Measuring Quantum Entropy

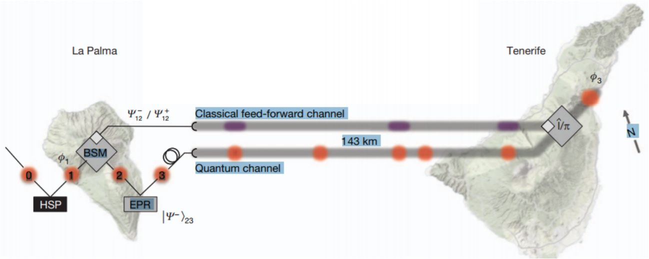
Jayadev Acharya[1] **Ibrahim Issa**[2] Nirmal Shende[1] Aaron Wagner[1]

[1]Cornell University [2]EPFL

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

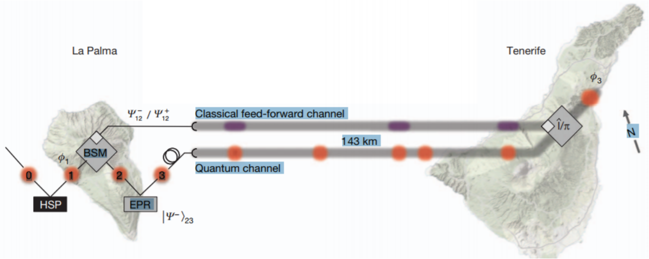
Quantum Information Processing, January 2018

Motivation



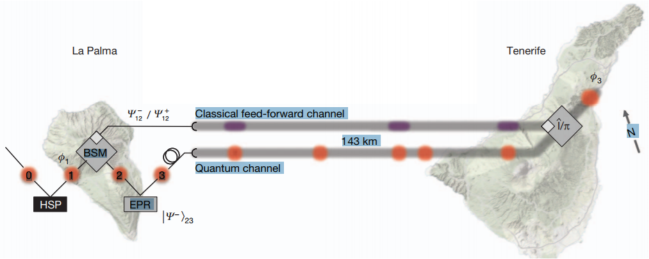
Quantum Teleportation Experiment [MHS+12]

Motivation



Quantum Teleportation Experiment [MHS+12]

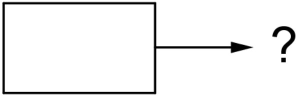
- Verified success by learning received state using 605 copies.



Quantum Teleportation Experiment [MHS+12]

- Verified success by learning received state using 605 copies.

- Quantum experiment



Main Question

(unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

Given a property f,
how many copies of ρ do we need to estimate f up to ε-accuracy?

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

Given a property f,
how many copies of ρ do we need to estimate f up to ε-accuracy?

-

$$\underbrace{\rho \otimes \rho \otimes \ldots \otimes \rho}$$

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

> Given a property f,
> how many copies of ρ do we need to estimate f up to ε-accuracy?

-

$$\underbrace{\rho \otimes \rho \otimes \ldots \otimes \rho}_{\text{Quantum measurement on } n \text{ copies}}$$

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d\times d}$

Given a property f,
how many copies of ρ do we need to estimate f up to ε-accuracy?

-

$$\underbrace{\rho \otimes \rho \otimes \ldots \otimes \rho}_{\text{Quantum measurement on } n \text{ copies}}$$

- Copy complexity:

$$C(f,d,\varepsilon) \stackrel{\text{def}}{=} \min \Big\{ n : \text{there exists } \hat{f} \text{ such that } \hat{f}(\rho^{\otimes n}) \text{ is an}$$

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

Given a property f,
how many copies of ρ do we need to estimate f up to ε-accuracy?

- $$\underbrace{\rho \otimes \rho \otimes \ldots \otimes \rho}_{\text{Quantum measurement on } n \text{ copies}}$$

- Copy complexity:

$$C(f, d, \varepsilon) \stackrel{\text{def}}{=} \min \Big\{ n : \text{there exists } \hat{f} \text{ such that } \hat{f}(\rho^{\otimes n}) \text{ is an } \pm \varepsilon \text{ estimate of } f(\rho) \text{ with probability} > 0.9 \Big\}$$

Main Question

- n copies of (unknown) mixed state $\rho \in \mathbb{C}^{d \times d}$

Given a property f,
how many copies of ρ do we need to estimate f up to ε-accuracy?

