Quantum query complexity of entropy estimation

Xiaodi Wu

QuICS, University of Maryland

MSR Redmond, July 19th, 2017

UNIVERSITY OF MARYLAND

Joint Center for Quantum Information and Computer Science

Outline

Outline

Motivation

Why do I come across this problem?

- Quantum property testing [e.g., survey [Montanaro, de Wolf]] (Quantum testers of classical properties !)

Motivation

Why do I come across this problem?

- Quantum property testing [e.g., survey [Montanaro, de Wolf]] (Quantum testers of classical properties !)
- Testing of *distributional properties*. A well-motivated branch in the classical property testing literature !! also connected to learning problems.

Motivation

Why do I come across this problem?

- Quantum property testing [e.g., survey [Montanaro, de Wolf]] (Quantum testers of classical properties !)
- Testing of *distributional properties*. A well-motivated branch in the classical property testing literature !! also connected to learning problems.
- Objects are *distributions*, rather than *boolean functions*. Might bring new insights or call for new techniques!

Entropies

Given any distribution p over a discrete set X, the *Shannon entropy* of this distribution p is defined by

$$H(p) := \sum_{x \in X : p(x) > 0} -p_x \log p_x.$$

Entropies

Given any distribution p over a discrete set X, the *Shannon entropy* of this distribution p is defined by

$$H(p) := \sum_{x \in X : p(x) > 0} -p_x \log p_x .$$

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

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

Problem Statement

For convenience, assume $X = [n]$.

Problem Statement

For convenience, assume $X = [n]$.

A natural question: Given access to samples obtained by taking independent draws from p, determine the necessary number of independent draws to **estimate $H(p)$ or $H_\alpha(p)$ within error** ϵ, with high probability.

Motivations: this is a theoretically appealing topic with intimate connections to statistics, information theory, learning theory, and algorithm design.

Our Question

Main Question: is there any **quantum speed-up** of estimation of Shannon and Rényi entropies ?

Our question aligns with the emerging topic called "quantum property testing" and focuses on investigating the quantum advantage in testing classical statistical properties.

Our Question

Main Question: is there any **quantum speed-up** of estimation of Shannon and Rényi entropies ?

Our question aligns with the emerging topic called "quantum property testing" and focuses on investigating the quantum advantage in testing classical statistical properties.

The first research paper on this topic is by Bravyi, Harrow, and Hassidim [BHH 11], where they have discovered quantum speed-ups of testing *uniformity, orthogonality, and statistical difference* on unknown distributions, followed by Chakraborty et al. [CFMdW 10].

Outline

Classical results

Classically, this fundamental question has been intensively studied in both the communities of theoretical computer science and information theory.

Classical results

Classically, this fundamental question has been intensively studied in both the communities of theoretical computer science and information theory.

If ϵ is a constant, then the classical query complexity

Classical results

Classically, this fundamental question has been intensively studied in both the communities of theoretical computer science and information theory.

If ϵ is a constant, then the classical query complexity

- for Shannon entropy [VV 11], [JVHW 15]: $\Theta(\frac{n}{\log n})$.

Classical results

Classically, this fundamental question has been intensively studied in both the communities of theoretical computer science and information theory.

If ϵ is a constant, then the classical query complexity

- for Shannon entropy [VV 11], [JVHW 15]: $\Theta(\frac{n}{\log n})$.
- for Rényi entropy [AOST 17]:

$$\begin{cases} O(\frac{n^{\frac{1}{\alpha}}}{\log n}) \text{ and } \Omega(n^{\frac{1}{\alpha}-o(1)}), & \text{when } 0 < \alpha < 1. \\ O(\frac{n}{\log n}) \text{ and } \Omega(n^{1-o(1)}), & \text{when } \alpha > 1, \alpha \notin \mathbb{N}. \\ \Theta(n^{1-\frac{1}{\alpha}}), & \text{when } \alpha > 1, \alpha \in \mathbb{N}. \end{cases}$$

Quantum results

Our results:

α	classical bounds	quantum bounds (this talk)
$0<\alpha<1$	$O(\frac{n^{\frac{1}{\alpha}}}{\log n})$, $\Omega(n^{\frac{1}{\alpha}-o(1)})$ [AOST 17]	$\tilde{O}(n^{\frac{1}{\alpha}-\frac{1}{2}})$, $\Omega(\max\{n^{\frac{1}{7\alpha}-o(1)}, n^{\frac{1}{3}}\})$
$\alpha=1$	$\Theta(\frac{n}{\log n})$ [VV 11, JVHW 15]	$\tilde{O}(\sqrt{n})$, $\Omega(n^{\frac{1}{3}})$
$\alpha>1, \alpha\notin\mathbb{N}$	$O(\frac{n}{\log n})$, $\Omega(n^{1-o(1)})$ [AOST 17]	$\tilde{O}(n^{1-\frac{1}{2\alpha}})$, $\Omega(\max\{n^{\frac{1}{3}}, \Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}})\})$
$\alpha=2$	$\Theta(\sqrt{n})$[AOST 17]	$\tilde{\Theta}(n^{\frac{1}{3}})$
$\alpha>2, \alpha\in\mathbb{N}$	$\Theta(n^{1-1/\alpha})$ [AOST 17]	$\tilde{O}(n^{\nu(1-1/\alpha)})$, $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}})$, $\nu<3/4$
$\alpha=\infty$	$\Theta(\frac{n}{\log n})$ [VV 11]	$\tilde{O}(Q(\lceil\log n\rceil\text{-distinctness}))$, $\Omega(\sqrt{n})$

Table 1: Summary of classical and quantum query complexity of $H_\alpha(p)$ for $\alpha>0$, assuming $\epsilon=\Theta(1)$.

