frac{16\log n}{\epsilon^2}$, with probability at least 0.15, i^* has a $\lceil\frac{16\log n}{\epsilon^2}\rceil$-collision in $\mathcal{S}$, while all other entries in $[n]$ do not (with failure probability at most $1/n^2$). In this case, with probability at least $\Omega(1)$, the distinctness quantum algorithm in Line 4 captures i^*, and the quantum counting (Theorem II.3) in Line 5 computes p_{i^*} within multiplicative error ϵ.

Theorem VII.1. *Algorithm 7 approximates $\max_{i\in[n]} p_i$ within a multiplicative error $0<\epsilon\le 1$ with success probability at least $\Omega(1)$ using $\tilde{O}\big(Q(\lceil\frac{16\log n}{\epsilon^2}\rceil\text{-distinctness})\big)$ quantum queries to p, where $Q(\lceil\frac{16\log n}{\epsilon^2}\rceil\text{-distinctness})$ is the quantum query complexity of the $\lceil\frac{16\log n}{\epsilon^2}\rceil$-distinctness problem.*

We first prove Lemma VII.1. output

Proof of Lemma VII.1. First, we prove (VII.1)[7]. In [61], it is shown that if $\lambda>0$ and $X\sim\mathrm{Poi}(\lambda)$, then for any $\nu>1$ we have

$$\Pr[X\ge\nu\lambda]\le\frac{e^{-\lambda}\lambda^{\nu\lambda}}{(\nu\lambda)!(1-1/\nu)}. \tag{VII.3}$$

Taking $\lambda=\frac{1}{\sqrt{1+\epsilon}}\cdot\frac{16\log n}{\epsilon^2}$ and $\nu=\sqrt{1+\epsilon}$, by Sterling's formula we have

$$\Pr\Big[X\ge\frac{16\log n}{\epsilon^2}\Big]\le\frac{e^{-\lambda}\lambda^{\nu\lambda}}{(\nu\lambda)!(1-1/\nu)} \tag{VII.4}$$

$$\approx\frac{2}{\epsilon}\frac{e^{-\lambda}\lambda^{\nu\lambda}}{\sqrt{2\pi\nu\lambda}(\nu\lambda/e)^{\nu\lambda}} \tag{VII.5}$$

$$\approx\sqrt{\frac{2}{\pi}}\frac{1}{\epsilon\sqrt{\lambda}}\left(\frac{e^{\epsilon/2}}{(1+\frac{\epsilon}{2})^{1+\epsilon/2}}\right)^{\lambda}. \tag{VII.6}$$

Because

$$\lim_{\epsilon\to0}\left(\frac{e^{\epsilon/2}}{(1+\frac{\epsilon}{2})^{1+\epsilon/2}}\right)^{8/\epsilon^2}$$

$$=\lim_{\epsilon\to0}\exp\Big[\frac{4}{\epsilon}-\Big(\frac{8}{\epsilon^2}+\frac{4}{\epsilon}\Big)\ln\Big(1+\frac{\epsilon}{2}\Big)\Big] \tag{VII.7}$$

$$=\lim_{\epsilon\to0}\exp\Big[\frac{4}{\epsilon}-\Big(\frac{8}{\epsilon^2}+\frac{4}{\epsilon}\Big)\Big(\frac{\epsilon}{2}-\frac{\epsilon^2}{8}+O(\epsilon^3)\Big)\Big] \tag{VII.8}$$

$$=\lim_{\epsilon\to0}\exp[-1+O(\epsilon)]=e^{-1}, \tag{VII.9}$$

we have

$$\left(\frac{e^{\epsilon/2}}{(1+\frac{\epsilon}{2})^{1+\epsilon/2}}\right)^{\lambda}\approx e^{-\frac{\epsilon^2}{8}\lambda}\approx\frac{1}{n^2}. \tag{VII.10}$$

Plugging (VII.10) into (VII.6), we have

$$\Pr\Big[X\ge\frac{16\log n}{\epsilon^2}\Big]\lesssim\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{16\log n}}\frac{1}{n^2}\le\frac{1}{n^2}. \tag{VII.11}$$

Now we prove (VII.2). A theorem of Ramanujan [62, Question 294] states that for any positive integer M,

$$\frac{1}{2}e^M=\sum_{m=0}^{M-1}\frac{M^m}{m!}+\theta(M)\cdot\frac{M^M}{M!}, \tag{VII.12}$$

where $\frac{1}{3}\le\theta(M)\le\frac{1}{2}\ \forall\, M\in\mathbb{N}$. Because $\sum_{m=0}^{\infty}\frac{M^m}{m!}=e^M$, by (VII.12) we have

$$\sum_{m=M+1}^{\infty}\frac{M^m}{m!}+\frac{2}{3}\cdot\frac{M^M}{M!}\ge\frac{1}{2}e^M. \tag{VII.13}$$

By Stirling's formula, $M!\ge\sqrt{2\pi M}\big(\frac{M}{e}\big)^M$. As a result,

$$\sum_{m=M+1}^{\infty}\frac{M^m}{m!}\ge\frac{1}{2}e^M-\frac{2}{3}\cdot\frac{M^M}{\sqrt{2\pi M}\big(\frac{M}{e}\big)^M} \tag{VII.14}$$

$$=\Big(\frac{1}{2}-\frac{1}{\sqrt{4.5\pi M}}\Big)e^M. \tag{VII.15}$$

We take $M=\lfloor\frac{16\log n}{\epsilon^2}\rfloor$. By (VII.15), we have

$$\Pr\Big[X\ge\frac{16\log n}{\epsilon^2}\Big]=e^{-\frac{16\log n}{\epsilon^2}}\sum_{m=M+1}^{\infty}\frac{(\frac{16\log n}{\epsilon^2})^m}{m!} \tag{VII.16}$$

$$\ge e^{-M-1}\sum_{m=M+1}^{\infty}\frac{M^m}{m!} \tag{VII.17}$$

$$\ge\frac{1}{2e}-\frac{1}{e\sqrt{4.5\pi M}}. \tag{VII.18}$$

Because $0<\epsilon\le1$, we have $M\ge\lfloor16\log2\rfloor=11$. Therefore,

$$\Pr\Big[X\ge\frac{16\log n}{\epsilon^2}\Big]\ge\frac{1}{2e}-\frac{1}{e\sqrt{4.5\pi\cdot11}}>0.15. \tag{VII.19}$$

□

Proof of Theorem VII.1. Denote σ to be the permutation on $[n]$ such that $p_{\sigma(1)}\ge p_{\sigma(2)}\ge\cdots\ge p_{\sigma(n)}$. Without loss of generality, we assume that $p_{\sigma(2)}\le\frac{p_{\sigma(1)}}{1+\epsilon}$; otherwise, $p_{\sigma(2)}$ is close enough to $p_{\sigma(1)}$ in the sense that applying quantum counting to $p_{\sigma(2)}$ within multiplicative error ϵ gives an approximation to $p_{\sigma(1)}$ within multiplicative error 2ϵ. We may assume that every call of the $\lceil\frac{16\log n}{\epsilon^2}\rceil$-distinctness quantum algorithm in Line 4 of Algorithm 7 succeeds if and only if a $\lceil\frac{16\log n}{\epsilon^2}\rceil$-collision exists, because this happens with probability at least $\big(1-\frac{\epsilon}{2\log n}\big)^{\log_{\sqrt{1+\epsilon}}n}\ge e^{-1}=\Omega(1)$; for convenience, this is always assumed in the result of the proof.

On the one hand, when

$$\frac{p_{\sigma(1)}\cdot16\lambda\log n}{\epsilon^2}<\frac{1}{\sqrt{1+\epsilon}}\cdot\frac{16\log n}{\epsilon^2}, \tag{VII.20}$$

by Lemma VII.1 we have $\Pr\big[M_i\ge\frac{16\log n}{\epsilon^2}\big]\le\frac{1}{n^2}\ \forall\, i\in[n]$, where M_i is the occurences of i. Therefore, by the union bound, with probability at least $1-n\cdot\frac{1}{n^2}=1-\frac{1}{n}$, there is no $\frac{16\log n}{\epsilon^2}$-collision in $\mathcal{S}$. Since the while loop only has at most $\log_{\sqrt{1+\epsilon}}n=O(\frac{\log n}{\epsilon})$ rounds and $(1-\frac{1}{n})^{\log n/\epsilon}=1-o(1)$, we may assume that as long as (VII.20) holds, Line 4 of Algorithm 7 always has a negative output and Line 5 enforces $\lambda\leftarrow\lambda\cdot\sqrt{1+\epsilon}$ and jumps to the start of the while loop.