-

$$\underbrace{\rho \otimes \rho \otimes \ldots \otimes \rho}_{\text{Quantum measurement on } n \text{ copies}}$$

- Copy complexity:

$$C(f, d, \varepsilon) \stackrel{\text{def}}{=} \min \Big\{ n : \text{there exists } \hat{f} \text{ such that } \hat{f}(\rho^{\otimes n}) \text{ is an } \pm \varepsilon \text{ estimate of } f(\rho) \text{ with probability} > 0.9 \Big\}$$

[Montanaro and de Wolf '13]

Related Work

- Testing if ρ is maximally mixed requires $\Theta(d/\varepsilon^2)$ copies [OW15]

Related Work

- Testing if ρ is maximally mixed requires $\Theta(d/\varepsilon^2)$ copies [OW15]
- Quantum tomography: output $\hat{\rho}$ close to ρ with respect to some distance measure.

Related Work

- Testing if ρ is maximally mixed requires $\Theta(d/\varepsilon^2)$ copies [OW15]
- Quantum tomography: output $\hat{\rho}$ close to ρ with respect to some distance measure.
  - Requires $\tilde{O}(d^2)$ copies [HHJ+17,OW16,OW17]

Related Work

- Testing if ρ is maximally mixed requires $\Theta(d/\varepsilon^2)$ copies [OW15]
- Quantum tomography: output $\hat{\rho}$ close to ρ with respect to some distance measure.
  - Requires $\tilde{O}(d^2)$ copies [HHJ+17,OW16,OW17]
- Testing if $\rho = \sigma$ for some known σ (next talk!)

Quantum Entropy

Given a property f, how many copies of ρ do we need to estimate f up to $\pm\varepsilon$-accuracy?

Quantum Entropy

Given a property f, how many copies of ρ do we need to estimate f up to $\pm\varepsilon$-accuracy?

Properties of interest in this talk

von Neumann and Rényi entropies

Quantum Entropy

Given a property f, how many copies of ρ do we need to estimate f up to $\pm\varepsilon$-accuracy?

Properties of interest in this talk

von Neumann and Rényi entropies

- Considered a measure of quantum entanglement [Car12]

Quantum Entropy

Given a property f, how many copies of ρ do we need to estimate f up to $\pm\varepsilon$-accuracy?

Properties of interest in this talk

von Neumann and Rényi entropies

- Considered a measure of quantum entanglement [Car12]

- von Neumann entropy:

$$S(\rho) = -\mathrm{tr}\,(\rho \log \rho)$$

Quantum Entropy

Given a property f, how many copies of ρ do we need to estimate f up to $\pm\varepsilon$-accuracy?

Properties of interest in this talk

von Neumann and Rényi entropies

- Considered a measure of quantum entanglement [Car12]

- von Neumann entropy:

$$S(\rho) = -\mathrm{tr}\,(\rho \log \rho)$$

- Rényi entropy ($\alpha > 0$, $\alpha \neq 1$):

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{tr}(\rho^\alpha)$$

Quantum Entropy: Property of the Spectrum

- Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_d)$ be the spectrum of ρ

Quantum Entropy: Property of the Spectrum

- Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_d)$ be the spectrum of ρ
- von Neumann entropy:

$$S(\rho) = \sum_{i=1}^{d} \boldsymbol{\eta}_i \log \frac{1}{\boldsymbol{\eta}_i}$$

Quantum Entropy: Property of the Spectrum

- Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_d)$ be the spectrum of ρ
- von Neumann entropy:

$$S(\rho) = \sum_{i=1}^{d} \boldsymbol{\eta}_i \log \frac{1}{\boldsymbol{\eta}_i}$$

- Rényi entropy ($\alpha > 0$, $\alpha \neq 1$):

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \sum_{i=1}^{d} \boldsymbol{\eta}_i^\alpha$$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N} \setminus \{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N} \backslash \{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- More efficient than quantum tomography

