Estimating Quantum Entropy

Jayadev Acharya, *Member, IEEE*, Ibrahim Issa, *Member, IEEE*, Nirmal V. Shende,
and Aaron B. Wagner, *Fellow, IEEE*

***Abstract*—The entropy of a quantum system is a measure of its randomness, and has applications in measuring quantum entanglement. We study the problem of estimating the von Neumann entropy, $S(\rho)$, and Rényi entropy, $S_\alpha(\rho)$ of an unknown mixed quantum state ρ in d dimensions, given access to independent copies of ρ. We provide algorithms with copy complexity $O(d^{2/\alpha})$ for estimating $S_\alpha(\rho)$ for $\alpha < 1$, and copy complexity $O(d^2)$ for estimating $S(\rho)$, and $S_\alpha(\rho)$ for non-integral $\alpha > 1$. These bounds are at least quadratic in d, which is the order dependence on the number of copies required for estimating the entire state ρ. For integral $\alpha > 1$, on the other hand, we provide an algorithm for estimating $S_\alpha(\rho)$ with a sub-quadratic copy complexity of $O(d^{2-2/\alpha})$, and we show the optimality of the algorithms by providing a matching lower bound.**

***Index Terms*—Quantum information, von Neumann entropy, Renyi entropy, entropy estimation, weak Schur sampling, copy complexity.**

I. Introduction

WE CONSIDER how to estimate the mixedness or noisiness of a quantum state using measurements of independent copies of the state. Mixed quantum states can arise in practice in various ways. Classical stochasticity can be intentionally introduced when the state is originally prepared. Pure states can become mixed by a quantum measurement. And the states of the subsystems of bipartite states can be mixed even when the overall bipartite state is pure, which forms the basis for purification.

In the third case, the level of mixedness of the subsystems indicates the level of entanglement in the pure, bipartite system. The possibility of entanglement of two separated systems is arguably the most curious, and the most powerful, way in which quantum systems differ from classical ones. Indeed, entanglement has been fruitfully exploited as a resource in a number of quantum information processing protocols (e.g., [1]–[5]). The subsystems of a pure bipartite state are pure if and only if the bipartite state itself is unentangled, and likewise they are maximally mixed if and only if the bipartite state is maximally entangled. Thus the mixedness of the subsystems' states can be used as a measure of entanglement of the bipartite system [4, Th. 8.6].

Mixedness can be measured in multiple ways. We shall use the von Neumann and (the family of) Rényi entropies, which correspond to the classical Shannon and (the family of) Rényi entropies of the eigenvalues of the density operator of the state, respectively. A density matrix (or operator) ρ is a complex positive semidefinite matrix with unit trace; thus its eigenvalues are nonnegative and sum to one. The von Neumann entropy of a density matrix ρ is

$$S(\rho) \overset{\text{def}}{=} -\mathrm{tr}(\rho \log \rho).$$

For $\alpha > 0, \alpha \neq 1$, the Rényi entropy of order α of ρ is

$$S_\alpha(\rho) \overset{\text{def}}{=} \frac{1}{1-\alpha} \log \mathrm{tr}\big(\rho^\alpha\big).$$

Similar to Shannon entropy, in the limit of $\alpha \to 1$, $\lim_{\alpha\to 1} S_\alpha(\rho) = S(\rho)$.

The classical Shannon and Rényi entropies are well-accepted measures of randomness, and can be derived axiomatically [6, pp. 25–27]. Both the classical and quantum versions can be justified operationally as a measure of compressibility [6]–[9]. The quantum versions have been explicitly proposed for quantifying entanglement in certain contexts [10].

In principle, both the von Neumann and Rényi entropies for a quantum state ρ can be computed if the state is known. We consider how to estimate these quantities for an unknown state given independent copies of the state, to which arbitrary quantum measurements followed by arbitrary classical computation can be applied. This problem arises when characterizing a completely unknown system and when one seeks to experimentally verify that a system is behaving as desired. Since generating independent copies of a state can be quite costly in the quantum setting [11], [12], it is desirable to minimize the number of copies of the state that are required to estimate the von Neumann and Rényi entropies to a desired precision and confidence. We thus adopt this *copy complexity* as our figure-of-merit.