The while loop keeps iterating until (VII.20) is violated. In the second iteration after (VII.20) is violated, we have

$$\frac{16\log n}{\epsilon^2}\le\frac{p_{\sigma(1)}\cdot16\lambda\log n}{\epsilon^2}<\sqrt{1+\epsilon}\cdot\frac{16\log n}{\epsilon^2}; \tag{VII.21}$$

[7]The tail bound of Poisson distributions is also studied elsewhere, for example, in [60, Exercise 4.7].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

since $p_{\sigma(2)} \le \frac{p_{\sigma(1)}}{1+\epsilon}$, we have

$$\frac{p_{\sigma(2)} \cdot 16\lambda \log n}{\epsilon^2} < \frac{1}{\sqrt{1+\epsilon}} \frac{16 \log n}{\epsilon^2}. \tag{VII.22}$$

As a result, by Lemma VII.1 we have

$$\Pr\left[M_{\sigma(1)} \ge \frac{16 \log n}{\epsilon^2}\right] > 0.15; \tag{VII.23}$$

$$\Pr\left[M_i \ge \frac{16 \log n}{\epsilon^2}\right] \le \frac{1}{n^2} \quad \forall\, i \in [n]/\{\sigma(1)\}. \tag{VII.24}$$

Therefore, the probability that Line 4 outputs $\sigma(1)$ in the second iteration after (VII.20) is violated is at least $0.15 \cdot \left(1 - \frac{n-1}{n^2}\right)^{n-1}$. In the first iteration after (VII.20) is violated, we still have $\Pr\left[M_i \ge \frac{16 \log n}{\epsilon^2}\right] \le \frac{1}{n^2}\ \forall\, i \in [n]/\{\sigma(1)\}$. Therefore, the probability that Line 4 outputs $\sigma(1)$ in the first or second iteration after (VII.20) is violated is at least

$$0.15 \cdot \left(1 - \frac{n-1}{n^2}\right)^{n-1} \cdot \left(1 - \frac{n-1}{n^2}\right)^{n-1} \ge \frac{0.15}{e^2} = \Omega(1). \tag{VII.25}$$

In all, with probability $\Omega(1)$, Line 4 of Algorithm 7 outputs $\sigma(1)$ correctly in the first or second iteration after (VII.20) is violated; after that, the quantum counting in Line 5 approximates $p_{\sigma(1)} = \max_{i \in [n]} p_i$ within multiplicative error ϵ. This establishes the correctness of Algorithm 7.

It remains to show that the quantum query complexity of Algorithm 7 is $\tilde{O}\big(Q(\lceil \frac{16 \log n}{\epsilon^2} \rceil\text{-distinctness})\big)$. Because there are at most $\log_{\sqrt{1+\epsilon}} n = O(\frac{\log n}{\epsilon})$ iterations in the while loop, the $\lceil \frac{16 \log n}{\epsilon^2} \rceil$-distinctness algorithm in Line 4 is called for at most $O(\frac{\log n}{\epsilon})$ times; if it gives a $\lceil \frac{16 \log n}{\epsilon^2} \rceil$-collision, because $\max_{i \in [n]} p_i \ge 1/n$, the quantum query complexity caused by Line 5 is $O(\frac{\sqrt{n}}{\epsilon})$ by Theorem II.3, which is smaller than the $\Omega(n^{2/3})$ quantum lower bound on the distinctness problems [44]. As a result, the query complexity of Algorithm 7 in total is at most

$$\begin{aligned} &O\Big(\frac{\log n}{\epsilon}\Big) \cdot Q\Big(\Big\lceil \frac{16 \log n}{\epsilon^2} \Big\rceil\text{-distinctness}\Big) + O\Big(\frac{\sqrt{n}}{\epsilon}\Big) \\ &= \tilde{O}\Big(Q\Big(\Big\lceil \frac{16 \log n}{\epsilon^2} \Big\rceil\text{-distinctness}\Big)\Big). \end{aligned} \tag{VII.26}$$

□

Remark VII.1. *In some special cases, Algorithm 7 already demonstrates provable quantum speedup. Recall the state-of-the-art quantum algorithm for k-distinctness is [42] by Belovs, which has query complexity $O(2^{k^2} n^{1-2^{k-2}/(2^k-1)})$; however, this is superlinear when $k = \Theta(\log n)$. Nevertheless, if we are promised that $H_\infty(p) \le f(n)$ for some $f(n) = o(\sqrt{\log n})$, then we can replace the n in Line 2 of Algorithm 7 by $e^{f(n)}$ and replace every $\lceil \frac{16 \log n}{\epsilon^2} \rceil$ by $\lceil \frac{16 f(n)}{\epsilon^2} \rceil$, and it can be shown that the quantum query complexity of min-entropy estimation is $\tilde{O}\big(e^{(\frac{3}{4}+o(1)) \cdot f(n)}\big)$, whereas the best classical algorithm takes $\tilde{\Theta}(e^{f(n)})$ queries. In this case, we obtain a $(\frac{3}{4}+o(1))$-quantum speedup, but the classical query complexity is already small ($e^{\sqrt{\log n}} = n^{1/\sqrt{\log n}} = o(n^c)$ for any $c > 0$).*

VIII. 0-Rényi entropy estimation

Motivations. Estimating the support size of distributions (i.e., the 0-Rényi entropy) is also important in various fields, ranging from vocabulary size estimation [23, 24], database attribute variation [25], password and security [26], diversity study in microbiology [27–29], etc. The study of support estimation was initiated by naturalist Corbet in 1940s, who spent two years at Malaya for trapping butterflies and recorded how many times he had trapped various butterfly species. He then asked the leading statistician at that time, Fisher, to predict how many new species he would observe if he returned to Malaya for another two years of butterfly trapping. Fisher answered by alternatively putting plus or minus sign for the number of species that showed up one, two, three times, and so on, which was proven to be an unbiased estimator [63].

Formally, assuming n independent samples are drawn from an unknown distribution, the goal of [63] is to estimate the number of hitherto unseen symbols that would be observed if $t \cdot n$ (t being a pre-determined parameter) additional independent samples were collected from the same distribution. Reference [63] solved the case $t = 1$, which was later improved to $t \le 1$ [64] and $t = O(\log n)$ [47]; the last work also showed that $t = \Theta(\log n)$ is the largest possible range to give an estimator with provable guarantee.

However, such estimation always assumes n samples; a more natural question is, *can we estimate the support of a distribution per se?* Specifically, given a discrete distribution p over a finite set X where p_x denotes the probability of $x \in X$, can we estimate its *support*, defined by

$$\mathrm{Supp}(p) := |\{x : x \in X,\, p_x > 0\}|, \tag{VIII.1}$$

with high precision and success probability?

Unfortunately, this is impossible in general because elements with negligible but nonzero probability will be very unlikely to appear in the samples, while still contribute to $\mathrm{Supp}(p)$. As an evidence, $\mathrm{Supp}(p)$ is the exponent of the 0-Rényi entropy of p, but the sample complexity of α-Rényi entropy goes to infinity when $\alpha \to 0^+$ by Theorem IX.1, both classically and quantumly.

To circumvent this difficulty, two related properties have been considered *as an alternative to estimate 0-Rényi entropy*:

- *Support coverage*: $S_n(p) := \sum_{x \in X} \big(1 - (1-p_x)^n\big)$, the expected number of elements observed when taking n samples. To estimate $S_n(p)$ within $\pm\epsilon n$, [64] showed that $n/2$ samples from p suffices for any constant ϵ; recently, [65] improved the sample complexity to $O\big(\frac{n}{\log n}\big)$, and [47, 66] also considered the dependence in ϵ by showing that $\Theta\big(\frac{n}{\log n} \cdot \log \frac{1}{\epsilon}\big)$ is a tight bound, as long as $\epsilon = \Omega(n^{-0.2})$.
- *Support size*: $\mathrm{Supp}(p)$, under the assumption that for any $x \in X$, $p_x = 0$ or $p_x \ge 1/m$ for some given $m \in \mathbb{N}$. Reference [67] proposed the problem and gave a lower bound $\Omega(m^{1-o(1)})$, and [10] gave an upper bound $O\big(\frac{m}{\log m} \cdot \frac{1}{\epsilon^2}\big)$. Recently, [46] and [47] both proved that $\Theta\big(\frac{m}{\log m} \cdot \log^2 \frac{1}{\epsilon}\big)$ is the tight bound for the problem (both optimal in m and ϵ).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Problem	classical bounds	quantum bounds (this paper)
Support coverage	$\Theta\big(\frac{n}{\log n}\cdot\log\frac{1}{\epsilon}\big)$ [47, 66] $[\epsilon=\Omega(n^{-0.2})]$	$\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^{1.5}}\big)$, $\Omega\big(\frac{n^{1/3}}{\epsilon^{1/6}}\big)$
Support size	$\Theta\big(\frac{m}{\log m}\cdot\log^2\frac{1}{\epsilon}\big)$ [46, 47] $[\epsilon=\Omega(1/m)]$	$\tilde{O}\big(\frac{\sqrt{m}}{\epsilon^{1.5}}\big)$, $\Omega\big(\frac{m^{1/3}}{\epsilon^{1/6}}\big)$

TABLE II: Summary of the classical and quantum query complexity of support coverage and size estimation.