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N}\backslash\{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- More efficient than quantum tomography
  - For $\alpha = 2$, $C(S_2, d, \varepsilon) = \Theta\left(\max\left\{\frac{\sqrt{d}}{\varepsilon^2}, \frac{d}{\varepsilon}\right\}\right)$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N} \backslash \{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- More efficient than quantum tomography
  - For $\alpha = 2$, $C(S_2, d, \varepsilon) = \Theta\left(\max\left\{\frac{\sqrt{d}}{\varepsilon^2}, \frac{d}{\varepsilon}\right\}\right)$
- Sublinear in the size of the input (d^2 is the natural dimensionality of the input)

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N}\backslash\{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- More efficient than quantum tomography
  - For $\alpha = 2$, $C(S_2, d, \varepsilon) = \Theta\left(\max\left\{\frac{\sqrt{d}}{\varepsilon^2}, \frac{d}{\varepsilon}\right\}\right)$
- Sublinear in the size of the input (d^2 is the natural dimensionality of the input)
- Classical estimation

$$C(H_\alpha, d, \varepsilon) = \Theta\left(\frac{d^{1-1/\alpha}}{\varepsilon^2}\right) \quad \text{[AOST15]}$$

Results: von Neumann Entropy

Theorem (AISW '17)

$$C(S, d, \varepsilon) \le O\left(\frac{d^2}{\varepsilon^2} + \frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right).$$

Results: von Neumann Entropy

Theorem (AISW '17)

$$C(S, d, \varepsilon) \leq O\left(\frac{d^2}{\varepsilon^2} + \frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right).$$

Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^2}{\varepsilon}\right)$ *copies to estimate von Neumann entropy.*

Results: von Neumann Entropy

Theorem (AISW '17)

$$C(S, d, \varepsilon) \leq O\left(\frac{d^2}{\varepsilon^2} + \frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right).$$

Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^2}{\varepsilon}\right)$ *copies to estimate von Neumann entropy.*

- Classical case [VV11,WY16,JVHW15]

$$C(H, d, \varepsilon) = \Theta\left(\frac{d}{\varepsilon \log d} + \frac{\log^2 d}{\varepsilon^2}\right)$$

Preliminaries: Unitarily Invariant Properties

- $f(\rho)$ is unitarily invariant if

$$f(U\rho U^\dagger) = f(\rho)$$

  for all unitary U

Preliminaries: Unitarily Invariant Properties

- $f(\rho)$ is unitarily invariant if

$$f(U\rho U^\dagger) = f(\rho)$$

  for all unitary U

  - Equivalently, f is a property of the spectrum $\boldsymbol{\eta}$ of ρ

Preliminaries: Unitarily Invariant Properties

- $f(\rho)$ is unitarily invariant if

$$f(U\rho U^\dagger) = f(\rho)$$

  for all unitary U

  - Equivalently, f is a property of the spectrum $\boldsymbol{\eta}$ of ρ
  - Examples: von Neumann and Rényi entropies, rank of ρ, etc.

Preliminaries: Unitarily Invariant Properties

- $f(\rho)$ is unitarily invariant if

$$f(U\rho U^\dagger) = f(\rho)$$

  for all unitary U

  - Equivalently, f is a property of the spectrum $\boldsymbol{\eta}$ of ρ
  - Examples: von Neumann and Rényi entropies, rank of ρ, etc.

A quantum measurement, called weak Schur sampling, is optimal for estimating unitarily invariant properties [KW01,CHW07,Har05,CHr06].

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \dots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \dots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\dots$

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

    $X_1^3 = 132.$

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

    $X_1^3 = 132.$

Histogram

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

$X_1^3 = 132.$ $\quad\quad$ $\boldsymbol{\lambda}_1 = 2$,

Histogram

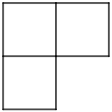
Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

$X_1^3 = 132.$

Histogram

$\boldsymbol{\lambda}_1 = 2$, $\boldsymbol{\lambda}_2 = 1$

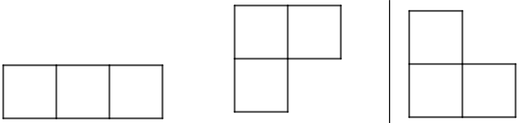


Young Diagram

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

$X_1^3 = 132.$ $\boldsymbol{\lambda}_1 = 2$, $\boldsymbol{\lambda}_2 = 1$ | $X_1^3 = 131.$



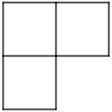
Histogram Young Diagram | Histogram

Preliminaries: Weak Schur Sampling

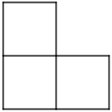
- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