Quantum results

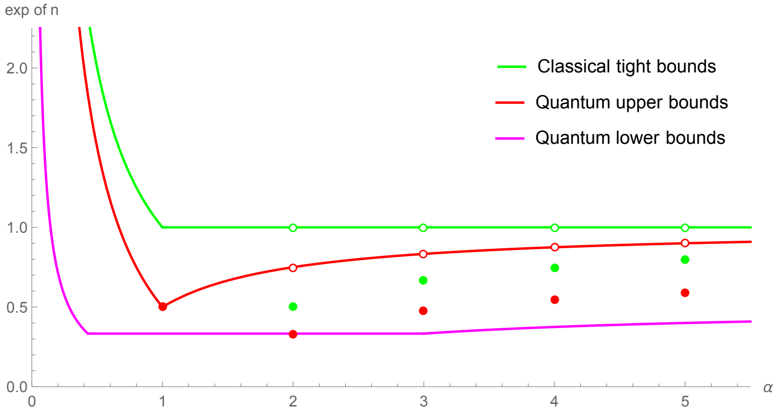


Figure 1: Visualization of classical and quantum query complexity of $H_\alpha(p)$. The x-axis represents α and the y-axis represents the exponent of n. Red curves and points represent quantum upper bounds. Green curves and points represent classical tight bounds. The Magenta curve represents quantum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

Quantum results: in my April talk at MSR

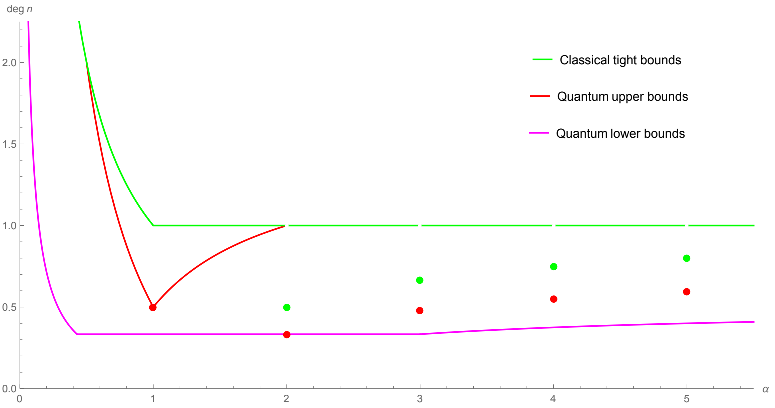


Figure 2: Visualization of classical and quantum query complexity of $H_\alpha(p)$. The x-axis represents α and the y-axis represents the exponent of n. Red curves and points represent quantum upper bounds. Green curves and points represent classical tight bounds. The Magenta curve represents quantum lower bounds.

Outline

Sample vs **Query** model

Sample vs **Query** model

Ref. [BHH 11] models any discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$ by an oracle $O_p\colon [S] \to [n]$. Any probability p_i $(i \in [n])$ is thus proportional to the size of pre-image of i under O_p:

$$p_i = \frac{|\{s \in [S] : O_p(s) = i\}|}{S} \quad \forall\, i \in [n].$$

If one samples s uniformly from $[S]$ and outputs $O_p(s)$, then one obtains a sample drawn from distribution p.

Sample vs **Query** model

Ref. [BHH 11] models any discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$ by an oracle $O_p\colon [S] \to [n]$. Any probability p_i ($i \in [n]$) is thus proportional to the size of pre-image of i under O_p:

$$p_i = \frac{|\{s \in [S] : O_p(s) = i\}|}{S} \quad \forall\, i \in [n].$$

If one samples s uniformly from $[S]$ and outputs $O_p(s)$, then one obtains a sample drawn from distribution p.

It is shown in [BHH 11] that the *query complexity* of the oracle model above and the *sample complexity* of independent samples are in fact equivalent classically.

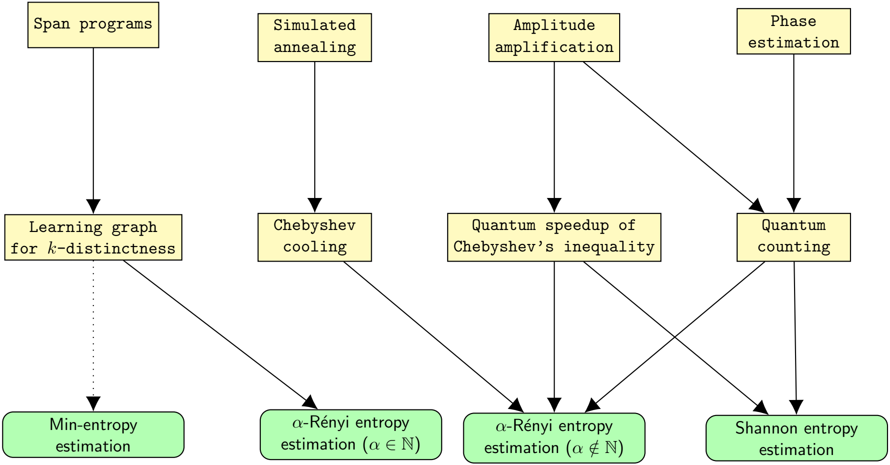
Quantum Query Model

Quantumly, O_p is transformed into a unitary operator $\hat{O}_p$ acting on $\mathbb{C}^S \otimes \mathbb{C}^{n+1}$ such that

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

This is more powerful than O_p because we may take a superposition of states as an input.

Roadmap of quantum entropy estimation: for all $\alpha > 0$



High-level Framework

A general distribution property estimation problem:

Given a discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$ and a function $f\colon (0,1] \to \mathbb{R}$, estimate $F(p) := \sum_{i\in[n]} p_i f(p_i)$ with small additive or multiplicative error with high success probability.

High-level Framework

A general distribution property estimation problem:

Given a discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$ and a function $f\colon (0,1] \to \mathbb{R}$, estimate $F(p) := \sum_{i\in[n]} p_i f(p_i)$ with small additive or multiplicative error with high success probability.

If $f(x) = -\log x$, $F(p)$ is the Shannon entropy $H(p)$ (*additive error*); if $f(x) = x^{\alpha-1}$ for some $\alpha > 0, \alpha \neq 1$, $H_\alpha(p) \propto \log F(p)$ (*multiplicative error*).