Our results are summarized as follows. We provide algorithms with copy complexity $O(d^{2/\alpha})$ for estimating $S_\alpha(\rho)$ for $\alpha < 1$, and copy complexity $O(d^2)$ for estimating $S(\rho)$, and $S_\alpha(\rho)$ for non-integral $\alpha > 1$. These bounds are at least

Manuscript received February 15, 2020; revised June 10, 2020; accepted July 22, 2020. Date of publication August 10, 2020; date of current version September 22, 2020. This work was supported in part by U.S. National Science Foundation under Grant CCF-1704443, Grant CCF-1513858, and Grant CCF-1815893; and in part by the U.S. Army Research Office under Grant W911NF-18-1-0426. This work was presented at the 21st Annual Conference on Quantum Information Processing, Delft, Jan. 2018, and at the IEEE International Symposium on Information Theory, Paris, Jul. 2019. *(Corresponding author: Jayadev Acharya.)*

Jayadev Acharya and Aaron B. Wagner are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: acharya@cornell.edu; wagner@cornell.edu).

Ibrahim Issa was with School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the School of Electrical and Computer Engineering, American University of Beirut, Beirut, Lebanon (e-mail: ii19@aub.edu.lb).

Nirmal V. Shende was with Cornell University, Ithaca, NY 14853 USA. He is now with Marvell Semiconductor, Santa Clara, CA 95054 USA (e-mail: nshende@marvell.com).

Digital Object Identifier 10.1109/JSAIT.2020.3015235

quadratic in d, which is the order dependence on the number of copies required for learning the entire state ρ. For integral $\alpha > 1$, on the other hand, we provide an algorithm for estimating $S_\alpha(\rho)$ with a sub-quadratic copy complexity of $O(d^{2-2/\alpha})$, and we show the optimality of the algorithms by providing a matching lower bound.

A. Related Work

Our work is related to symmetric distribution property estimation in classical setting and the property estimation of quantum states (as in the set-up of this paper). We briefly mention some closely related works. The reader is referred to Montanaro and Wolf [13] and Wright [14] for additional references.

1) Symmetric Property Estimation of Discrete Distributions: For the Shannon entropy $H(p) \overset{\text{def}}{=} -\sum_x p(x)\log p(x)$, a long line of work culminated in [15]–[17] showing that the sample complexity of estimating Shannon entropy to additive ε is $\Theta\Big(\frac{d}{\varepsilon \log d} + \frac{\log^2 d}{\varepsilon^2}\Big)$.

The problem of estimating Rényi entropy $H_\alpha(p) \overset{\text{def}}{=} \log\left(\sum_x p(x)^\alpha\right)/(1-\alpha)$, was studied in [18]–[20]. The sample complexity dependence in the classical setting seems to suggest the same qualitative behavior as our results. They show that for $\alpha < 1$, the sample complexity is $O\Big(\frac{d^{1/\alpha}}{\varepsilon^{1/\alpha}\log d}\Big)$, and for $\alpha > 1, \alpha \notin \mathbb{N}$, it is $O\Big(\frac{d}{\varepsilon^{1/\alpha}\log d}\Big)$. Moreover, their information-theoretic lower bounds show that the exponent of d cannot be improved by any algorithm. Finally for $\alpha \in \{2, 3, \ldots\}$, they show that the sample complexity $\Theta\Big(\frac{d^{1-1/\alpha}}{\varepsilon^2}\Big)$ up to constant factors. We note that this complexity is indeed one of the terms in our copy complexity for integral α, which is the dominant term for large n.

2) Quantum Property Estimation of Mixed States: There are now many works on the related problem of quantum property *testing*, where the goal is to find the copy complexity of deciding whether a mixed state has a certain property of interest, and on the problem of quantum tomography, where the goal is to learn the entire density matrix ρ. The copy complexity of quantum tomography is quadratic in d, and the complexity for tomography in various distance measures have been studied in [21]–[23]. Reference [24, Ch. 6] provides a number of results on universal quantum information processing, which considers various problems including tomography in the asymptotic setting where n is large.

Testing whether ρ has a particular unitarily invariant property of interest was studied in [25] for a number of properties. Recently, [26] obtained tight bounds on the copy complexity of testing whether an unknown density matrix is equal to a known density matrix. The optimal measurement schemes for some of these problems can be computationally expensive. Testing properties under simpler *local measurements* was studied recently in [27]. Reference [28] considers testing various properties with particular emphasis on the setting where only local measurements on the copies is allowed.

In a personal communication, Bavarian *et al.* [29] claim an algorithm with copy complexity $O(d^2/\varepsilon)$ for the von Neumann entropy estimation, which is an ε multiplicative factor improvement over our bound.

3) Quantum Algorithms for Classical and Quantum Distribution Properties: Testing and estimating distribution properties using quantum queries has been considered by various authors. Problems of testing properties such as uniformity, identity, closeness under the regular quantum query model, and conditional quantum query models have been studied in [30]–[32].