Quantumly, we give upper and lower bounds on both support coverage and support size estimation, summarized in Table II.

Support coverage estimation. We give the following upper bound on support coverage estimation; its lower bound is given in Proposition IX.2.

Algorithm 8: Estimate the support coverage $S_n(p)$.

1 **Regard the following subroutine as $\mathcal{A}$:**
2 Draw a sample $i\in X$ according to p;
3 Use **EstAmp** with $M=2^{\lceil\log_2(\sqrt{n/\epsilon})\rceil}$ queries to obtain an estimation $\tilde{p}_i$ of p_i;
4 Output $\tilde{x}_i=\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}$ if $\tilde{p}_i\neq 0$; otherwise, output n;
5 Use $\mathcal{A}$ for $\Theta\big(\frac{1}{\epsilon}\log^{3/2}\big(\frac{1}{\epsilon}\big)\log\log\big(\frac{1}{\epsilon}\big)\big)$ executions in Theorem II.1 and output an estimation $\tilde{S}_n(p)$ of $S_n(p)$;

Theorem VIII.1. *Algorithm 8 approximates* $\frac{S_n(p)}{n}:=\frac{\sum_{x\in X}(1-(1-p_x)^n)}{n}$ *within an additive error* $0<\epsilon\le O(1)$ *with success probability at least* $2/3$ *using* $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^{1.5}}\big)$ *quantum queries to* p.

Proof. We prove this theorem in two steps. The **first** step is to show that the expectation of the subroutine $\mathcal{A}$'s output (denoted $\tilde{E}:=\sum_{i\in X}p_i\cdot\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}$) satisfies $|\tilde{E}-E|=O(\epsilon n)$, where $E:=\sum_{i\in X}p_i\cdot\frac{1-(1-p_i)^n}{p_i}=S_n(p)$.

To achieve this, it suffices to prove that for each $i\in X$,

$$\mathbb{E}\Big[\Big|\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}-\frac{1-(1-p_i)^n}{p_i}\Big|\Big]=O(\epsilon n). \tag{VIII.2}$$

We write $p_i=\sin^2(\theta_i\pi)$. Assume $k\in\mathbb{Z}$ such that $k\le M\theta_i<k+1$. By Theorem II.3, for any $l\in\{1,2,\dots,\max\{k-1,M-k-1\}\}$, the output of **EstAmp** taking M queries satisfies

$$\Pr\Big[\tilde{p}_i=\sin^2\big(\frac{(k\pm(l+1))\pi}{M}\big)\Big]\le\frac{1}{4l^2}. \tag{VIII.3}$$

We first consider the case when $\tilde{p}_i>p_i$, and $\tilde{p}_i=\sin^2\big(\frac{(k+l+1)\pi}{M}\big)$ for some $l\in\mathbb{N}$. For convenience, denote $f(x)=\frac{1-(1-x)^n}{x}$ where $x\in(0,1]$. Because

$$f'(x)=\frac{nx(1-x)^{n-1}+(1-x)^n-1}{x^2} \tag{VIII.4}$$

$$\le\frac{nx+(1-nx)-1}{x^2}=0, \tag{VIII.5}$$

f is a decreasing function on $(0,1]$. Therefore,

$$\Big|\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}-\frac{1-(1-p_i)^n}{p_i}\Big|$$

$$\le\sin^2\frac{k\pi}{M}\cdot\frac{1-\cos^{2n}\frac{(k+l+1)\pi}{M}}{\sin^2\frac{(k+l+1)\pi}{M}}-\Big(1-\cos^{2n}\frac{k\pi}{M}\Big) \tag{VIII.6}$$

$$=\frac{\sin^2\frac{(k+l+1)\pi}{M}\cos^{2n}\frac{k\pi}{M}-\sin^2\frac{k\pi}{M}\cos^{2n}\frac{(k+l+1)\pi}{M}}{\sin^2\frac{(k+l+1)\pi}{M}}$$

$$+\frac{\sin^2\frac{k\pi}{M}-\sin^2\frac{(k+l+1)\pi}{M}}{\sin^2\frac{(k+l+1)\pi}{M}}. \tag{VIII.7}$$

By Taylor expansion, we have

$$\sin^2\frac{k\pi}{M}=\frac{k^2\pi^2}{M^2}+O\Big(\frac{k^6}{M^6}\Big) \tag{VIII.8}$$

$$\sin^2\frac{(k+l+1)\pi}{M}=\frac{(k+l+1)^2\pi^2}{M^2}+O\Big(\frac{(k+l)^6}{M^6}\Big), \tag{VIII.9}$$

and

$$\cos^{2n}\frac{k\pi}{M}=\Big(1-\frac{k^2\pi^2}{2M^2}+O\big(\frac{k^4}{M^4}\big)\Big)^{2n} \tag{VIII.10}$$

$$=\Big(1-\frac{k^2\pi^2\epsilon}{2n}+O\big(\frac{k^2\epsilon^2}{n^2}\big)\Big)^{2n} \tag{VIII.11}$$

$$=1-k^2\pi^2\epsilon+O\big(\epsilon^2k^2\big); \tag{VIII.12}$$

similarly

$$\cos^{2n}\frac{(k+l+1)\pi}{M}=1-(k+l+1)^2\pi^2\epsilon+O\big(\epsilon^2(k+l)^2\big). \tag{VIII.13}$$

Plugging (VIII.9), (VIII.12), and (VIII.13) into (VIII.7) and noticing that the tail in (VIII.9) has $1/M^6$, much smaller than that of (VIII.12) and (VIII.13), we have

$$\Big|\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}-\frac{1-(1-p_i)^n}{p_i}\Big|$$

$$=\frac{(k+l+1)^2(1-k^2\pi^2\epsilon)-k^2(1-(k+l+1)^2\pi^2\epsilon)}{(k+l+1)^2}$$

$$+\frac{k^2-(k+l+1)^2}{(k+l+1)^2}+O(\epsilon^2k^2)$$

$$-\frac{k^2}{(k+l+1)^2}O(\epsilon^2(k+l)^2) \tag{VIII.14}$$

$$=0+O\big(\epsilon^2(k+l)^2\big) \tag{VIII.15}$$

$$\le O(\epsilon n), \tag{VIII.16}$$

where (VIII.16) holds because $k+l\le M$ and $M=\Theta(\sqrt{n/\epsilon})$. Similarly, for the case $\tilde{p}_i<p_i$, we have

$$\Big|\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}-\frac{1-(1-p_i)^n}{p_i}\Big|\le O(\epsilon n). \tag{VIII.17}$$

In all, summing all $l\in\{1,2,\dots,\max\{k-1,M-k-1\}\}$ in cases $\tilde{p}_i>p_i$ and $\tilde{p}_i<p_i$ and by (VIII.3), the expectation of the deviation in (VIII.2) is at most

$$\mathbb{E}\Big[\Big|\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}-\frac{1-(1-p_i)^n}{p_i}\Big|\Big]$$

$$\le 2\cdot\sum_{l=1}^{M}\frac{1}{4l^2}\cdot O(\epsilon n)\le\frac{\pi^2}{3}\cdot O(\epsilon n)=O(\epsilon n). \tag{VIII.18}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Therefore, (VIII.2) follows and $|\tilde{E}-E| = O(\epsilon n)$. By rescaling M by a constant, without loss of generality we have $|\tilde{E}-E| \le \epsilon n/2$.

The **second** step is to bound the variance of the random variable, which is

$$\sum_{i\in X} p_i \cdot \Big(\frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}\Big)^2 - \Big(\sum_{i\in X} p_i \cdot \frac{1-(1-\tilde{p}_i)^n}{\tilde{p}_i}\Big)^2 \le \sum_{i\in X} p_i \cdot n^2 = n^2, \tag{VIII.19}$$

because $1-(1-\tilde{p}_i)^n \le n\tilde{p}_i$ by $0 \le \tilde{p}_i \le 1$. As a result of Theorem II.1, we can approximate $\tilde{E}$ up to additive error $\epsilon n/2$ with failure probability at most $1/3$ using

$$O\Big(\frac{n}{\epsilon n}\log^{3/2}\Big(\frac{n}{\epsilon n}\Big)\log\log\Big(\frac{n}{\epsilon n}\Big)\Big)\cdot 2^{\lceil \log_2(\sqrt{n/\epsilon})\rceil} = \tilde{O}\Big(\frac{\sqrt{n}}{\epsilon^{1.5}}\Big) \tag{VIII.20}$$

quantum queries. Together with $|\tilde{E}-E| \le \epsilon n/2$, Algorithm 8 approximates $E = S_n(p)$ up to additive error ϵn with failure probability at most $1/3$; in other words, Algorithm 8 approximates $E = S_n(p)/n$ up to ϵ with success probability at least $2/3$. □

Support size estimation. We give the following upper bound on support size estimation; its lower bound is given in Proposition IX.3.