High-level Framework

A general distribution property estimation problem:

Given a discrete distribution $p = (p_i)_{i=1}^n$ on $[n]$ and a function $f\colon (0,1] \to \mathbb{R}$, estimate $F(p) := \sum_{i\in[n]} p_i f(p_i)$ with small additive or multiplicative error with high success probability.

If $f(x) = -\log x$, $F(p)$ is the Shannon entropy $H(p)$ (*additive error*); if $f(x) = x^{\alpha-1}$ for some $\alpha > 0, \alpha \neq 1$, $H_\alpha(p) \propto \log F(p)$ (*multiplicative error*).

Inspired by BHH, we formulate a framework for approximating $F(p)$.

High-level Framework

Algorithm: Estimate $F(p) = \sum_i p_i f(p_i)$.

Set $l, M \in \mathbb{N}$;

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

- Draw a sample $i \in [n]$ according to p;
- Use amplitude estimation with M queries to obtain an estimation $\tilde{p}_i$ of p_i;
- Output $X = f(\tilde{p}_i)$;

Use $\mathcal{A}$ for l times in *"quantum speedup of Chebyshev's inequality"* and outputs an estimation $\tilde{F}(\tilde{p})$ of $F(p)$;

High-level Framework

Algorithm: Estimate $F(p) = \sum_i p_i f(p_i)$.

Set $l, M \in \mathbb{N}$;

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

- Draw a sample $i \in [n]$ according to p;
- Use amplitude estimation with M queries to obtain an estimation $\tilde{p}_i$ of p_i;
- Output $X = f(\tilde{p}_i)$;

Use $\mathcal{A}$ for l times in *"quantum speedup of Chebyshev's inequality"* and outputs an estimation $\tilde{F}(\tilde{p})$ of $F(p)$;

Intuitively, performance depends on $\mathbb{E}(\tilde{F}(\tilde{p}))$ and $\sigma(\tilde{F}(\tilde{p}))$.

High-level Framework

Algorithm: Estimate $F(p) = \sum_i p_i f(p_i)$.

Set $l, M \in \mathbb{N}$;

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

- Draw a sample $i \in [n]$ according to p;
- Use amplitude estimation with M queries to obtain an estimation $\tilde{p}_i$ of p_i;
- Output $X = f(\tilde{p}_i)$;

Use $\mathcal{A}$ for l times in *"quantum speedup of Chebyshev's inequality"* and outputs an estimation $\tilde{F}(\tilde{p})$ of $F(p)$;

Intuitively, performance depends on $\mathbb{E}(\tilde{F}(\tilde{p}))$ and $\sigma(\tilde{F}(\tilde{p}))$.

Want: $\mathbb{E}(\tilde{F}(\tilde{p}))$ close to $F(p)$ and $\sigma(\tilde{F}(\tilde{p}))$ small.

High-level Framework

Algorithm: Estimate $F(p) = \sum_i p_i f(p_i)$.

Set $l, M \in \mathbb{N}$;

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

- Draw a sample $i \in [n]$ according to p;
- Use amplitude estimation with M queries to obtain an estimation $\tilde{p}_i$ of p_i;
- Output $X = f(\tilde{p}_i)$;

Use $\mathcal{A}$ for l times in *"quantum speedup of Chebyshev's inequality"* and outputs an estimation $\tilde{F}(\tilde{p})$ of $F(p)$;

Intuitively, performance depends on $\mathbb{E}(\tilde{F}(\tilde{p}))$ and $\sigma(\tilde{F}(\tilde{p}))$.

Want: $\mathbb{E}(\tilde{F}(\tilde{p}))$ close to $F(p)$ and $\sigma(\tilde{F}(\tilde{p}))$ small.

Pros: intuitively correct; (tedious) calculation only !?

High-level Framework

Algorithm: Estimate $F(p) = \sum_i p_i f(p_i)$.

Set $l, M \in \mathbb{N}$;
Regard the following subroutine as $\mathcal{A}$:
  Draw a sample $i \in [n]$ according to p;
  Use amplitude estimation with M queries to obtain an estimation $\tilde{p}_i$ of p_i;
  Output $X = f(\tilde{p}_i)$;
Use $\mathcal{A}$ for l times in *"quantum speedup of Chebyshev's inequality"* and outputs an estimation $\tilde{F}(\tilde{p})$ of $F(p)$;

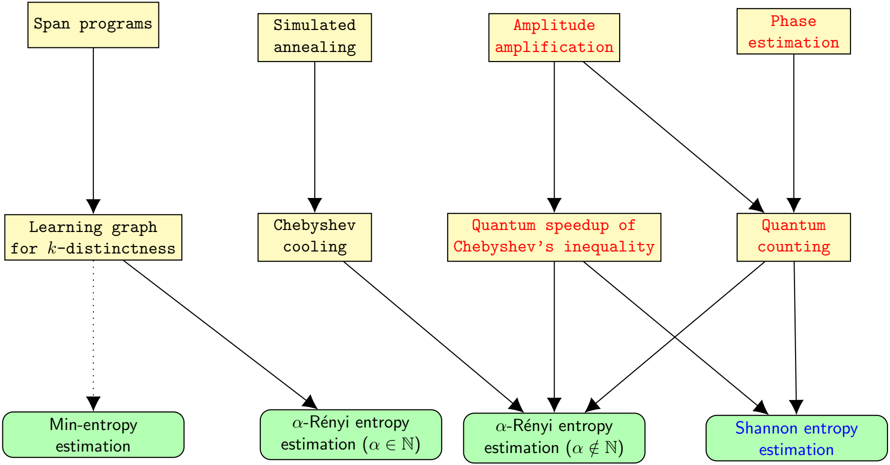
Intuitively, performance depends on $\mathbb{E}(\tilde{F}(\tilde{p}))$ and $\sigma(\tilde{F}(\tilde{p}))$.

Want: $\mathbb{E}(\tilde{F}(\tilde{p}))$ close to $F(p)$ and $\sigma(\tilde{F}(\tilde{p}))$ small.

Pros: intuitively correct; (tedious) calculation only !?

Cons: very limited quantum speedup with only the **basic** framework.