$X_1^3 = 132.$ $\boldsymbol{\lambda}_1 = 2$, $\boldsymbol{\lambda}_2 = 1$ | $X_1^3 = 131.$ $\boldsymbol{\lambda}_1 = 2$,

Histogram



Young Diagram



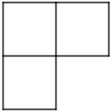
Histogram

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Consider a distribution over $\{1, 2, \ldots, d\}$, where i has probability $\boldsymbol{\eta}_i$.
  - Draw X^n independently from this distribution.
  - Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \ldots$ be such that
    - $\boldsymbol{\lambda}_1$ is the length of the longest non-decreasing subsequence of X^n
    - $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ is the *largest sum of lengths* of two disjoint non-decreasing subsequences
    - $\ldots$
  - Example: $d = n = 3$

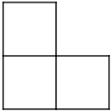
$X_1^3 = 132.$ $\boldsymbol{\lambda}_1 = 2$, $\boldsymbol{\lambda}_2 = 1$

Histogram

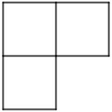


Young Diagram

$X_1^3 = 131.$ $\boldsymbol{\lambda}_1 = 2$, $\boldsymbol{\lambda}_2 = 1$



Histogram



Young Diagram

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Denote the whole procedure as $\boldsymbol{\lambda} \sim SW_{\boldsymbol{\eta}}$

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Denote the whole procedure as $\boldsymbol{\lambda} \sim SW_{\boldsymbol{\eta}}$
- $\boldsymbol{\lambda}$ is a partition of n, denoted by $\boldsymbol{\lambda} \vdash n$ and represented by a Young diagram:

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Denote the whole procedure as $\boldsymbol{\lambda} \sim SW_{\boldsymbol{\eta}}$
- $\boldsymbol{\lambda}$ is a partition of n, denoted by $\boldsymbol{\lambda} \vdash n$ and represented by a Young diagram:

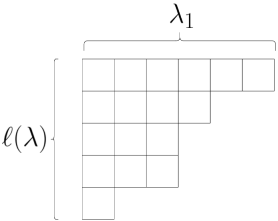


Figure: English Young diagram for the partition $\boldsymbol{\lambda} = (6, 4, 3, 3, 1)$.

Preliminaries: Weak Schur Sampling

- Weak Schur sampling output:
  - Denote the whole procedure as $\boldsymbol{\lambda} \sim SW_{\boldsymbol{\eta}}$
- $\boldsymbol{\lambda}$ is a partition of n, denoted by $\boldsymbol{\lambda} \vdash n$ and represented by a Young diagram:

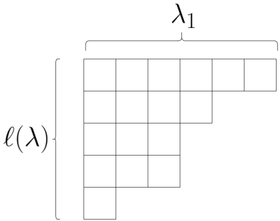


Figure: English Young diagram for the partition $\boldsymbol{\lambda} = (6, 4, 3, 3, 1)$.

- Since $\boldsymbol{\eta} - X^n - \boldsymbol{\lambda}(X^n)$ is a Markov chain,

Estimating a quantum state property is at least as hard as estimating the same property in the classical setting

Preliminaries: Definitions

- "Semi-standard" Young tableau: filled diagram with strictly increasing columns (top-to-bottom) and non-decreasing rows (left-to-right)

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 3 & \multicolumn{1}{|c}{} \\ \cline{1-1}\end{array}$$

Preliminaries: Definitions

- "Semi-standard" Young tableau: filled diagram with strictly increasing columns (top-to-bottom) and non-decreasing rows (left-to-right)

1	1
3	

- "Standard" Young tableau: semi-standard with strictly increasing rows

1	2
3	

Preliminaries: Definitions

- "Semi-standard" Young tableau: filled diagram with strictly increasing columns (top-to-bottom) and non-decreasing rows (left-to-right)

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}$$

- "Standard" Young tableau: semi-standard with strictly increasing rows

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}$$

- Schur polynomial

$$s_{\boldsymbol{\lambda}}(x_1, x_2, \ldots, x_d) = \sum_{\substack{T\text{:semi-standard tableau} \\ \text{of shape } \boldsymbol{\lambda}}} \prod_{i=1}^{d} x_i^{\{\#\text{ of occurences of } i \text{ in } T\}}$$