Algorithm 9: Estimate $\mathrm{Supp}(p)$, under the promise that $p_x = 0$ or $p_x \ge 1/m$ for any $x \in X$.

1 Call Algorithm 8 with $n = \lceil m\log(2/\epsilon)\rceil$ and error $\frac{\epsilon}{2\log(2/\epsilon)}$, and denote the output as $\tilde{S}_n(p)$;
2 Denote $\widetilde{\mathrm{Supp}}(p) := \lceil \tilde{S}_n(p)\rceil$. Output $\widetilde{\mathrm{Supp}}(p)$ as an estimation of $\mathrm{Supp}(p)$;

Theorem VIII.2. *Under the promise that for any $x \in X$, $p_x = 0$ or $p_x \ge 1/m$, Algorithm 9 approximates $\mathrm{Supp}(p)/m$ within an additive error $0 < \epsilon \le O(1)$ with success probability at least $2/3$ using $\tilde{O}\big(\frac{\sqrt{m}}{\epsilon^{1.5}}\big)$ quantum queries to p.*

Proof. For convenience, denote $X_{1/m} := \{x \in X : p_x \ge 1/m\}$. Then $\mathrm{Supp}(p) = |X_{1/m}|$ by the promise, and

$$1-\frac{\epsilon}{2} \le 1-(1-p_x)^n \le 1 \qquad \forall\, x \in X_{1/m}; \tag{VIII.21}$$

$$1-(1-p_x)^n = 0 \qquad \forall\, x \notin X_{1/m}; \tag{VIII.22}$$

As a result,

$$S_n(p) = \sum_{x\in X} 1-(1-p_x)^n \in \Big[\big(1-\frac{\epsilon}{2}\big)\,\mathrm{Supp}(p), \mathrm{Supp}(p)\Big]. \tag{VIII.23}$$

Furthermore, by the correctness of Algorithm 8, with probability at least $2/3$ we have

$$|\tilde{S}_n(p) - S_n(p)| \le \frac{\epsilon}{2\log(2/\epsilon)}\cdot n = \frac{m\epsilon}{2}. \tag{VIII.24}$$

Together with (VIII.23),

$$\tilde{S}_n(p) \in \Big[\big(1-\frac{\epsilon}{2}\big)\,\mathrm{Supp}(p) - \frac{m\epsilon}{2}, \mathrm{Supp}(p) + \frac{m\epsilon}{2}\Big] \subseteq \Big[\mathrm{Supp}(p) - m\epsilon, \mathrm{Supp}(p) + \frac{m\epsilon}{2}\Big]. \tag{VIII.25}$$

Therefore, with probability at least $2/3$, $\frac{\lceil \tilde{S}_n(p)\rceil}{m}$ approximates $\frac{\mathrm{Supp}(p)}{m}$ up to ϵ with success probability at least $2/3$. □

IX. Quantum lower bounds

In this section, we prove Theorem I.2, which is rewritten below:

Theorem IX.1. *Any quantum algorithm that approximates $H_\alpha(p)$ of distribution p on $[n]$ within additive error ϵ with success probability at least 2/3 must use*

- $\Omega(\sqrt{n}+n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ *quantum queries when $\alpha = 0$, assuming $1/n \le \epsilon \le 1/12$.*
- $\tilde{\Omega}(n^{\frac{1}{7\alpha}-o(1)}/\epsilon^{\frac{2}{7}})$ *quantum queries when $0 < \alpha < \frac{3}{7}$.*
- $\Omega(n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ *quantum queries when $\frac{3}{7} \le \alpha \le 3$ and $\alpha \ne 1$, assuming $1/n \le \epsilon \le 1/2$.*
- $\Omega(\sqrt{n}+n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ *quantum queries when $\alpha = 1$, assuming $1/n \le \epsilon \le 1/2$.*
- $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}}/\epsilon)$ *quantum queries when $3 \le \alpha < \infty$.*
- $\Omega(\sqrt{n}/\epsilon)$ *quantum queries when $\alpha = \infty$.*

Because we use different techniques for different ranges of α, we divide the proofs into three categories.

A. Reduction from classical lower bounds ($0 < \alpha < \frac{3}{7}$)

We prove that the quantum lower bound when $0 < \alpha < \frac{3}{7}$ is indeed $\Omega(n^{\frac{1}{7\alpha}-o(1)}/\epsilon^{\frac{2}{7}})$, as claimed in Theorem IX.1.

Proof. First, by [20], we know that $\Omega(n^{\frac{1}{\alpha}-o(1)}/\epsilon^2)$ is a lower bound on the classical query complexity of α-Rényi entropy estimation. On the other hand, reference [52] shows that for any problem that is invariant under permuting inputs and outputs and that has sufficiently many outputs, the quantum query complexity is at least the seventh root of the classical randomized query complexity (up to poly-logarithmic factors). Our query oracle $O_p\colon [S] \to [n]$ has n outputs with tend to infinity when n is large; the distribution p is invariant under permutations on $[S]$ since $p_i = |\{s \in [S] : O_p(s) = i\}|/S$ is invariant for all i; Rényi entropy is invariant under permutations on $[n]$ since it does not depend on the order of p_i. Therefore, our problem satisfies the requirements from [52], and $\Omega(n^{\frac{1}{7\alpha}-o(1)}/\epsilon^{\frac{2}{7}})$ is a lower bound on the quantum query complexity of α-Rényi entropy estimation. □

B. Exploitation of the collision lower bound ($\alpha = 0$ and $\frac{3}{7} \le \alpha \le 3$)

We prove lower bounds on entropy estimation by further exploiting the famous collision lower bound [44, 53]. First, we define the following problem:

Definition IX.1 (l-pairs distinctness)**.** *Given positive integers n and l such that $1 \le l \le n/2$, and a function $f\colon [n] \to [n]$. Under the promise that either f is 1-to-1 or their exists l*

pairwise different pairs $(x_{i_1}, y_{i_1}), \ldots, (x_{i_l}, y_{i_l}) \in [n] \times [n]$ *such that* $x_{i_j} \neq y_{i_j}$ *but* $f(x_{i_j}) = f(y_{i_j})$ *for all* $j \in [l]$*, the* l*-pairs distinctness problem is to determine which is the case, with success probability at least* $2/3$*.*

Note that when $l = 1$, l-pairs distinctness reduces to the element distinctness problem, whose quantum query complexity is $\Theta(n^{2/3})$ [41, 44]; when $l = n/2$, l-pairs distinctness reduces to the collision problem, whose quantum query complexity is $\Theta(n^{1/3})$ [44, 53]. Inspired by the reduction from the collision lower bound to the element distinctness lower bound in [44], we prove a more general quantum lower bound for l-pairs distinctness:

Proposition IX.1. *The quantum query complexity of* l*-pairs distinctness is at least* $\Omega(n^\alpha)$*, where* $l^\alpha = n^{\frac{2}{3}-\alpha}$*.*

Proof. Assume the contrary that the quantum query complexity of l-pairs distinctness is $o(n^\alpha)$. Consider a function $f\colon [n] \to [n]$ that is promised to be either 1-to-1 or 2-to-1. By [53], it takes $\Omega(n^{1/3})$ quantum queries to decide whether f is 1-to-1 or 2-to-1.