Recently Li and Wu [33] studied the quantum query complexity of estimating entropy of discrete distributions. They provide bounds on the query complexity for estimating von Neumann entropy, and Rényi entropy. For certain values of α, the bounds on query complexity can in fact be at times quadratically better than the corresponding sample complexity bounds. Subramanian and Hsieh [34] consider the problem of estimating α-Rényi entropy under a different model, in particular, the purified quantum query access model. In this scenario, a quantum oracle generates a purified version of the input state $\rho \in \mathbb{C}^{d\times d}$, with dimensions $(d+a)\times(d+a)$. The authors then provide a quantum sampling method to estimate Rényi entropy using $\tilde{O}((da\alpha/\delta\epsilon)^2)$ queries. Other recent papers have considered more practical considerations for the estimation of quantum properties, where they have access to the purified quantum state. In particular, Cincio *et al.* [35], Johri *et al.* [36], and Subasi *et al.* [37] demonstrate short-depth quantum circuits that improve and generalize the Swap test in order to estimate $Tr(\rho^k)$, with special emphasis for the case $k = 2$.

B. Organization

The paper is organized as follows. The next section contains a precise formulation and statement of our results. Section III provides the preliminary results needed for setting up the paper. Section IV proves our bounds for integral, but non-unity, order Rényi entropy. Section V proves the upper bounds for von Neumann entropy, and Section VI proves the upper bounds for Rényi entropy of non-integral orders.

II. Formulation and Description of Results

A. Property Estimation

A *property* $f(\rho)$ maps a mixed state ρ to $\mathbb{R}$. Given n and d, an *estimator* is a set of measurement matrices $\{M_m\}_{m=1}^\infty$ for the state space $\mathbb{C}^{d^n\times d^n}$ and a "classical processor" $g(\cdot)$, which maps the natural numbers to $\mathbb{R}$. Given n copies of a state ρ, the estimator proceeds by applying the measurement $\{M_m\}_{m=1}^\infty$ to the state $\rho^{\otimes n}$ and then applying $g(\cdot)$ to the resulting outcome. Given a property f, accuracy parameter ε, error parameter δ, and access to n independent copies of a mixed state ρ, we seek an estimator $\hat{f}$ such that with probability at least $1-\delta$

$$\left|f(\rho) - \hat{f}\big(\rho^{\otimes n}\big)\right| < \varepsilon.$$

The *copy complexity* of f is

$$C(f, d, \varepsilon, \delta) \overset{\text{def}}{=} \min\Big\{n : \exists \hat{f} : \forall \rho, \hat{f} \text{ is a } \pm\varepsilon \text{ estimate of } f(\rho) \text{ with probability} > 1-\delta\Big\},$$

the minimum number of copies required to solve the problem. Throughout this paper we will consider δ to be a constant, say 1/3. We can boost the error to any δ by repeating the estimation task $O(\log(1/\delta))$ times and taking the median of the outcomes. We denote

$$C(f, d, \varepsilon) \overset{\text{def}}{=} C(f, d, \varepsilon, 1/3). \tag{1}$$

B. Unitarily Invariant Properties

Suppose $U(d)$ is the set of all $d \times d$ unitary matrices.

Definition 1: A property $f(\rho)$ is called *unitarily invariant* if $f(U\rho U^\dagger) = f(\rho)$ for all $U \in U(d)$.

Unitarily invariant properties are functions of only the spectrum (that is, the multiset of eigenvalues) of the density matrix. Since density matrices are positive semi-definite with unit trace, we can view the eigenvalues as a distribution over some set. Unitarily invariant properties are analogous to symmetric properties in classical distributions.

For a density matrix with eigenvalues $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_d$, we have $S(\rho) = -\sum_i \boldsymbol{\eta}_i \log \boldsymbol{\eta}_i$, and $\mathcal{S}_\alpha(\rho) = \log(\sum_i \boldsymbol{\eta}_i^\alpha)/(1-\alpha)$. Quantum entropy can be viewed as the classical entropy of a distribution defined by $\boldsymbol{\eta}$, and in particular is unitarily invariant.

Working with unitarily invariant properties is greatly simplified by the following powerful result (see [13, Sec. 4.2.2] and the references therein).

Lemma 1: A quantum measurement called *weak Schur sampling* is optimal for estimating unitarily invariant properties.

Weak Schur sampling is discussed in Section III-A1.

C. Our Results

We will use the standard asymptotic notation. We are interested in characterizing the dependence of $C(S, d, \varepsilon)$, and $C(S_\alpha, d, \varepsilon)$, as a function of d and ε. We assume the parameter α to be a constant, and focus on only the growth rate as a function of d and ε.