Preliminaries: Definitions

- "Semi-standard" Young tableau: filled diagram with strictly increasing columns (top-to-bottom) and non-decreasing rows (left-to-right)

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}$$

- "Standard" Young tableau: semi-standard with strictly increasing rows

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}$$

- Schur polynomial

$$s_{\boldsymbol{\lambda}}(x_1, x_2, \ldots, x_d) = \sum_{\substack{T\text{:semi-standard tableau} \\ \text{of shape } \boldsymbol{\lambda}}} \prod_{i=1}^{d} x_i^{\{\#\text{ of occurences of } i \text{ in } T\}}$$

$$SW_{\boldsymbol{\eta}}(\boldsymbol{\lambda}) = (\#\text{of standard Young tableaus of shape } \boldsymbol{\lambda}) \cdot s_{\boldsymbol{\lambda}}(\boldsymbol{\eta})$$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N} \backslash \{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N}\backslash\{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- Classical estimation

$$C(H_\alpha, d, \varepsilon) = \Theta\left(\frac{d^{1-1/\alpha}}{\varepsilon^2}\right) \quad \text{[AOST15]}$$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N}\backslash\{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- Classical estimation

$$C(H_\alpha, d, \varepsilon) = \Theta\left(\frac{d^{1-1/\alpha}}{\varepsilon^2}\right) \quad \text{[AOST15]}$$

Results: Rényi Entropy for Integral α

Theorem (AISW '17)

For $\alpha \in \mathbb{N}\backslash\{1\}$,

$$C(S_\alpha, d, \varepsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- Classical estimation

$$C(H_\alpha, d, \varepsilon) = \Theta\left(\frac{d^{1-1/\alpha}}{\varepsilon^2}\right) \quad \text{[AOST15]}$$

Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Design two density matrices ρ and σ with spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, respectively, such that

Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Design two density matrices ρ and σ with spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, respectively, such that

$$|S_\alpha(\boldsymbol{\eta}) - S_\alpha(\boldsymbol{\nu})| > 2\varepsilon$$

Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Design two density matrices ρ and σ with spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, respectively, such that

$$|S_\alpha(\boldsymbol{\eta}) - S_\alpha(\boldsymbol{\nu})| > 2\varepsilon$$

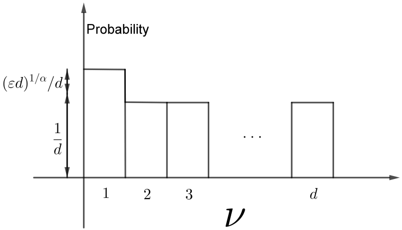
- Subproblem: generate $\boldsymbol{\lambda}$ from either $\boldsymbol{\eta}$ or $\boldsymbol{\nu}$, and distinguish between the two cases

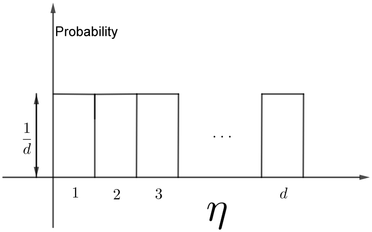
Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Design two density matrices ρ and σ with spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, respectively, such that

$$|S_\alpha(\boldsymbol{\eta}) - S_\alpha(\boldsymbol{\nu})| > 2\varepsilon$$

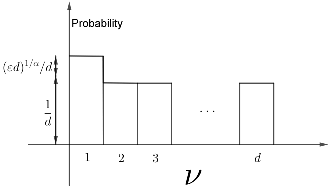
- Subproblem: generate $\boldsymbol{\lambda}$ from either $\boldsymbol{\eta}$ or $\boldsymbol{\nu}$, and distinguish between the two cases





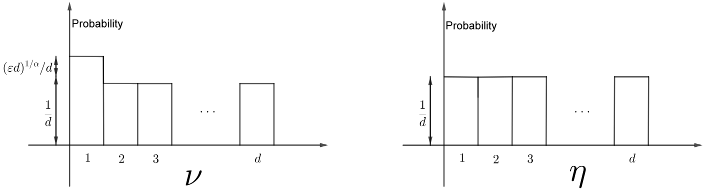
Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Generate $\boldsymbol{\lambda}$ from either $\boldsymbol{\eta}$ or $\boldsymbol{\nu}$, and distinguish between the two cases



Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Generate $\boldsymbol{\lambda}$ from either $\boldsymbol{\eta}$ or $\boldsymbol{\nu}$, and distinguish between the two cases

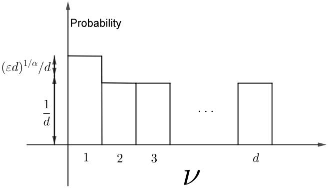


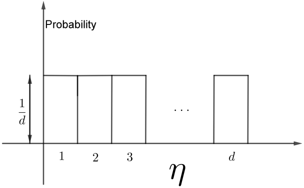
- Unless $n = \Omega\left(\frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right)$, we cannot tell them apart since

$$2d_{TV}\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right)^2 \leq \chi^2\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right) \leq 0.01$$

Rényi Entropy $S_\alpha(\rho)$ for Integral α: Converse

- Generate $\boldsymbol{\lambda}$ from either $\boldsymbol{\eta}$ or $\boldsymbol{\nu}$, and distinguish between the two cases

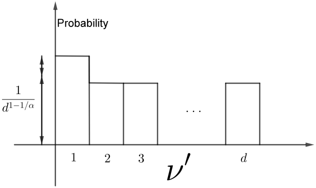


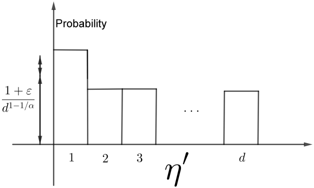


- Unless $n = \Omega\left(\frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right)$, we cannot tell them apart since

$$2d_{TV}\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right)^2 \leq \chi^2\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right) \leq 0.01$$

- In classical case,





Results: von Neumann Entropy

Theorem (AISW '17)

$$C(S, d, \varepsilon) \leq O\left(\frac{d^2}{\varepsilon^2} + \frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right).$$

Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^2}{\varepsilon}\right)$ *copies to estimate von Neumann entropy.*

Results: von Neumann Entropy

Theorem (AISW '17)

$$C(S, d, \varepsilon) \le O\left(\frac{d^2}{\varepsilon^2} + \frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right).$$

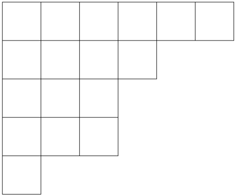
Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^2}{\varepsilon}\right)$ copies to estimate von Neumann entropy.

Empirical Estimator (EYD)

- Assume the eigenvalues of ρ are sorted: $\boldsymbol{\eta}_1 \geq \boldsymbol{\eta}_2 \geq \dots$

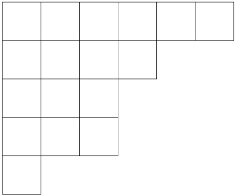
Empirical Estimator (EYD)

- Assume the eigenvalues of ρ are sorted: $\boldsymbol{\eta}_1 \geq \boldsymbol{\eta}_2 \geq \dots$
- $\boldsymbol{\lambda} \sim SW_\eta$



Empirical Estimator (EYD)

- Assume the eigenvalues of ρ are sorted: $\boldsymbol{\eta}_1 \geq \boldsymbol{\eta}_2 \geq \ldots$
- $\boldsymbol{\lambda} \sim SW_\eta$

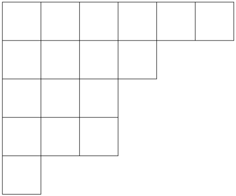


- The empirical Young diagram algorithm sets

$$\widehat{\boldsymbol{\eta}}_i := \frac{\boldsymbol{\lambda}_i}{n}$$

Empirical Estimator (EYD)

- Assume the eigenvalues of ρ are sorted: $\boldsymbol{\eta}_1 \geq \boldsymbol{\eta}_2 \geq \dots$
- $\boldsymbol{\lambda} \sim SW_\eta$