Denote $\mathcal{S}$ to be a subset of $[n]$, where $|\mathcal{S}| = \lceil 2\sqrt{nl} \rceil$ and the elements in $\mathcal{S}$ are chosen uniformly at random. If f is 1-to-1, then f restricted on $\mathcal{S}$, denoted $f|_{\mathcal{S}}$, is still 1-to-1 on $\mathcal{S}$. If f is 2-to-1, denote the set of its images as $\{a_1, \ldots, a_{n/2}\}$. For any $j \in [n/2]$, denote X_j to be a binary random variable that equals to 1 when the collision pair of a_j appears in $\mathcal{S}$, and equals to 0 otherwise. Then for any $j, k \in [n], j \neq k$,

$$\Pr[X_j = 1] = \frac{\binom{|\mathcal{S}|}{2}}{\binom{n}{2}}, \quad \Pr[X_j X_k = 1] = \frac{\binom{|\mathcal{S}|}{4}}{\binom{n}{4}}. \tag{IX.1}$$

Denote $X = \sum_{j=1}^{n/2} X_j$, which is the number of collision pairs in $\mathcal{S}$. By linearity of expectation,

$$\mathbb{E}[X] = \frac{n}{2} \cdot \frac{|\mathcal{S}|(|\mathcal{S}|-1)}{n(n-1)} \gtrsim 2l. \tag{IX.2}$$

On the other hand,

$$\begin{aligned} \mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 && \text{(IX.3)} \\ &= \sum_{j=1}^{n/2} \mathbb{E}[X_j] + \sum_{j \neq k} \mathbb{E}[X_j X_k] - \mathbb{E}[X]^2 && \text{(IX.4)} \\ &\leq \frac{n}{2} \cdot \frac{\binom{|\mathcal{S}|}{2}}{\binom{n}{2}} + \frac{n}{2}\Big(\frac{n}{2} - 1\Big) \cdot \frac{\binom{|\mathcal{S}|}{4}}{\binom{n}{4}} - \frac{n^2}{4} \cdot \frac{\binom{|\mathcal{S}|}{2}^2}{\binom{n}{2}^2} && \text{(IX.5)} \\ &\lesssim 2l. && \text{(IX.6)} \end{aligned}$$

Therefore, by Chebyshev's inequality,

$$\Pr[X < l] \leq \Pr\big[X \leq \mathbb{E}[X] - 2\sqrt{2l}\big] \leq 1/4. \tag{IX.7}$$

In other words, with probability at least $3/4$, $f|_{\mathcal{S}}$ on $\mathcal{S}$ has at least l collision pairs. By our assumption, it takes $o(|\mathcal{S}|^\alpha) = o(n^{\alpha/2} \cdot n^{1/3-\alpha/2}) = o(n^{1/3})$ quantum queries to decide whether $f|_{\mathcal{S}}$ is 1-to-1 or has l collision pairs, which suffices to decide whether f is 1-to-1 or 2-to-1. However, this contradicts with the $\Omega(n^{1/3})$ quantum lower bound for the collision problem [53]. □

1) $\alpha = 0$*:* For 0-Rényi entropy estimation, we use Proposition IX.1 to give quantum lower bounds for both support coverage estimation and support size estimation (both defined in Section VIII).

Proposition IX.2. *The quantum query complexity of support coverage estimation is* $\Omega\big(\sqrt{n} + \frac{n^{1/3}}{\epsilon^{1/6}}\big)$*, for all* $\frac{1}{n} \leq \epsilon \leq \frac{1}{12}$*.*

Proof. On the one hand, [45, Theorem 55] proved that for $\epsilon = O(1)$, the quantum query complexity of support coverage estimation is $\Omega(\sqrt{n})$. Therefore, it suffices to prove that $\Omega\big(\frac{n^{1/3}}{\epsilon^{1/6}}\big)$ is a quantum lower bound when $\frac{1}{n} \leq \epsilon \leq \frac{1}{12}$.

Because $\frac{1}{n} \leq \epsilon \leq \frac{1}{12}$, we may denote $\epsilon = n^r$ where $r \in [-1, 0]$. Consider two distributions p_1 and p_2 encoded by $O_{p_1}, O_{p_2} : [n] \to X$ ($S = n$ in (I.7)), where the nonzero probabilities in p_1 are $1/n$ for n times, and the nonzero probabilities in p_2 are $2/n$ for $l = \lceil 6n\epsilon \rceil$ times and $1/n$ for $n - 2l$ times. In other words, O_{p_1} is injective, and O_{p_2} has l collision pairs but otherwise injective. On the one hand, by Proposition IX.1, it takes $\Omega(n^\alpha)$ quantum queries to distinguish between O_{p_1} and O_{p_2}, where

$$\begin{aligned} & l^\alpha = n^{\frac{2}{3}-\alpha} \\ \Rightarrow\ & \alpha = \frac{2}{3(2+r)} + O\Big(\frac{1}{\log n}\Big) \geq \frac{1}{3} - \frac{r}{6} \quad \forall\, r \in [-1, 0]. \end{aligned} \tag{IX.8}$$

As a result, $n^\alpha \geq n^{1/3 - r/6} = \frac{n^{1/3}}{\epsilon^{1/6}}$.

On the other hand,

$$\frac{S_n(p_1)}{n} = \frac{n \cdot \big(1 - (1-1/n)^n\big)}{n} \approx 1 - \frac{1}{e}; \tag{IX.9}$$

$$\begin{aligned} \frac{S_n(p_2)}{n} &= \frac{l \cdot \big(1 - (1-2/n)^n\big) + (n-2l) \cdot \big(1 - (1-1/n)^n\big)}{n} \\ &\approx 1 - \frac{1}{e} - \frac{l(1-1/e)^2}{n}. \end{aligned} \tag{IX.10}$$

As a result,

$$\Big|\frac{S_n(p_1)}{n} - \frac{S_n(p_2)}{n}\Big| \approx \frac{l(1-1/e)^2}{n} > 2\epsilon. \tag{IX.11}$$

Therefore, if a quantum algorithm can estimate support coverage with error ϵ, it can distinguish between p_1 and p_2 with success probability at least $2/3$. In conclusion, the quantum query complexity of support coverage estimation is $\Omega\big(\frac{n^{1/3}}{\epsilon^{1/6}}\big)$. □

Similar to the proof of Proposition IX.2, we can prove (with details omitted):

Proposition IX.3. *The quantum query complexity of support size estimation is* $\Omega\big(\frac{m^{1/3}}{\epsilon^{1/6}}\big)$*, for all* $\frac{1}{m} \leq \epsilon \leq \frac{1}{4}$*.*

2) $\frac{3}{7} \leq \alpha \leq 3$*:* Using Proposition IX.1, we show that the quantum query complexity of entropy estimation when $\frac{3}{7} \leq \alpha \leq 3$ is also $\Omega(n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$, as long as $\frac{1}{n} \leq \epsilon \leq \frac{1}{2}$.

Proof. We consider the case $\alpha = 1$, i.e., Shannon entropy estimation; the proof for other $\alpha \in [\frac{3}{7}, 3]$ is basically identical.

Consider two distributions p_1 and p_2 encoded by $O_{p_1}, O_{p_2} : [n] \to X$ ($S = n$ in (I.7)), where the nonzero probabilities in p_1 are $1/n$ for n times, and the nonzero probabilities in p_2 are $2/n$ for $l = \lceil n\epsilon / \log 2 \rceil$ times and $1/n$ for $n - 2l$ times.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

In other words, O_{p_1} is injective, and O_{p_2} has l collision pairs but otherwise injective. On the one hand, similar to the proof of Proposition IX.2, it takes it takes $\Omega(n^{1/3-\epsilon/6})$ quantum queries to distinguish between O_{p_1} and O_{p_2}.

On the other hand,

$$H(p_1) = n \cdot \frac{1}{n} \log n = \log n; \tag{IX.12}$$

$$H(p_2) = l \cdot \frac{2}{n} \log \frac{n}{2} + (n-2l) \cdot \frac{1}{n} \log n = \log n - \frac{2l}{n} \log 2. \tag{IX.13}$$

As a result,

$$|H(p_1) - H(p_2)| = \frac{2l}{n} \log 2 \geq 2\epsilon. \tag{IX.14}$$

Therefore, if a quantum algorithm can estimate support coverage with error ϵ, it can distinguish between p_1 and p_2 with success probability at least $2/3$. In conclusion, the quantum query complexity of support coverage estimation is $\Omega\big(\frac{n^{1/3}}{\epsilon^{1/6}}\big)$. □

Note that when $\alpha = 1$ (Shannon entropy), [45, Corollary 64] showed that the quantum query complexity is at least $\Omega(\sqrt{n})$ when $\epsilon = O(1)$. Together with the proof above, $\Omega(\sqrt{n} + n^{\frac{1}{3}}/\epsilon^{\frac{1}{6}})$ is a quantum lower bound for Shannon entropy estimation.

C. Polynomial method ($3 \leq \alpha \leq \infty$)

We use the polynomial method [43] to show quantum lower bounds for entropy estimation when $3 \leq \alpha \leq \infty$. Inspired by the symmetrization technique in [53], we obtain a bivariate polynomial whose degree is at most two times the corresponding quantum query complexity. Next, similar to [54], we apply Paturi's lemma [68] to give a lower bound on the degree of the polynomial. To be more specific, we prove:

Proposition IX.4. *The quantum query complexity of estimating min-entropy with error ϵ is $\Omega(\frac{\sqrt{n}}{\epsilon})$.*

Proposition IX.5. *When the constant α satisfies $1 < \alpha < \infty$, the quantum query complexity of estimating α-Rényi entropy with error ϵ is $\Omega(\frac{\alpha n^{\frac{1}{2}-\frac{1}{2\alpha}}}{\epsilon})$.*

Without loss of generality, we assume that the oracle O_p in (I.7) satisfies $n|S$, otherwise consider the oracle $O'_p\colon [Sn] \to [n]$ such that $O'_p(s+Sl) = O_p(s)$ for all $s \in [S]$ and $l \in [n]$; this gives an oracle for the same distribution.