Our results are summarized in Table I and Table II. Similar to the sample complexity of estimating Rényi entropies of classical distributions from samples, our bounds are also dependent on whether α is less than one, and whether it is an integer. (See [19, Tab. 1], and Section I-A1 for the sample complexity in classical settings.) We organize our results as a function of α as follows.

When interpreting our results it is useful to recall that the copy complexity of quantum tomography, where the goal is to estimate the density matrix ρ in trace distance is $O(d^2)$.

1) Integral $\alpha > 1$: We obtain our most optimistic and conclusive results in this case. In Theorem 1, we show that $C(S_\alpha, d, \varepsilon) = \Theta\Big(\max\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\}\Big)$. We note that the lower bounds here hold for *all estimators*, not just of the estimators used in the upper bound. Furthermore, these bounds are sub-quadratic in d. Namely, we can estimate the Rényi entropy of integral orders even before we have enough copies to perform full tomography. The upper bounds are established by analyzing certain polynomials from representation theory that are related to the central characters of the symmetric group. For the lower bound, we design the spectrums of two mixed states such that their Rényi entropy differ by at least ε, but require a large copy complexity to distinguish between them. We use various properties of Schur polynomials and other properties of integer partitions [38], [39].

TABLE I
Copy Complexity of $\mathcal{S}_\alpha(\rho)$ for Integral $\alpha > 1$

Upper Bound	Lower Bound
$O\left(\max\left\{\frac{d^{2-\frac{2}{\alpha}}}{\varepsilon^{\frac{2}{\alpha}}}, \frac{d^{1-\frac{1}{\alpha}}}{\varepsilon^2}\right\}\right)$	$\Omega\left(\max\left\{\frac{d^{2-\frac{2}{\alpha}}}{\varepsilon^{\frac{2}{\alpha}}}, \frac{d^{1-\frac{1}{\alpha}}}{\varepsilon^2}\right\}\right)$

TABLE II
Copy Complexity of the Empirical Estimator

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

Remark 1: The first term in the complexity dominates when $\varepsilon < 1/\sqrt{d}$, and is identical to the sample complexity of estimating Rényi entropy in the classical setting.

2) $\alpha < 1$: We analyze the Empirical Young Diagram (EYD) algorithm [40]–[43] for estimating $\mathcal{S}_\alpha(\rho)$ for $\alpha < 1$. The EYD algorithm is similar to using a plug-in estimate of the empirical distribution to estimate properties in classical distribution property estimation. We show that $C(S_\alpha, d, \varepsilon) = O(d^{2/\alpha}/\varepsilon^{2/\alpha})$. Since $\alpha < 1$, the EYD algorithm requires more copies than is required for tomography. We show elsewhere [44] that the EYD algorithm requires at least $\Omega(d^{1+1/\alpha}/\varepsilon^{1/\alpha})$ copies to estimate the entropy, showing that the super-quadratic dependence on d is necessary for the EYD algorithm. The upper bound is proved in Theorem 4. For comparison, in the classical setting the exponent of d is almost $1/\alpha$.

3) von Neumann Entropy, $\alpha = 1$: Again using the EYD algorithm, in Theorem 2 we show that $C(S, d, \varepsilon) = O(d^2/\varepsilon^2)$. We formulate an optimization problem whose solutions are an upper bound on the bias of the empirical estimate, and we bound the variance by proving that the estimator has a small bounded difference constant. Elsewhere [44], we show a lower bound of $\Omega(d^2/\varepsilon)$ for the EYD estimator to estimate the entropy of the maximally mixed state. This complexity is still similar to that of full quantum tomography.

4) Non-Integral $\alpha > 1$: Using the EYD algorithm, in Theorem 3, we show that $C(S_\alpha, d, \varepsilon) = O(d^2/\varepsilon^2)$. Elsewhere [44], we provide a lower bound of $\Omega(d^2/\varepsilon)$ for the EYD estimator.

III. Preliminaries

We list some of the definitions and results we use in the paper.

Lemma 2: The *total variation* distance, *KL divergence*, and χ^2 *distance* between distributions p and q over $\mathcal{X}$ satisfy

$$\begin{aligned} 2d_{TV}(p,q)^2 &= 2\left(\sup_{A\subset\mathcal{X}} (p(A) - q(A))\right)^2 = \frac{1}{2}\|p-q\|_1^2 \\ &\le d_{KL}(p,q) \\ &\le \chi^2(p,q) \overset{\text{def}}{=} \sum_{x\in\mathcal{X}} \frac{(p(x)-q(x))^2}{q(x)}. \end{aligned}$$

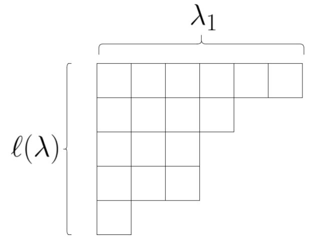

Fig. 1. English Young diagram for the partition $\boldsymbol{\lambda} = (6, 4, 3, 3, 1)$.