- The empirical Young diagram algorithm sets

$$\widehat{\boldsymbol{\eta}}_i := \frac{\boldsymbol{\lambda}_i}{n}$$

- It is known that [ARS88,KW01]

$$\mathbf{E}_{SW_\eta}\left[\frac{\boldsymbol{\lambda}}{n}\right] \xrightarrow{n\to\infty} \boldsymbol{\eta}$$

von Neumann Entropy

- The empirical estimate $\widehat{S(\boldsymbol{\lambda})}$ is

$$\widehat{S(\boldsymbol{\lambda})} = \sum_{i=1}^{d} \frac{\boldsymbol{\lambda}_i}{n} \log \frac{n}{\boldsymbol{\lambda}_i}$$

von Neumann Entropy

- The empirical estimate $\widehat{S(\boldsymbol{\lambda})}$ is

$$\widehat{S(\boldsymbol{\lambda})} = \sum_{i=1}^{d} \frac{\boldsymbol{\lambda}_i}{n} \log \frac{n}{\boldsymbol{\lambda}_i}$$

Theorem

The empirical entropy estimate satisfies:

$$\mathbf{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] \leq O\left(\frac{d^4}{n^2} + \frac{d^2}{n} + \frac{\log^2 n}{n}\right).$$

Hence,

$$C(S, d, \varepsilon) \leq O\left(\frac{d^2}{\varepsilon^2} + \frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right).$$

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

- $\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) \leq \frac{\log^2 n}{n}$

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

- $\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) \leq \frac{\log^2 n}{n}$
  - Proof similar to Lipschitzness of empirical entropy

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

- $\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) \leq \frac{\log^2 n}{n}$
    - Proof similar to Lipschitzness of empirical entropy

- Bias is small by concentration results [OW17]

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

- $\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) \leq \frac{\log^2 n}{n}$
  - Proof similar to Lipschitzness of empirical entropy
- Bias is small by concentration results [OW17]
  - $\chi^2(\boldsymbol{\lambda}/n, \boldsymbol{\eta}) \leq d^2/n$

von Neumann Entropy - Proof Sketch

$$\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] = \underbrace{\left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2}_{\text{bias}} + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right)$$

- $\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) \leq \frac{\log^2 n}{n}$
  - Proof similar to Lipschitzness of empirical entropy

- Bias is small by concentration results [OW17]
  - $\chi^2(\boldsymbol{\lambda}/n, \boldsymbol{\eta}) \leq d^2/n$
  - $|\boldsymbol{\eta}_i - \mathbb{E}\left[(\widehat{\boldsymbol{\eta}}_i)\right]| \leq 2\sqrt{\frac{\min\{1, \boldsymbol{\eta}_i d\}}{n}}$

Lower Bounds for Empirical Estimator

- Strong error probability bound:

Lower Bounds for Empirical Estimator

- Strong error probability bound:

Theorem (AISW '17)

For small enough ε, if ρ is maximally mixed and $n \leq O(d^2/\varepsilon^2)$, then

$$\Pr\left(\sum_{i=1}^{d}\left|\frac{\boldsymbol{\lambda}_i}{n} - \frac{1}{d}\right| > \varepsilon\right) > 1 - e^{-\Omega(d)}.$$

Lower Bounds for Empirical Estimator

- Strong error probability bound:

Theorem (AISW '17)

For small enough ε, if ρ is maximally mixed and $n \leq O(d^2/\varepsilon^2)$, then

$$\Pr\left(\sum_{i=1}^{d}\left|\frac{\boldsymbol{\lambda}_i}{n} - \frac{1}{d}\right| > \varepsilon\right) > 1 - e^{-\Omega(d)}.$$

  - Previously the best known lower bound was a constant (0.01) [OW15].

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy, and any Rényi entropy of order $\alpha > 1$.

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy, and any Rényi entropy of order $\alpha > 1$.

  - Our upper bound for the empirical estimator $O(d^2/\varepsilon^2)$ is therefore tight in terms of the dependence on d

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy, and any Rényi entropy of order $\alpha > 1$.

- Our upper bound for the empirical estimator $O(d^2/\varepsilon^2)$ is therefore tight in terms of the dependence on d
- Proof: suppose ρ is maximally mixed. Then $S(\rho) = S_\alpha(\rho) = \log d$

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy, and any Rényi entropy of order $\alpha > 1$.

- Our upper bound for the empirical estimator $O(d^2/\varepsilon^2)$ is therefore tight in terms of the dependence on d
- Proof: suppose ρ is maximally mixed. Then $S(\rho) = S_\alpha(\rho) = \log d$

$$\log d - \widehat{S_\alpha}(\boldsymbol{\lambda}) \geq \log d - \widehat{S}(\boldsymbol{\lambda})$$

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy, and any Rényi entropy of order $\alpha > 1$.