We consider the special case where the probabilities $\{p_i\}_{i=1}^n$ takes at most two different values; to integrate the probabilities, we assume the existence of two integers c, d where $c \in \{1, \ldots, n-1\}$, such that $p_i = \frac{1}{n} - \frac{d}{S}$ for $n-c$ different i's in $\{1, \ldots, n\}$, and $p_i = \frac{1}{n} + \frac{(n-c)d}{cS}$ for the other c i's in $\{1, \ldots, n\}$.

Proof of Proposition IX.4. Following the symmetrization technique in [53], we obtain a bivariate polynomial $Q(c,d)$ where such that the degree of Q is at most two times the query complexity of min-entropy estimation, and:

- $c \in \{1, \ldots, n-1\}$ and $d \in \{-\lfloor \frac{Sc}{n(n-c)} \rfloor, \ldots, \frac{S}{n}\}$. This is because $p_i \geq 0$ for all $i \in [n]$.
- $0 \leq Q(c,d) \leq 1$ if $c|nd$. Only if $c|nd$, $S \cdot \big(\frac{1}{n} + \frac{(n-c)d}{cS}\big)$ is an integer and the distribution $\{p_i\}_{i=1}^n$ is valid under our model in (I.7).

Furthermore, we consider the property testing problem of determining whether $\max_i p_i = \frac{1}{n}$ or $\max_i p_i \geq \frac{1+\epsilon}{n}$, where the accept probability should be at most $1/3$ for the former case and at least $2/3$ for the latter case. As a result,

- $0 \leq Q(c,0) \leq 1/3$: In this case, $p_i = \frac{1}{n}$ for all $i \in [n]$.
- $2/3 \leq Q(c,d) \leq 1$ if $c|nd$, $\frac{(n-c)d}{Sc} \geq \frac{\epsilon}{n}$: In this case, $\exists\, i$ such that $p_i = \frac{1}{n} + \frac{(n-c)d}{cS} \geq \frac{1+\epsilon}{n}$.
- $2/3 \leq Q(c,d) \leq 1$ if $c|nd$, $d \leq -\frac{\epsilon S}{n}$: In this case, $\exists\, i$ such that $p_i = \frac{1}{n} - \frac{d}{S} \geq \frac{1+\epsilon}{n}$.

Therefore, we have

- $0 \leq Q(1,d) \leq 1$ for $d \in \{-\lfloor \frac{S}{n(n-1)} \rfloor, \ldots, \frac{S}{n}\}$;
- $0 \leq Q(1,0) \leq 1/3$;
- $2/3 \leq Q(1,d) \leq 1$ for $d \in \{-\lfloor \frac{S}{n(n-1)} \rfloor, \ldots, -\lceil \frac{\epsilon S}{n} \rceil\} \cup \{\lceil \frac{\epsilon S}{n(n-1)} \rceil, \ldots, \frac{S}{n}\}$.

Using Paturi's lower bound [68], we have

$$\deg_d Q(1,d) \geq \Omega\left(\frac{\sqrt{\lfloor \frac{S}{n(n-1)} \rfloor \cdot \frac{S}{n}}}{\lceil \epsilon S/n(n-1) \rceil}\right) = \Omega\Big(\frac{\sqrt{n}}{\epsilon}\Big). \tag{IX.15}$$

Therefore, $\deg Q(c,d) \geq \deg_d Q(1,d) = \Omega(\sqrt{n}/\epsilon)$. □

Proof of Proposition IX.5. The proof is similar to that of Proposition IX.4. Following the symmetrization technique, we still obtain a bivariate polynomial $Q(c,d)$ where such that the degree of Q is at most two times the query complexity of min-entropy estimation, and $c \in \{1, \ldots, n-1\}, d \in \{-\lfloor \frac{Sc}{n(n-c)} \rfloor, \ldots, \frac{S}{n}\}$, $0 \leq Q(c,d) \leq 1$ if $c|nd$. Furthermore, we consider the property testing problem of determining whether $\sum_{i\in[n]} p_i^\alpha \leq \frac{2}{n^{\alpha-1}}$ or $\sum_{i\in[n]} p_i^\alpha \geq \frac{2+2\epsilon}{n^{\alpha-1}}$, where the accept probability should be at most $1/3$ for the former case and at least $2/3$ for the latter case. We also assume $c = 1$. On the one hand, when $0 \leq d \leq \lfloor \frac{n^{1/\alpha}-1}{n-1} \cdot \frac{S}{n} \rfloor$, we have $\frac{1}{n} + \frac{(n-1)d}{S} \leq \frac{1}{n^{1-1/\alpha}}$, and

$$\begin{aligned}\sum_{i\in[n]} p_i^\alpha &\leq \Big(\frac{1}{n^{1-1/\alpha}}\Big)^\alpha + (n-1)\Big(\frac{1}{n-1}\Big(1 - \frac{1}{n^{1-1/\alpha}}\Big)\Big)^\alpha \\ &\leq \frac{2}{n^{\alpha-1}}.\end{aligned} \tag{IX.16}$$

On the other hand, because $(1+m)^\alpha \approx 1 + m\alpha$ when $m = o(1)$, we have

$$\begin{aligned}&(n-1)\Big(\frac{1}{n-1}\Big(1 - \frac{1}{n^{1-1/\alpha}} - \frac{3\epsilon}{\alpha n^{1-1/\alpha}}\Big)\Big)^\alpha \\ &\quad - \frac{2+2\epsilon}{n^{\alpha-1}} + \Big(\frac{1}{n^{1-1/\alpha}} + \frac{3\epsilon}{\alpha n^{1-1/\alpha}}\Big)^\alpha \\ &= \frac{1}{(n-1)^{\alpha-1}}\Big(1 - \frac{1}{n^{1-1/\alpha}} - \frac{3\epsilon}{\alpha n^{1-1/\alpha}}\Big)^\alpha \\ &\quad - \frac{2+2\epsilon}{n^{\alpha-1}} + \frac{1}{n^{\alpha-1}}\Big(1 + \frac{3\epsilon}{\alpha}\Big)^\alpha \qquad \text{(IX.17)} \\ &\approx \frac{1}{n^{\alpha-1}}\Big(\epsilon - \frac{\alpha+3\epsilon}{n^{1-1/\alpha}}\Big) \qquad \text{(IX.18)} \\ &\geq 0 \qquad \text{(IX.19)}\end{aligned}$$

for large enough n. As a result, when $d \geq \lceil \frac{(1+3\epsilon/\alpha)n^{1/\alpha}-1}{n-1} \cdot \frac{S}{n} \rceil$, we have $\sum_{i\in[n]} p_i^\alpha \geq \frac{2+2\epsilon}{n^{\alpha-1}}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2883306, IEEE Transactions on Information Theory

Therefore, we have

- $0 \le Q(1,d) \le 1$ for $d \in \{0,\dots,\frac{S}{n}\}$;
- $0 \le Q(1,d) \le 1/3$ for $d \in \{0,\dots,\lfloor \frac{n^{1/\alpha}-1}{n-1} \cdot \frac{S}{n} \rfloor\}$;
- $2/3 \le Q(1,d) \le 1$ for $d \in \{\lceil \frac{(1+3\epsilon/\alpha)n^{1/\alpha}-1}{n-1} \cdot \frac{S}{n} \rceil,\dots,\frac{S}{n}\}$.

Using Paturi's lower bound [68], we have

$$\deg_d Q(1,d) \ge \Omega\left(\frac{\sqrt{\lfloor \frac{n^{1/\alpha}-1}{n-1} \cdot \frac{S}{n} \rfloor \left(\frac{S}{n} - \lfloor \frac{n^{1/\alpha}-1}{n-1} \cdot \frac{S}{n} \rfloor\right)}}{\lceil \frac{(3\epsilon/\alpha)n^{1/\alpha}}{n-1} \cdot \frac{S}{n} \rceil}\right) = \Omega\Big(\frac{\alpha n^{\frac{1}{2}-\frac{1}{2\alpha}}}{\epsilon}\Big). \quad \text{(IX.20)}$$

Therefore, $\deg Q(c,d) \ge \deg_d Q(1,d) = \Omega(\alpha n^{\frac{1}{2}-\frac{1}{2\alpha}}/\epsilon)$. □

Technically, our proofs only focus on the degree in d for $c = 1$, but in general it is possible to prove a better lower bound when analyzing the degree of the polynomial in c and d together. We leave this as an open problem.

ACKNOWLEDGMENT

We thank Andrew M. Childs for discussions that inspired the proof of Theorem VI.1, and general suggestions on our manuscript; we thank Yanjun Han for introducing us classical references related to Shannon and Rényi entropy estimation, in particular his papers [13, 14, 58]. We also thank anonymous reviewers for helpful comments on earlier versions of this paper. TL acknowledges support from NSF CCF-1526380. XW acknowledges support from NSF CCF-1755800 and CCF-1816695.