The inequalities follow from Pinsker's inequality and the concavity of logarithms, respectively.

A. Schur Polynomials and Power-Sum Polynomials

A *partition* $\boldsymbol{\lambda}$ of n is a collection of non-negative integers $\boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \dots$ that sum to n. We write $\boldsymbol{\lambda} \vdash n$ and we denote by Λ_n the *set of all partitions* of n. The number of positive integers in $\boldsymbol{\lambda}$ is denoted by $\ell(\boldsymbol{\lambda})$, which is called its *length*. A partition $\boldsymbol{\lambda}$ can be depicted with an English *Young diagram*, which consists of a row of λ_1 boxes above a row of λ_2 boxes, etc., as shown in Fig. 1. The partition associated with a Young diagram is called its *shape*. Note that the number of rows in the Young diagram of $\boldsymbol{\lambda}$ is $\ell(\boldsymbol{\lambda})$ and the total number of boxes is n. A *Young tableau* over alphabet $[d] = \{1, \dots, d\}$ is a Young diagram in which each box has been filled with an element of $[d]$. A Young tableau is called *standard* if it is strictly increasing left-to-right across each row and top-to-bottom down each column. A Young tableau is *semistandard* if it is strictly increasing top-to-bottom down each column and nondecreasing left-to-right across each row. Given $\boldsymbol{\lambda} \vdash n$ and d, the *Schur polynomial* is a polynomial in the variables $x_1, x_2, \dots, x_d$ defined by

$$s_{\boldsymbol{\lambda}}(x) \stackrel{\text{def}}{=} \sum_T \prod_{i=1}^{d} x_i^{\#(T,i)}, \qquad (2)$$

where the sum is over the set of all semistandard Young Tableaus T over alphabet $[d]$ corresponding to the partition $\boldsymbol{\lambda}$ and $\#(T, i)$ is the number of times i appears in T. Schur polynomials turn out to be symmetric, namely that they are invariant to the ordering of the variables $x_1, \dots, x_d$ [38], [45].

We will also use power sum polynomials. For $\alpha \in \mathbb{R}_{\geq 0}$ and a distribution $\boldsymbol{\eta}$ on $[d]$,[1] let

$$M_\alpha(\boldsymbol{\eta}) \stackrel{\text{def}}{=} \sum_{i=1}^{d} \boldsymbol{\eta}_i^\alpha.$$

Given $\boldsymbol{\lambda} \vdash r$, the *power sum* polynomial is defined as

$$M_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) = \prod_{i=1}^{\ell(\boldsymbol{\lambda})} M_{\lambda_i}(\boldsymbol{\eta}).$$

Reference [19, Lemma 1] describes several useful inequalities for the power sums of distributions.

Schur polynomials and power-sum polynomials are related through a change of basis. There exists a function $\chi_\cdot(\cdot) : \Lambda_n^2 \mapsto \mathbb{R}$ such that [45, Th. 7.17.3]

$$M_{\boldsymbol{\mu}}(\cdot) = \sum_{\boldsymbol{\lambda}} \chi_{\boldsymbol{\lambda}}(\boldsymbol{\mu}) s_{\boldsymbol{\lambda}}(\cdot). \qquad (3)$$

The $\chi_\cdot(\cdot)$ function in fact comprises the characters of the irreducible representations of the symmetric group on $[n] = \{1, \dots, n\}$ [45, Sec. 7.18], although this fact is not needed. The $\chi_{\boldsymbol{\lambda}}(\boldsymbol{\mu})$ function is defined combinatorially in [45] and is difficult to compute in general [46], although we shall only be interested in a particular $\boldsymbol{\mu}$, as follows. Let $\dim(\boldsymbol{\lambda})$ denote the number of standard Young tableaus over alphabet $[n]$ with shape $\boldsymbol{\lambda}$. For $\boldsymbol{\lambda} \vdash n$ and $\boldsymbol{\mu} \vdash r$ define

$$p^\#_{\boldsymbol{\mu}}(\boldsymbol{\lambda}) \stackrel{\text{def}}{=} \begin{cases} n^{\underline{r}} \cdot \frac{\chi_{\boldsymbol{\lambda}}(\boldsymbol{\mu} \cup 1^{n-r})}{\dim(\boldsymbol{\lambda})} & \text{if } n \geq r, \\ 0 & \text{otherwise,} \end{cases}$$

where $n^{\underline{r}}$ is the *falling power*, i.e., $n^{\underline{r}} = n \cdot (n-1) \cdot (n-2) \cdot (n-r+1)$ and $\boldsymbol{\mu} \cup 1^{n-r}$ denotes the partition of $[n]$ consisting of $\boldsymbol{\mu}$ followed by $n - r$ ones.