Roadmap of quantum entropy estimation: for all $\alpha > 0$



Quantum algorithm for Shannon entropy estimation

Two **new** ingredients (c.f. BHH) when $f(p) = -\log p$

Quantum algorithm for Shannon entropy estimation

Two **new** ingredients (c.f. BHH) when $f(p) = -\log p$

- Montanaro's *quantum speedup of Monte Carlo* methods. Let $\mathcal{A}$ be a quantum algorithm outputting X such that $\mathrm{Var}(X) \leq \sigma^2$. For ϵ s.t. $0 < \epsilon < 4\sigma$, by using $\tilde{O}(\sigma/\epsilon)$ times of $\mathcal{A}$ and $\mathcal{A}^{-1}$, one can output estimate $\tilde{\mathbb{E}}(X)$ of $\mathbb{E}(X)$ s.t.

$$\Pr\left[|\tilde{\mathbb{E}}(X) - \mathbb{E}(X)| \geq \epsilon\right] \leq 1/4.$$

  Classically, one needs to use $\Theta(\sigma^2/\epsilon^2)$ times of $\mathcal{A}$.

Quantum algorithm for Shannon entropy estimation

Two **new** ingredients (c.f. BHH) when $f(p) = -\log p$

- Montanaro's *quantum speedup of Monte Carlo* methods. Let $\mathcal{A}$ be a quantum algorithm outputting X such that $\mathrm{Var}(X) \leq \sigma^2$. For ϵ s.t. $0 < \epsilon < 4\sigma$, by using $\tilde{O}(\sigma/\epsilon)$ times of $\mathcal{A}$ and $\mathcal{A}^{-1}$, one can output estimate $\tilde{\mathbb{E}}(X)$ of $\mathbb{E}(X)$ s.t.

$$\Pr\left[|\tilde{\mathbb{E}}(X) - \mathbb{E}(X)| \geq \epsilon\right] \leq 1/4.$$

  Classically, one needs to use $\Theta(\sigma^2/\epsilon^2)$ times of $\mathcal{A}$.
- a *fine-tuned analysis* to bound the value of $\log(1/p)$ when $p \to 0$. Analyze the full distribution of amplitude amplification.

Quantum algorithm for Shannon entropy estimation

Two **new** ingredients (c.f. BHH) when $f(p) = -\log p$

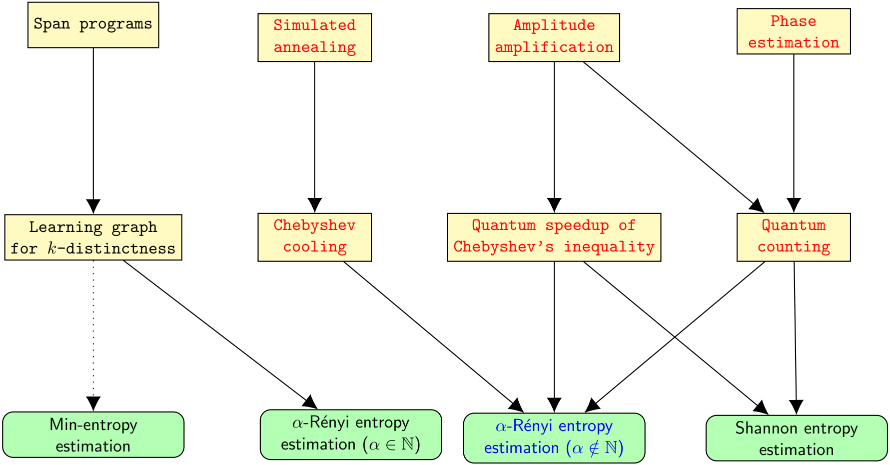
- Montanaro's *quantum speedup of Monte Carlo* methods. Let $\mathcal{A}$ be a quantum algorithm outputting X such that $\mathrm{Var}(X) \le \sigma^2$. For ϵ s.t. $0 < \epsilon < 4\sigma$, by using $\tilde{O}(\sigma/\epsilon)$ times of $\mathcal{A}$ and $\mathcal{A}^{-1}$, one can output estimate $\tilde{\mathbb{E}}(X)$ of $\mathbb{E}(X)$ s.t.

  $$\Pr\left[|\tilde{\mathbb{E}}(X) - \mathbb{E}(X)| \ge \epsilon\right] \le 1/4.$$

  Classically, one needs to use $\Theta(\sigma^2/\epsilon^2)$ times of $\mathcal{A}$.
- a *fine-tuned analysis* to bound the value of $\log(1/p)$ when $p \to 0$. Analyze the full distribution of amplitude amplification.

Complexity: Classical $\Theta(n/\log(n))$ vs Quantum $\tilde{O}(\sqrt{n})$.

Roadmap of quantum entropy estimation: for all $\alpha > 0$



Quantum algorithm for α-Rényi entropy: $\alpha \notin N$

Issue w/ the **basic** framework for Rényi entropy

multiplicative errors (Montanaro's algorithm yields no speedup in the worst case) $\rightarrow$ quantum advantage when $1/2 < \alpha < 2$.

Quantum algorithm for α-Rényi entropy: $\alpha \notin N$

Issue w/ the **basic** framework for Rényi entropy

multiplicative errors (Montanaro's algorithm yields no speedup in the worst case) $\rightarrow$ quantum advantage when $1/2 < \alpha < 2$.

New observations

- given $\mathbb{E}[X] \in [a, b]$ s.t. $b = O(a)$, we develop a variant of Montanaro's algorithm with a quadratic speedup.

Quantum algorithm for α-Rényi entropy: $\alpha \notin N$

Issue w/ the **basic** framework for Rényi entropy

multiplicative errors (Montanaro's algorithm yields no speedup in the worst case) $\rightarrow$ quantum advantage when $1/2 < \alpha < 2$.

New observations

- given $\mathbb{E}[X] \in [a, b]$ s.t. $b = O(a)$, we develop a variant of Montanaro's algorithm with a quadratic speedup.
- Let $P_\alpha(p) = \sum_i p_i^\alpha$, $\alpha_1, \alpha_2 > 0$ s.t. $\alpha_1/\alpha_2 = 1 \pm 1/\log(n)$,

$$P_{\alpha_1}(p) = \Theta(P_{\alpha_2}(p)^{\alpha_1/\alpha_2})$$

Use P_{α_2} to estimate $[a, b]$ for P_{α_1} where α_2 is closer to 1.