APPENDIX A
THEOREM II.2: MULTIPLICATIVE QUANTUM CHEBYSHEV INEQUALITY

The main technique that we use is Lemma 4 in [2], which approximates a random variable with an additive error as long as its second-moment is bounded:

Lemma A.1 (Lemma 4 in [2])**.** *Assume $\mathcal{A}$ is a quantum algorithm that outputs a random variable X. Then for ϵ where $0 < \epsilon < 1/2$ (multiplicative error), by using $O((1/\epsilon)\log^{3/2}(1/\epsilon)\log\log(1/\epsilon))$ executions of $\mathcal{A}$ and $\mathcal{A}^{-1}$, Algorithm 2 in [2] outputs an estimate $\tilde{\mathbb{E}}[X]$ of $\mathbb{E}[X]$ such that*[8]

$$\Pr\left[\left|\tilde{\mathbb{E}}[X] - \mathbb{E}[X]\right| \ge \epsilon(\sqrt{\mathbb{E}[X^2]}+1)^2\right] \le 1/50. \quad \text{(A.1)}$$

Based on Lemma A.1 and inspired by Algorithm 3 and Theorem 5 in [2], we propose Algorithm 10.

Proof of Theorem II.2. Because $\mathrm{Var}[X] \le \sigma^2\mathbb{E}[X]^2 \le \sigma^2 b^2$, by Chebyshev's inequality we have

$$\Pr\left[\left|\widetilde{m} - \mathbb{E}[X/\sigma b]\right| \ge 4\right] \le 1/16. \quad \text{(A.2)}$$

Therefore, with probability at least $15/16$ we have $|\widetilde{m} - \mathbb{E}[X/\sigma b]| \le 4$. Denote $X_B = \frac{X}{\sigma b} - \widetilde{m}$, which is the random variable output by $\mathcal{B}$; $X_{B,+} := \max\{X_B, 0\}$ is then the output of $\mathcal{B}_+$ and $X_{B,-} := \min\{X_B, 0\}$ is the output of $\mathcal{B}_-$. Assuming $|\widetilde{m} - \mathbb{E}[X/\sigma b]| \le 4$, we have

$$\mathbb{E}[X_B^2] = \mathbb{E}\Big[\Big((\frac{X}{\sigma b} - \mathbb{E}[\frac{X}{\sigma b}]) + \big(\mathbb{E}[\frac{X}{\sigma b}] - \widetilde{m}\big)\Big)^2\Big] \quad \text{(A.3)}$$

$$\le 2\mathbb{E}\Big[\Big(\frac{X}{\sigma b} - \mathbb{E}[\frac{X}{\sigma b}]\Big)^2\Big] + 2\mathbb{E}\Big[\Big(\mathbb{E}[\frac{X}{\sigma b}] - \widetilde{m}\Big)^2\Big] \quad \text{(A.4)}$$

$$\le 2(1^2 + 4^2) = 34. \quad \text{(A.5)}$$

Therefore, $\mathbb{E}\big[(X_B/6)^2\big] \le 34/36 < 1$, hence $\mathbb{E}\big[(X_{B,+}/6)^2\big] < 1$ and $\mathbb{E}\big[(-X_{B,-}/6)^2\big] < 1$. By Lemma A.1, we have

$$\big|\widetilde{\mu}_- - \mathbb{E}[-X_{B,-}/6]\big| \le \frac{\epsilon a}{12\sigma b} \quad \text{(A.6)}$$

$$\big|\widetilde{\mu}_+ - \mathbb{E}[X_{B,+}/6]\big| \le \frac{\epsilon a}{12\sigma b}, \quad \text{(A.7)}$$

both with failure probability at most $1/50$. Because

$$\mathbb{E}[X] = \sigma b\big(\widetilde{m} + \mathbb{E}[X_B]\big) = \sigma b\big(\widetilde{m} + \mathbb{E}[X_{B,+}] - \mathbb{E}[-X_{B,-}]\big), \quad \text{(A.8)}$$

with probability at least $15/16 \cdot (1-1/50)^2 > 9/10$, we have

$$\big|\tilde{\mathbb{E}}[X] - \mathbb{E}[X]\big| \le \sigma b \cdot \big(6\big|\widetilde{\mu}_- - \mathbb{E}[-X_{B,-}/6]\big| + 6\big|\widetilde{\mu}_+ - \mathbb{E}[X_{B,+}/6]\big|\big) \quad \text{(A.9)}$$

$$\le \sigma b \cdot 2 \cdot 6 \cdot \frac{\epsilon a}{12\sigma b} = \epsilon a \le \epsilon\mathbb{E}(X). \quad \text{(A.10)}$$

□

Algorithm 10: Estimate $\mathbb{E}[X]$ within multiplicative error ϵ.

1. Run the algorithm that gives a, b such that $\mathbb{E}[X] \in [a,b]$;
2. Set $\mathcal{A}' = \mathcal{A}/\sigma b$;
3. Run $\mathcal{A}'$ once and denote $\widetilde{m}$ to be the output. Set $\mathcal{B} = \mathcal{A}' - \widetilde{m}$;
4. Let $\mathcal{B}_-$ be the algorithm that calls $\mathcal{B}$ once; if $\mathcal{B}$ outputs $x \ge 0$ then $\mathcal{B}_-$ outputs 0, and if $\mathcal{B}$ outputs $x < 0$ then $\mathcal{B}_-$ outputs x. Similarly, let $\mathcal{B}_+$ be the algorithm such that if $\mathcal{B}$ outputs $x < 0$ then $\mathcal{B}_+$ outputs 0, and if $\mathcal{B}$ outputs $x \ge 0$ then $\mathcal{B}_+$ outputs x;
5. Apply Lemma A.1 to $-\mathcal{B}_-/6$ and $\mathcal{B}_+/6$ with error $\frac{\epsilon a}{48\sigma b}$ and failure probability $1/50$, and obtain estimates $\widetilde{\mu}_-$ and $\widetilde{\mu}_+$, respectively;
6. Output $\tilde{\mathbb{E}}[X] = \sigma b(\widetilde{m} - 6\widetilde{\mu}_- + 6\widetilde{\mu}_+)$;

[8] The original error probability in (A.1) is $1/5$, but it can be improved to $1/50$ by rescaling the parameters in Lemma 4 in [2] up to a constant.

REFERENCES

[1] S. Bravyi, A. W. Harrow, and A. Hassidim, "Quantum algorithms for testing properties of distributions," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3971–3981, 2011.

[2] A. Montanaro, "Quantum speedup of Monte Carlo methods," in *Proc. R. Soc. A*, vol. 471, no. 2181. The Royal Society, 2015, p. 20150301.

[3] D. Ron, "Algorithmic and analysis techniques in property testing," *Foundations and Trends in Theoretical Computer Science*, vol. 5, no. 2, pp. 73–205, 2010.

[4] P. Valiant, "Testing symmetric properties of distributions," *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1927–1968, 2011.

[5] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[6] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 547–561.

[7] T. Batu, S. Dasgupta, R. Kumar, and R. Rubinfeld, "The complexity of approximating the entropy," *SIAM Journal on Computing*, vol. 35, no. 1, pp. 132–150, 2005.

[8] L. Paninski, "Estimation of entropy and mutual information," *Neural Computation*, vol. 15, no. 6, pp. 1191–1253, 2003.

[9] ——, "Estimating entropy on m bins given fewer than m samples," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2200–2203, 2004.

[10] G. Valiant and P. Valiant, "Estimating the unseen: an $n/log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs," in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*. ACM, 2011, pp. 685–694.

[11] ——, "The power of linear estimators," in *52nd Annual Symposium on Foundations of Computer Science*. IEEE, 2011, pp. 403–412.

[12] Y. Wu and P. Yang, "Minimax rates of entropy estimation on large alphabets via best polynomial approximation," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3702–3720, 2016.

[13] J. Jiao, K. Venkat, Y. Han, and T. Weissman, "Minimax estimation of functionals of discrete distributions," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2835–2885, 2015.

[14] ——, "Maximum likelihood estimation of functionals of discrete distributions," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6774–6798, 2017.

[15] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[16] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, 1995.

[17] P. C. van Oorschot and M. J. Wiener, "Parallel collision search with cryptanalytic applications," *Journal of Cryptology*, vol. 12, no. 1, pp. 1–28, 1999.

[18] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[19] R. Impagliazzo and D. Zuckerman, "How to recycle random bits," in *30th Annual Symposium on Foundations of Computer Science*. IEEE, 1989, pp. 248–253.

[20] J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi, "Estimating Rényi entropy of discrete distributions," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 38–56, 2017.