1) Weak Schur Sampling (WSS): We describe some of the key results about weak Schur sampling (WSS) that we will use in this paper. The readers is referred to [13, Sec. 4.2.2], [14, Ch. 3], and references therein for further details.

Weak Schur Sampling is a measurement scheme indexed by partitions of n. It takes n independent copies of a mixed state ρ (denoted $\rho^{\otimes n}$), and outputs a $\boldsymbol{\lambda} \vdash n$. The output distribution over partitions is called the Schur-Weyl distribution, denoted $SW_{\boldsymbol{\eta}}$, and the probability of $\boldsymbol{\lambda} \vdash n$ is given by

$$SW_{\boldsymbol{\eta}}(\boldsymbol{\lambda}) = \dim(\boldsymbol{\lambda}) \cdot s_{\boldsymbol{\lambda}}(\boldsymbol{\eta}), \qquad (4)$$

where, recall from the previous section that $\dim(\boldsymbol{\lambda})$ is the number of standard Young tableaus of shape $\boldsymbol{\lambda}$, and $s_{\boldsymbol{\lambda}}(\boldsymbol{\eta})$ is the Schur polynomial with variables $\boldsymbol{\eta}$, and shape $\boldsymbol{\lambda}$. Since Schur polynomials are symmetric, this probability is only a function of the multiset of eigenvalues, namely a function of the eigenvalue spectrum.

An alternate combinatorial characterization of the output of WSS is available (e.g., [47]). Suppose ρ is a mixed state with the multiset of eigenvalues $\{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_d\}$.

1) Consider a distribution over $[d]$, where i has probability $\boldsymbol{\eta}_i$.
2) Draw $X^n = X_1, \dots, X_n$ independently from this distribution.
3) Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \dots$, be such that for any $k > 0$, $\boldsymbol{\lambda}_1 + \dots + \boldsymbol{\lambda}_k$ is equal to the *largest sum of lengths* of k disjoint non-decreasing subsequences of X^n.

The output distribution of this process is the same as the distribution over partitions given by weak Schur sampling [14]. Furthermore, the output distribution of the above procedure is *independent* of the ordering of $\boldsymbol{\eta}_i$'s and only depends on the multiset of the eigenvalues [14]. We call this the *longest increasing subsequence (LIS)* interpretation of weak Schur sampling.

The $p^\#_{\boldsymbol{\mu}}(\boldsymbol{\lambda})$ polynomial defined in the last section is useful to us due to the following lemma, which states that the (normalized) polynomial $p^\#_{(r)}(\boldsymbol{\lambda})$ is an unbiased estimator of the

[1] Power sums can are usually defined for general vectors. We will consider them only for distributions in this paper.

rth moment of $\boldsymbol{\eta}$, where (r) is the partition of n that has one r, and $n - r$ one's. The lemma follows from the definitions and results already mentioned, and is implicit in [48], [49], and explicit in [14, Proposition 3.8.3].

Lemma 3: Fix a distribution $\boldsymbol{\eta}$, a natural number r, and any partition $\boldsymbol{\mu}$ of r. If $\boldsymbol{\lambda}$ is randomly generated according to the distribution in (4) then

$$\mathbb{E}\big[p^{\#}_{\boldsymbol{\mu}}(\boldsymbol{\lambda})\big] = n^{\underline{r}} \cdot M_{\boldsymbol{\mu}}(\boldsymbol{\eta}) = n^{\underline{r}} \cdot \prod_i M_{\mu_i}(\boldsymbol{\eta}). \tag{5}$$

In the special case that $\boldsymbol{\mu} = (r)$, a partition with only one part, we have

$$\mathbb{E}\Big[p^{\#}_{(r)}(\boldsymbol{\lambda})\Big] = n^{\underline{r}} \cdot M_r(\boldsymbol{\eta}). \tag{6}$$

Proof: Plugging in the probability of $\boldsymbol{\lambda}$ from (4), and the definition of $p^{\#}_{(r)}(\boldsymbol{\lambda})$ from Section III-A, and finally using (3) gives the lemma. ∎

2) The EYD Algorithm and Classical Plug-In Estimation: The EYD algorithm is a simple algorithm for estimating $f(\rho)$. The algorithm works in two steps:

1) Compute the empirical distribution, which assigns probability $\boldsymbol{\lambda}_i/n$ to the symbol i.
2) Output the property f of a mixed state with eigenvalues equal to $\boldsymbol{\lambda}_i/n$.