Quantum algorithm for α-Rényi entropy: $\alpha \notin N$

Issue w/ the **basic** framework for Rényi entropy

multiplicative errors (Montanaro's algorithm yields no speedup in the worst case) $\rightarrow$ quantum advantage when $1/2 < \alpha < 2$.

New observations

- given $\mathbb{E}[X] \in [a, b]$ s.t. $b = O(a)$, we develop a variant of Montanaro's algorithm with a quadratic speedup.
- Let $P_\alpha(p) = \sum_i p_i^\alpha$, $\alpha_1, \alpha_2 > 0$ s.t. $\alpha_1/\alpha_2 = 1 \pm 1/\log(n)$,

$$P_{\alpha_1}(p) = \Theta(P_{\alpha_2}(p)^{\alpha_1/\alpha_2})$$

  Use P_{α_2} to estimate $[a, b]$ for P_{α_1} where α_2 is closer to 1.
- Recursively solve the P_{α_2} case until $\alpha_2 \approx 1$ where a speed-up is already known.

Quantum algorithm for α-Rényi entropy: $\alpha \notin N$

Recursively call roughly $O(\log(n))$ times until $\alpha \approx 1$

- $\alpha > 1$: $\alpha \to \alpha(1 + \frac{1}{\log n})^{-1} \to \alpha(1 + \frac{1}{\log n})^{-2} \cdots$;
- $0 < \alpha < 1$: $\alpha \to \alpha(1 - \frac{1}{\log n})^{-1} \to \alpha(1 - \frac{1}{\log n})^{-2} \cdots$.

Quantum algorithm for α-Rényi entropy: $\alpha \notin N$

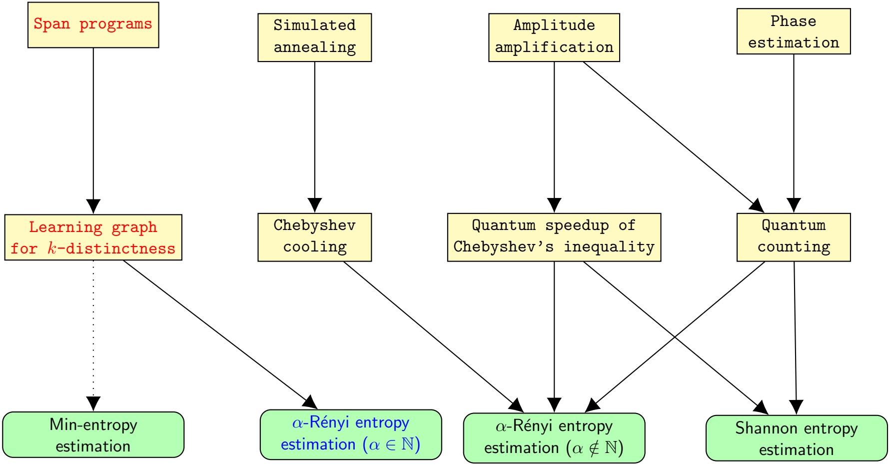
Recursively call roughly $O(\log(n))$ times until $\alpha \approx 1$

- $\alpha > 1$: $\alpha \to \alpha(1 + \frac{1}{\log n})^{-1} \to \alpha(1 + \frac{1}{\log n})^{-2} \cdots$;
- $0 < \alpha < 1$: $\alpha \to \alpha(1 - \frac{1}{\log n})^{-1} \to \alpha(1 - \frac{1}{\log n})^{-2} \cdots$.

Similarity and difference: cooling schedules in simulated annealing, volume estimation

- multi-section, multiplicative factors, similar design principle.
- adapt this design principle to our context.

Roadmap of quantum entropy estimation: for all $\alpha > 0$



Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

In the **sample** model, this problem has a good empirical estimator based on α-th frequency moment.

Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

In the **sample** model, this problem has a good empirical estimator based on α-th frequency moment.

Frequency moments

A sequence $a_1, \ldots, a_M \in [n]$ are given with the occurrences of $1, \ldots, n$ to be $m_1, \ldots, m_n$ respectively. You are asked to give a good approximation of $F_k = \sum_{i \in [n]} m_i^k$ for $k \in \mathbb{N}$.

Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

In the **sample** model, this problem has a good empirical estimator based on α-th frequency moment.

Frequency moments

A sequence $a_1, \ldots, a_M \in [n]$ are given with the occurrences of $1, \ldots, n$ to be $m_1, \ldots, m_n$ respectively. You are asked to give a good approximation of $F_k = \sum_{i \in [n]} m_i^k$ for $k \in \mathbb{N}$.

Empirical Estimation

Key observation:

$$\sum_{i \in [n]} p_i^\alpha \approx \sum_{i \in [n]} (\frac{m_i}{M})^k = F_k / M^k$$

Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

In the **sample** model, this problem has a good empirical estimator based on α-th frequency moment.

Frequency moments

A sequence $a_1, \ldots, a_M \in [n]$ are given with the occurrences of $1, \ldots, n$ to be $m_1, \ldots, m_n$ respectively. You are asked to give a good approximation of $F_k = \sum_{i \in [n]} m_i^k$ for $k \in \mathbb{N}$.

Empirical Estimation

Key observation:

$$\sum_{i \in [n]} p_i^{\alpha} \approx \sum_{i \in [n]} \left(\frac{m_i}{M}\right)^k = F_k / M^k$$

Note: this is not the best classical estimator. Why?

Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

Want to use this empirical estimator in the query model !!

Issues and Solutions

- **Issue 1**: How to generate M samples? cannot query M times (the same complexity as classical).

Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

Want to use this empirical estimator in the query model !!