- Our upper bound for the empirical estimator $O(d^2/\varepsilon^2)$ is therefore tight in terms of the dependence on d
- Proof: suppose ρ is maximally mixed. Then $S(\rho) = S_\alpha(\rho) = \log d$

$$\begin{aligned} \log d - \widehat{S_\alpha}(\boldsymbol{\lambda}) &\geq \log d - \widehat{S}(\boldsymbol{\lambda}) \\ &= d_{KL}\left(\frac{\boldsymbol{\lambda}}{n}, \boldsymbol{\eta}\right) \end{aligned}$$

Lower Bounds for Empirical Estimator

- An important consequence of the concentration result

Theorem

For small enough ε, the empirical estimator requires $\Omega(d^2/\varepsilon)$ copies to estimate von Neumann entropy, and any Rényi entropy of order $\alpha > 1$.

- Our upper bound for the empirical estimator $O(d^2/\varepsilon^2)$ is therefore tight in terms of the dependence on d
- Proof: suppose ρ is maximally mixed. Then $S(\rho) = S_\alpha(\rho) = \log d$

$$\begin{aligned} \log d - \widehat{S_\alpha}(\boldsymbol{\lambda}) &\geq \log d - \widehat{S}(\boldsymbol{\lambda}) \\ &= d_{KL}\left(\frac{\boldsymbol{\lambda}}{n}, \boldsymbol{\eta}\right) \\ &\geq 2d_{TV}^2\left(\frac{\boldsymbol{\lambda}}{n}, \boldsymbol{\eta}\right) \end{aligned}$$

Results: Non-integral Rényi Entropy

Theorem (AISW '17)

For $\alpha > 1$,

$$C(S, d, \varepsilon) \leq O\left(\frac{d^2}{\varepsilon^2}\right).$$

Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^2}{\varepsilon}\right)$ copies to estimate $S_\alpha(\rho)$.

Results: Non-integral Rényi Entropy

Theorem (AISW '17)

For $\alpha > 1$,

$$C(S, d, \varepsilon) \leq O\left(\frac{d^2}{\varepsilon^2}\right).$$

Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^2}{\varepsilon}\right)$ *copies to estimate* $S_\alpha(\rho)$.

Theorem (AISW '17)

For $\alpha < 1$,

$$C(S, d, \varepsilon) \leq O\left(\frac{d^{2/\alpha}}{\varepsilon^{2/\alpha}}\right).$$

Moreover, the empirical estimator (EYD) requires $\Omega\left(\frac{d^{1+1/\alpha}}{\varepsilon^{1/\alpha}}\right)$ *copies to estimate* $S_\alpha(\rho)$.

Summary

- Integral Rényi entropy with $\alpha > 1$:

$$C(S_\alpha, d, \epsilon) = \Theta\left(\max\left\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right\}\right)$$

- von Neumann and non-integral Rényi entropies:

Table: Copy complexity of empirical estimators

α	Upper Bound	Lower Bound
von Neumann	$O(d^2/\varepsilon^2)$	$\Omega(d^2/\varepsilon)$
$\alpha > 1$	$O(d^2/\varepsilon^2)$	$\Omega(d^2/\varepsilon)$
$\alpha < 1$	$O(d^{2/\alpha}/\varepsilon^{2/\alpha})$	$\Omega(d^{1+1/\alpha}/\varepsilon^{1/\alpha})$

Open Questions

- Exact characterization of the copy complexity of von Neumann entropy, and Rényi entropy for non-integer α

Open Questions

- Exact characterization of the copy complexity of von Neumann entropy, and Rényi entropy for non-integer α
- Characterization of the copy complexity for restricted quantum measurements

Open Questions

- Exact characterization of the copy complexity of von Neumann entropy, and Rényi entropy for non-integer α
- Characterization of the copy complexity for restricted quantum measurements
  - Example: non-adaptive measurements

Open Questions

- Exact characterization of the copy complexity of von Neumann entropy, and Rényi entropy for non-integer α
- Characterization of the copy complexity for restricted quantum measurements
  - Example: non-adaptive measurements
- Conjecture: $S_2(\rho)$ is the easiest entropy to estimate