The EYD algorithm is a quantum analogue of the classical empirical/plug-in estimator. An observation from the non-decreasing subsequence interpretation of weak-Schur sampling is that for any sequence X^n, the distribution $\boldsymbol{\lambda}_i/n$ majorizes the corresponding empirical distribution. This follows from the fact that the length of longest k disjoint non-decreasing sub-sequences is always at least the sum of the k largest N_i's, where N_i is the number of appearances of i in X^n.

Lemma 4: Consider the sorted plug-in distribution $\hat{p}$ of X^n, and the distribution $\boldsymbol{\lambda}_i/n$ obtained from X^n by the WSS procedure. $\boldsymbol{\lambda}/n$ majorizes $\hat{p}$, namely, for all j, $\sum_{i=1}^{j} \boldsymbol{\lambda}_i/n \geq \sum_{i=1}^{j} \hat{p}(i)$.

B. Proving Upper Bounds on the Copy Complexity

Consider $\alpha \neq 1$ and $\hat{\varepsilon} \in (0, 1)$. Suppose $\widehat{M_\alpha(\boldsymbol{\eta})}$ satisfies

$$\Big|\widehat{M_\alpha(\boldsymbol{\eta})} - M_\alpha(\boldsymbol{\eta})\Big| \leq \hat{\varepsilon} M_\alpha(\boldsymbol{\eta}).$$

Then

$$\begin{aligned}
&\left|\frac{1}{1-\alpha}\log \widehat{M_\alpha(\boldsymbol{\eta})} - \mathcal{S}_\alpha(\rho)\right| \\
&= \left|\frac{1}{1-\alpha}\log\frac{\widehat{M_\alpha(\boldsymbol{\eta})}}{M_\alpha(\boldsymbol{\eta})}\right| \\
&\leq \left|\frac{1}{1-\alpha}\max\big\{\log\big(1+\hat{\varepsilon}\big), \big|\log\big(1-\hat{\varepsilon}\big)\big|\big\}\right| \\
&\leq \left|\frac{\log\big(1-\hat{\varepsilon}\big)}{1-\alpha}\right|.
\end{aligned}$$

Therefore, to obtain a ±ε estimate of $\mathcal{S}_\alpha(\boldsymbol{\eta})$, it suffices to derive a $1 - e^{-\varepsilon|1-\alpha|}$ multiplicative estimate of $M_\alpha(\boldsymbol{\eta})$. Note that $1 - e^{-\varepsilon|1-\alpha|} \geq \frac{\varepsilon|1-\alpha|}{1+\varepsilon|1-\alpha|}$ since $e^{-x} \leq \frac{1}{x+1}$ for $x > -1$. Moreover, in the regime in which ε does not grow with d, $\frac{\varepsilon|1-\alpha|}{1+\varepsilon|1-\alpha|} = \Theta(\varepsilon)$. Therefore, in the remainder of the paper, we will be interested in $1+\varepsilon$ multiplicative estimators.

Algorithm 1 Estimating Rényi Entropy for Integral α's

```
1: Input: $n$ independent copies of the state $\rho$, and $\alpha \in \mathbb{N}$
2: Run weak Schur sampling to obtain $\boldsymbol{\lambda} \vdash n$.
3: Let $(\alpha)$ be the partition of $\alpha$ with one part.
4: Compute $p^{\#}_{(\alpha)}(\boldsymbol{\lambda}) = n^{\underline{\alpha}} \cdot \frac{\chi_{\boldsymbol{\lambda}}((\alpha)\cup 1^{n-\alpha})}{\dim(\boldsymbol{\lambda})}$.
5: Output: $\frac{1}{1-\alpha}\log\left(\frac{p^{\#}_{(\alpha)}(\lambda)}{n^{\underline{\alpha}}}\right)$.
```

Finally note by Markov's inequality that for any X,

$$\Pr\Big(|X - \mathbb{E}[X]|^2 > 9 \cdot \mathrm{Var}(X)\Big) < \frac{1}{9}.$$

Since $\mathbb{E}\Big[p^{\#}_{(\alpha)}(\lambda)\Big] = n^{\underline{\alpha}} M_\alpha(\boldsymbol{\eta})$ (by Lemma 3), then we get a $1+\varepsilon$ multiplicative estimator of $M_\alpha(\boldsymbol{\eta})$ with probability at least $8/9$ if

$$\Big(\varepsilon \cdot \mathbb{E}\Big[p^{\#}_{(\alpha)}(\lambda)\Big]\Big)^2 \geq 9 \cdot \mathrm{Var}\Big(p^{\#}_{(\alpha)}(\lambda)\Big). \tag{7}$$

IV. Measuring $\mathcal{S}_\alpha(\rho)$ for Integral α

Our main result for integral $\alpha > 1$ is the following tight bound (up to constant factors) on the copy complexity of estimating $\mathcal{S}_\alpha(\rho)$.