Issues and Solutions

- **Issue 1**: How to generate M samples? cannot query M times (the same complexity as classical).
  Key observation: treat our quantum oracle O_p as a sequence of S samples. The α-frequency moment of $O_p(1), \ldots, O_p(S)$ is exactly $S^\alpha \sum_{i\in[n]} p_i^\alpha$.

Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

Want to use this empirical estimator in the query model !!

Issues and Solutions

- **Issue 1**: How to generate M samples? cannot query M times (the same complexity as classical).
  Key observation: treat our quantum oracle O_p as a sequence of S samples. The α-frequency moment of $O_p(1), \ldots, O_p(S)$ is exactly $S^\alpha \sum_{i \in [n]} p_i^\alpha$.
- **Issue 2**: quantum algorithm for α-frequency moment with query complexity (e.g., $o(n^{\frac{3}{4}(1-\frac{1}{\alpha})})$ [Montanaro 16]) where the "n" is actually "S" in this context.

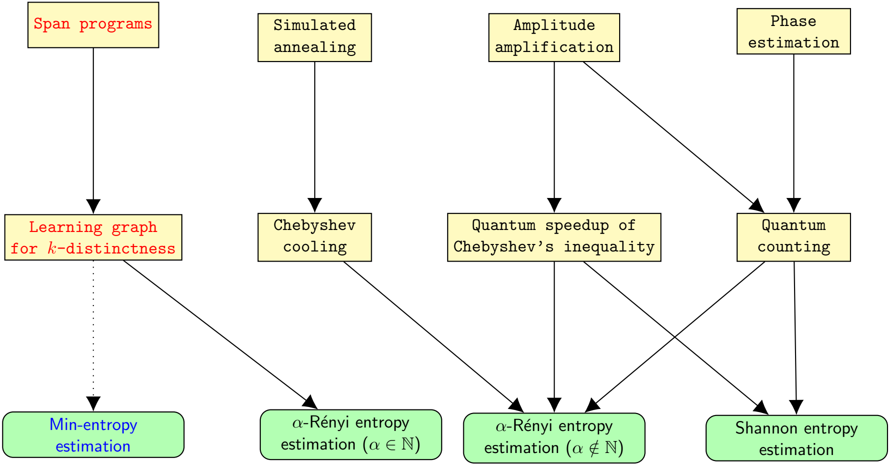
Quantum algorithm for α-Rényi entropy: $\alpha \geq 2 \in N$

Want to use this empirical estimator in the query model !!

Issues and Solutions

- **Issue 1**: How to generate M samples? cannot query M times (the same complexity as classical).
  Key observation: treat our quantum oracle O_p as a sequence of S samples. The α-frequency moment of $O_p(1), \ldots, O_p(S)$ is exactly $S^\alpha \sum_{i \in [n]} p_i^\alpha$.
- **Issue 2**: quantum algorithm for α-frequency moment with query complexity (e.g., $o(n^{\frac{3}{4}(1-\frac{1}{\alpha})})$ [Montanaro 16]) where the "n" is actually "S" in this context.
  Key observation: Roughly replace S by αn in the algorithm and redo the analysis.

Roadmap of quantum entropy estimation: for all $\alpha > 0$



The min-entropy (i.e., $\alpha = +\infty$) case: find the $\max_i p_i$

How about using integer α algorithm?

Intuitively, exists some α s.t., $H_\alpha(p)$ (sub-linear for any α) is a good enough approximation of $H_{\min}(p)$.

The min-entropy (i.e., $\alpha = +\infty$) case: find the $\max_i p_i$

How about using integer α algorithm?

Intuitively, exists some α s.t., $H_\alpha(p)$ (sub-linear for any α) is a good enough approximation of $H_{\min}(p)$.

NO, $\tilde{O}(\cdot)$ hides an exponential dependence on α.

The min-entropy (i.e., $\alpha = +\infty$) case: find the $\max_i p_i$

How about using integer α algorithm?

Intuitively, exists some α s.t., $H_\alpha(p)$ (sub-linear for any α) is a good enough approximation of $H_{\min}(p)$.

NO, $\tilde{O}(\cdot)$ *hides an exponential dependence on* α*.*

Another reduction to $\log(n)$-distinctness

- "Poissonized sampling": the numbers of occurrences of ith point are *pair-wise independent Poisson distributions*, parameterized by a guess value λ and p_i, $i \in [n]$.

The min-entropy (i.e., $\alpha = +\infty$) case: find the $\max_i p_i$

How about using integer α algorithm?

Intuitively, exists some α s.t., $H_\alpha(p)$ (sub-linear for any α) is a good enough approximation of $H_{\min}(p)$.

NO, $\tilde{O}(\cdot)$ hides an exponential dependence on α.

Another reduction to $\log(n)$-distinctness

- "Poissonized sampling": the numbers of occurrences of ith point are *pair-wise independent Poisson distributions*, parameterized by a guess value λ and p_i, $i \in [n]$.
- $\log(n)$ is a natural cut-off threshold for Poisson distribution.

The min-entropy (i.e., $\alpha = +\infty$) case: find the $\max_i p_i$

How about using integer α algorithm?

Intuitively, exists some α s.t., $H_\alpha(p)$ (sub-linear for any α) is a good enough approximation of $H_{\min}(p)$.

NO, $\tilde{O}(\cdot)$ hides an exponential dependence on α.

Another reduction to $\log(n)$-distinctness

- "Poissonized sampling": the numbers of occurrences of ith point are *pair-wise independent Poisson distributions*, parameterized by a guess value λ and p_i, $i \in [n]$.
- $\log(n)$ is a natural cut-off threshold for Poisson distribution.
- When $\lambda \cdot \max_i p_i$ passes this threshold, a $\log(n)$-collision can be found w.h.p.. Otherwise, update λ and try again.

Quantum lower bounds

Any quantum algorithm that approximates α-Rényi entropy of a discrete distribution on $[n]$ with success probability at least $2/3$ must use

- $\Omega(n^{\frac{1}{7\alpha}-o(1)})$ quantum queries when $0 < \alpha < \frac{3}{7}$.
- $\Omega(n^{\frac{1}{3}})$ quantum queries when $\frac{3}{7} \leq \alpha \leq 3$.
- $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}})$ quantum queries when $\alpha \geq 3$.
- $\Omega(\sqrt{n})$ quantum queries when $\alpha = +\infty$.