[21] S. P. Vadhan, "Pseudorandomness," *Foundations and Trends® in Theoretical Computer Science*, vol. 7, no. 1–3, pp. 1–336, 2012.

[22] R. V. L. Hartley, "Transmission of information," *Bell Labs Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928.

[23] B. Efron and R. Thisted, "Estimating the number of unseen species: How many words did Shakespeare know?" *Biometrika*, vol. 63, no. 3, pp. 435–447, 1976.

[24] R. Thisted and B. Efron, "Did Shakespeare write a newly-discovered poem?" *Biometrika*, vol. 74, no. 3, pp. 445–455, 1987.

[25] P. J. Haas, J. F. Naughton, S. Seshadri, and L. Stokes, "Sampling-based estimation of the number of distinct values of an attribute," in *Proceedings of 21th International Conference on Very Large Data Bases*, vol. 95, 1995, pp. 311–322.

[26] D. Florencio and C. Herley, "A large-scale study of web password habits," in *Proceedings of the 16th International Conference on World Wide Web*. ACM, 2007, pp. 657–666.

[27] I. Kroes, P. W. Lepp, and D. A. Relman, "Bacterial diversity within the human subgingival crevice," *Proceedings of the National Academy of Sciences*, vol. 96, no. 25, pp. 14 547–14 552, 1999.

[28] B. J. Paster, S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe, and F. E. Dewhirst, "Bacterial diversity in human subgingival plaque," *Journal of Bacteriology*, vol. 183, no. 12, pp. 3770–3783, 2001.

[29] J. B. Hughes, J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan, "Counting the uncountable: statistical approaches to estimating microbial diversity," *Applied and Environmental Microbiology*, vol. 67, no. 10, pp. 4399–4406, 2001.

[30] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[31] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[32] O. Catoni, *Statistical Learning Theory and Stochastic Optimization: Ecole D'Eté de Probabilités de Saint-Flour XXXI-2001*. Springer, 2004.

[33] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv:1312.6114*, 2013.

[34] A. Montanaro and R. de Wolf, "A survey of quantum property testing," *Theory of Computing Graduate Surveys*, no. 7, pp. 1–81, 2016.

[35] S. Chakraborty, E. Fischer, A. Matsliah, and R. de Wolf, "New results on quantum property testing," in *Thirtieth International Conference on Foundations of Software Technology and Theoretical Computer Science*, 2010, p. 145.

[36] R. O'Donnell and J. Wright, "Quantum spectrum testing," in *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*. ACM, 2015, pp. 529–538.

[37] ——, "Efficient quantum tomography," in *Proceedings of the 48th Annual ACM Symposium on Theory of*

Computing. ACM, 2016, pp. 899–912.

[38] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," in *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*. ACM, 2016, pp. 913–925.

[39] R. O'Donnell and J. Wright, "Efficient quantum tomography II," in *Proceedings of the Forty-ninth Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 2017, pp. 962–974.

[40] J. Acharya, I. Issa, N. V. Shende, and A. B. Wagner, "Measuring quantum entropy," *arXiv:1711.00814*, 2017.

[41] A. Ambainis, "Quantum walk algorithm for element distinctness," *SIAM Journal on Computing*, vol. 37, no. 1, pp. 210–239, 2007.

[42] A. Belovs, "Learning-graph-based quantum algorithm for k-distinctness," in *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science*. IEEE, 2012, pp. 207–216.

[43] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf, "Quantum lower bounds by polynomials," *Journal of the ACM (JACM)*, vol. 48, no. 4, pp. 778–797, 2001.

[44] S. Aaronson and Y. Shi, "Quantum lower bounds for the collision and the element distinctness problems," *Journal of the ACM (JACM)*, vol. 51, no. 4, pp. 595–605, 2004.

[45] M. Bun, R. Kothari, and J. Thaler, "The polynomial method strikes back: Tight quantum query bounds via dual polynomials," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 2018, pp. 297–310.

[46] Y. Wu and P. Yang, "Chebyshev polynomials, moment matching, and optimal estimation of the unseen," *arXiv:1504.01227*, 2015.

[47] A. Orlitsky, A. T. Suresh, and Y. Wu, "Optimal prediction of the number of unseen species," *Proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. 13 283–13 288, 2016.

[48] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, "Quantum amplitude amplification and estimation," *Contemporary Mathematics*, vol. 305, pp. 53–74, 2002.

[49] D. Štefankovič, S. Vempala, and E. Vigoda, "Adaptive simulated annealing: A near-optimal connection between sampling and counting," *Journal of the ACM (JACM)*, vol. 56, no. 3, p. 18, 2009.

[50] A. Montanaro, "The quantum complexity of approximating the frequency moments," *Quantum Information & Computation*, vol. 16, no. 13&14, pp. 1169–1190, 2016.

[51] L. Le Cam, *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

[52] S. Aaronson and A. Ambainis, "The need for structure in quantum speedups," *Theory of Computing*, vol. 10, no. 6, pp. 133–166, 2014.

[53] S. Kutin, "Quantum lower bound for the collision problem with small range," *Theory of Computing*, vol. 1, no. 1, pp. 29–36, 2005.

[54] A. Nayak and F. Wu, "The quantum query complexity of approximating the median and related statistics," in *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*. ACM, 1999, pp. 384–393.

[55] P. Høyer, T. Lee, and R. Spalek, "Negative weights make adversaries stronger," in *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*. ACM, 2007, pp. 526–535.

[56] A. Belovs and A. Rosmanis, "Adversary lower bounds for the collision and the set equality problems," *arXiv:1310.5185*, 2013.

[57] P. Dagum, R. Karp, M. Luby, and S. Ross, "An optimal algorithm for Monte Carlo estimation," *SIAM Journal on Computing*, vol. 29, no. 5, pp. 1484–1496, 2000.

[58] Y. Han, J. Jiao, and T. Weissman, "Minimax rate-optimal estimation of divergences between discrete distributions," *arXiv:1605.09124*, 2016.

[59] Y. Bu, S. Zou, Y. Liang, and V. V. Veeravalli, "Estimation of KL divergence: optimal minimax rate," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2648–2674, 2018.

[60] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2005.

[61] P. W. Glynn, "Upper bounds on Poisson tail probabilities," *Operations Research Letters*, vol. 6, no. 1, pp. 9–14, 1987.

[62] G. H. Hardy, P. V. Seshu Aiyar, and B. M. Wilson, *Collected papers of Srinivasa Ramanujan*. AMS, 1927.

[63] R. A. Fisher, A. S. Corbet, and C. B. Williams, "The relation between the number of species and the number of individuals in a random sample of an animal population," *The Journal of Animal Ecology*, pp. 42–58, 1943.

[64] I. J. Good and G. H. Toulmin, "The number of new species, and the increase in population coverage, when a sample is increased," *Biometrika*, vol. 43, no. 1-2, pp. 45–63, 1956.

[65] J. Zou, G. Valiant, P. Valiant, K. Karczewski, S.-O. Chan, K. Samocha, M. Lek, S. Sunyaev, M. Daly, and D. G. MacArthur, "Quantifying unobserved protein-coding variants in human populations provides a roadmap for large-scale sequencing projects," *Nature communications*, vol. 7, 2016.

[66] J. Acharya, H. Das, A. Orlitsky, and A. T. Suresh, "A unified maximum likelihood approach for estimating symmetric properties of discrete distributions," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 11–21.

[67] S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith, "Strong lower bounds for approximating distribution support size and the distinct elements problem," *SIAM Journal on Computing*, vol. 39, no. 3, pp. 813–842, 2009.

[68] R. Paturi, "On the degree of polynomials that approximate symmetric Boolean functions (preliminary version)," in *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*. ACM, 1992, pp. 468–474.

Tongyang Li is a Ph.D. candidate at the Department of Computer Science, University of Maryland. He received B.E. from Institute for Interdisciplinary Information Sciences, Tsinghua University and B.S. from Department of Mathematical Sciences, Tsinghua University, both in 2015; he also received a Master degree from Department of Computer Science, University of Maryland in 2018. He is a recipient of the IBM Ph.D. Fellowship and the NSF QISE-NET Triplet Award. His research focuses on quantum algorithms, including topics such as quantum machine learning, quantum query complexity, and quantum simulation.

Xiaodi Wu received the B.S. degree in mathematics and physics in 2008 from Tsinghua University, the Ph.D. degree in computer science from University of Michigan, Ann Arbor in 2013. He was a postdoctoral associate at MIT (2013 – 2015) and a Simons research fellow at the Simons Institute, UC Berkeley (2014). He then worked as an assistant professor in the computer and information science department at the University of Oregon (2015 – 2017). Since the summer of 2017, he has been an assistant professor in the department of computer science and the institute for advanced computer studies at the University of Maryland, College Park.