Quantum lower bounds

Any quantum algorithm that approximates α-Rényi entropy of a discrete distribution on $[n]$ with success probability at least $2/3$ must use

- $\Omega(n^{\frac{1}{7\alpha}-o(1)})$ quantum queries when $0 < \alpha < \frac{3}{7}$.
- $\Omega(n^{\frac{1}{3}})$ quantum queries when $\frac{3}{7} \leq \alpha \leq 3$.
- $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}})$ quantum queries when $\alpha \geq 3$.
- $\Omega(\sqrt{n})$ quantum queries when $\alpha = +\infty$.

Techniques:

- *Reductions* to the collision, Hamming weight, symmetry function problems.

Quantum lower bounds

Any quantum algorithm that approximates α-Rényi entropy of a discrete distribution on $[n]$ with success probability at least $2/3$ must use

- $\Omega(n^{\frac{1}{7\alpha}-o(1)})$ quantum queries when $0 < \alpha < \frac{3}{7}$.
- $\Omega(n^{\frac{1}{3}})$ quantum queries when $\frac{3}{7} \leq \alpha \leq 3$.
- $\Omega(n^{\frac{1}{2}-\frac{1}{2\alpha}})$ quantum queries when $\alpha \geq 3$.
- $\Omega(\sqrt{n})$ quantum queries when $\alpha = +\infty$.

Techniques:

- *Reductions* to the collision, Hamming weight, symmetry function problems.
- The *polynomial method* inspired by the collision lower bound for a better error dependence.

Outline

Questions

k-distinctness for super-constant k?

A problem interesting on its own. Our reduction to the min-entropy case just adds another motivation!

Questions

k-distinctness for super-constant k?

A problem interesting on its own. Our reduction to the min-entropy case just adds another motivation!

No quantum speed-up from existing algorithms by Ambainis and Belovs when $k = \log(n)$

Questions

k-distinctness for super-constant k?

A problem interesting on its own. Our reduction to the min-entropy case just adds another motivation!

No quantum speed-up from existing algorithms by Ambainis and Belovs when $k = \log(n)$

differences between quantum and classical?

At a high level, we want testers with correct expectations but small variances. Many techniques are proposed classically, and so different from ours.

Questions

k-distinctness for super-constant k?

A problem interesting on its own. Our reduction to the min-entropy case just adds another motivation!

No quantum speed-up from existing algorithms by Ambainis and Belovs when $k = \log(n)$

differences between quantum and classical?

At a high level, we want testers with correct expectations but small variances. Many techniques are proposed classically, and so different from ours.

Can we leverage the design principle of classical testers?

Thank You!!

Q & A

Classical estimators for Shannon entropy

Classical estimators for Shannon entropy

A first choice: empirical estimator. If we take M samples and occurences of $1, \ldots, n$ are $m_1, \ldots, m_n$ respectively, then empirically the distribution is $(\frac{m_1}{M}, \ldots, \frac{m_n}{M})$, and

$$H_{\mathrm{emp}}(p) = -\sum_{i \in [n]} \frac{m_i}{M} \log \frac{m_i}{M}.$$

Classical estimators for Shannon entropy

A first choice: empirical estimator. If we take M samples and occurences of $1, \ldots, n$ are $m_1, \ldots, m_n$ respectively, then empirically the distribution is $(\frac{m_1}{M}, \ldots, \frac{m_n}{M})$, and

$$H_{\mathrm{emp}}(p) = -\sum_{i \in [n]} \frac{m_i}{M} \log \frac{m_i}{M}.$$

To approximate $H(p)$ within error ϵ with high probability, need $M = \Theta(\frac{n}{\epsilon^2})$.

Classical estimators for Shannon entropy

A first choice: empirical estimator. If we take M samples and occurences of $1, \ldots, n$ are $m_1, \ldots, m_n$ respectively, then empirically the distribution is $(\frac{m_1}{M}, \ldots, \frac{m_n}{M})$, and

$$H_{\text{emp}}(p) = -\sum_{i \in [n]} \frac{m_i}{M} \log \frac{m_i}{M}.$$

To approximate $H(p)$ within error ϵ with high probability, need $M = \Theta(\frac{n}{\epsilon^2})$.

Not that bad compared to the best estimator, where $M = \Theta(\frac{n}{\epsilon^2 \log n})$.

Classical estimators for Shannon entropy

Construction of the best estimator:

Classical estimators for Shannon entropy

Construction of the best estimator:

- [VV 11]: A very clever (but complicated) application of linear programming under Poissonized samples (will explain later)

Classical estimators for Shannon entropy

Construction of the best estimator:

- [VV 11]: A very clever (but complicated) application of linear programming under Poissonized samples (will explain later)
- [JVHW 15]: polynomial approximation

Classical estimators for Rényi entropy

Classical estimators for Rényi entropy

Still, we can first consider the empirical estimator: $H_{\alpha,\mathrm{emp}}(p) = \frac{1}{1-\alpha} \log \sum_{i\in[n]} \left(\frac{m_i}{M}\right)^{\alpha}$.

Classical estimators for Rényi entropy

Still, we can first consider the empirical estimator: $H_{\alpha,\mathrm{emp}}(p) = \frac{1}{1-\alpha}\log\sum_{i\in[n]}\left(\frac{m_i}{M}\right)^{\alpha}$.

In [AOST 17], it is shown that the query complexity of the empirical estimator is

$$\begin{cases} \Theta(n^{\frac{1}{\alpha}}), & \text{when } 0 < \alpha < 1. \\ \Theta(n), & \text{when } \alpha > 1. \end{cases}$$

Classical estimators for Rényi entropy

Still, we can first consider the empirical estimator: $H_{\alpha,\mathrm{emp}}(p) = \frac{1}{1-\alpha} \log \sum_{i\in[n]} \left(\frac{m_i}{M}\right)^{\alpha}$.

In [AOST 17], it is shown that the query complexity of the empirical estimator is

$$\begin{cases} \Theta(n^{\frac{1}{\alpha}}), & \text{when } 0 < \alpha < 1. \\ \Theta(n), & \text{when } \alpha > 1. \end{cases}$$

Recall that the classical query complexity of Rényi entropy:

$$\begin{cases} O(\frac{n^{\frac{1}{\alpha}}}{\log n}) \text{ and } \Omega(n^{\frac{1}{\alpha}-o(1)}), & \text{when } 0 < \alpha < 1. \\ O(\frac{n}{\log n}) \text{ and } \Omega(n^{1-o(1)}), & \text{when } \alpha > 1, \alpha \notin \mathbb{N}. \\ \Theta(n^{1-\frac{1}{\alpha}}), & \text{when } \alpha > 1, \alpha \in \mathbb{N}. \end{cases}$$

Classical estimators for Rényi entropy

Main difference happens only when $\alpha > 1, \alpha \in \mathbb{N}$.

Classical estimators for Rényi entropy

Main difference happens only when $\alpha > 1, \alpha \in \mathbb{N}$.

This is not a minor point. In particular, 2-Rényi entropy (also known as the *collision entropy*) has classical query complexity $\Theta(\sqrt{n})$, which is much less than that of Shannon entropy ($\Theta(\frac{n}{\log n})$).

Classical estimators for Rényi entropy

Main difference happens only when $\alpha > 1, \alpha \in \mathbb{N}$.

This is not a minor point. In particular, 2-Rényi entropy (also known as the *collision entropy*) has classical query complexity $\Theta(\sqrt{n})$, which is much less than that of Shannon entropy ($\Theta(\frac{n}{\log n})$).

Collision entropy gives a rough estimation of Shannon entropy, and it measures the quality of random number generators and key derivation in cryptographic applications.

Classical estimators for Rényi entropy

When $\alpha > 1$ and $\alpha \in \mathbb{N}$, [AOST 17] proposes the following "bias-corrected" estimator:

$$H_{\alpha,\mathrm{bias}}(p) = \frac{1}{1-\alpha} \log \sum_{i \in [n]} \frac{m_i(m_i - 1) \cdots (m_i - \alpha + 1)}{M^{\alpha}}.$$

This is actually the best estimator with query complexity $\Theta(n^{1-\frac{1}{\alpha}})$.

Classical estimators for Rényi entropy

When $\alpha > 1$ and $\alpha \in \mathbb{N}$, [AOST 17] proposes the following "bias-corrected" estimator:

$$H_{\alpha,\mathrm{bias}}(p) = \frac{1}{1-\alpha} \log \sum_{i\in[n]} \frac{m_i(m_i-1)\cdots(m_i-\alpha+1)}{M^\alpha}.$$

This is actually the best estimator with query complexity $\Theta(n^{1-\frac{1}{\alpha}})$.

The analysis is mainly based on a "Possionization" technique.

An intuition of Possionization

An intuition of Possionization

When you are taking M samples where M is fixed, then the occurences of $1, \ldots, n$ are not independent. But if you take $M \sim \mathrm{Poi}(\lambda)$, then the occurence of i is $m_i \sim \mathrm{Poi}(\lambda p_i)$, and all m_i are pairwise independent.

An intuition of Possionization

When you are taking M samples where M is fixed, then the occurences of $1, \ldots, n$ are not independent. But if you take $M \sim \mathrm{Poi}(\lambda)$, then the occurence of i is $m_i \sim \mathrm{Poi}(\lambda p_i)$, and all m_i are pairwise independent.

Moreover, with high probability $\lambda/2 \leq M \leq 2\lambda$, so Possionization does not influence query complexity up to a constant.

An intuition of Possionization

When you are taking M samples where M is fixed, then the occurences of $1, \ldots, n$ are not independent. But if you take $M \sim \mathrm{Poi}(\lambda)$, then the occurence of i is $m_i \sim \mathrm{Poi}(\lambda p_i)$, and all m_i are pairwise independent.

Moreover, with high probability $\lambda/2 \leq M \leq 2\lambda$, so Possionization does not influence query complexity up to a constant.

A key reason to use bias-corrected estimator: if a random variable $X \sim \mathrm{Poi}(\lambda)$, then $\forall\, r \in \mathbb{N}$,

$$\mathbb{E}[X(X-1)\cdots(X-r+1)] = \lambda^r.$$

An intuition of Possionization

When you are taking M samples where M is fixed, then the occurences of $1, \ldots, n$ are not independent. But if you take $M \sim \mathrm{Poi}(\lambda)$, then the occurence of i is $m_i \sim \mathrm{Poi}(\lambda p_i)$, and all m_i are pairwise independent.

Moreover, with high probability $\lambda/2 \leq M \leq 2\lambda$, so Possionization does not influence query complexity up to a constant.

A key reason to use bias-corrected estimator: if a random variable $X \sim \mathrm{Poi}(\lambda)$, then $\forall\, r \in \mathbb{N}$,

$$\mathbb{E}[X(X-1)\cdots(X-r+1)] = \lambda^r.$$

This property helps a lot for the analysis in [AOST 17] based on Chernoff-Hoeffding inequality.

References

G. Valiant and P. Valiant.

Estimating the unseen: an n/log (n)-sample estimator for entropy and support size, shown optimal via new CLTs.

Proceedings of the forty-third annual ACM Symposium on Theory of Computing, ACM, 2011.

J. Jiao, K. Venkat, Y. Han, and T. Weissman.

Minimax estimation of functionals of discrete distributions.

IEEE Transactions on Information Theory, 61.5 (2015): 2835-2885.

J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi.

Estimating Rényi entropy of discrete distributions.

IEEE Transactions on Information Theory, 63.1 (2017): 38-56.

S. Bravyi, A. W. Harrow, and A. Hassidim.

Quantum algorithms for testing properties of distributions..

IEEE Transactions on Information Theory, 57.6 (2011): 3971-3981.

A. Montanaro.

Quantum speedup of Monte Carlo methods.

Proc. R. Soc. A, Vol. 471. No. 2181. The Royal Society, 2015.