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

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

where the constants depend only on α and is independent of d.

A. Achievability

Our estimator is simple and is described in Algorithm 1.

Note that we could remove the $n^{\underline{\alpha}}$ terms from the algorithm's description in Steps 4 and 5, but these polynomials with said factors have a number of applications in representation theory to study the symmetric group, and we keep the notations and definitions intact.

To prove the theorem, we bound the expectation and concentration of $p^{\#}_{(\alpha)}(\lambda)$.

Lemma 5: There is a constant C_α depending only on α such that

$$\mathbb{E}\Big[p^{\#}_{(\alpha)}(\boldsymbol{\lambda})\Big] = n^{\underline{\alpha}} M_\alpha(\boldsymbol{\eta}), \tag{8}$$

$$\mathrm{Var}\Big(p^{\#}_{(\alpha)}(\boldsymbol{\lambda})\Big) \leq C_\alpha \cdot n^\alpha\Big(1 + n^{\alpha-1} M_{2\alpha-1}(\boldsymbol{\eta})\Big). \tag{9}$$

1) Proof of Theorem 1 Using Lemma 5: We want (7) to hold, which happens if

$$\big(\varepsilon n^{\underline{\alpha}} M_\alpha(\boldsymbol{\eta})\big)^2 \geq 9C_\alpha \cdot n^\alpha\Big(1 + n^{\alpha-1} M_{2\alpha-1}(\boldsymbol{\eta})\Big). \tag{10}$$

We claim that $n = \Theta\Big(\max\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\}\Big)$ is sufficient for (10) to hold. Note that for a fixed α, and large enough n, $n^{\underline{\alpha}} = \Theta(n^\alpha)$ and let $\tilde{c}_\alpha > 0$ be such that $n^{\underline{\alpha}} \geq \sqrt{\tilde{c}_\alpha} n^\alpha$.

Now suppose $n \geq c_\alpha \max\{\frac{d^{1-1/\alpha}}{\varepsilon^2}, \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\}$ for some constant $c_\alpha \geq \max\{(18C_\alpha/\tilde{c}_\alpha)^{1/\alpha}, 18C_\alpha/\tilde{c}_\alpha\}$. Then

$$\begin{aligned}\left(\varepsilon n^{\underline{\alpha}} M_\alpha(\boldsymbol{\eta})\right)^2 &\geq \tilde{c}_\alpha \varepsilon^2 n^{2\alpha} M_\alpha(\boldsymbol{\eta})^2\\ &\geq \frac{\tilde{c}_\alpha}{2}\varepsilon^2 n^{2\alpha}\left(\frac{1}{d^{2\alpha-2}} + \frac{M_{2\alpha-1}(\boldsymbol{\eta})}{d^{1-1/\alpha}}\right) \quad (11)\\ &\geq \frac{\tilde{c}_\alpha}{2}\varepsilon^2 n^{2\alpha}\left(\frac{c_\alpha^\alpha}{\varepsilon^2 n^\alpha} + \frac{c_\alpha M_{2\alpha-1}(\boldsymbol{\eta})}{n\varepsilon^2}\right)\\ &= \frac{\tilde{c}_\alpha}{2} n^\alpha \left(c_\alpha^\alpha + c_\alpha n^{\alpha-1} M_{2\alpha-1}(\boldsymbol{\eta})\right)\\ &\geq 9C_\alpha \cdot n^\alpha \left(1 + n^{\alpha-1} M_{2\alpha-1}(\boldsymbol{\eta})\right), \quad (12)\end{aligned}$$

where (11) follows from the fact that $M_\alpha(\boldsymbol{\eta}) \geq d^{1-\alpha}$ and $M_{2\alpha-1}(\boldsymbol{\eta}) \leq d^{1-1/\alpha} M_\alpha(\boldsymbol{\eta})^2$ [19, Lemma 1(*i*,*v*)], and (12) follows from the assumption that $n \geq c_\alpha \frac{d^{1-1/\alpha}}{\varepsilon^2}$ and $n \geq c_\alpha \frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}$.

2) Proof of Lemma 5: Equation (8) has already been established in Lemma 3. It remains to bound the variance of the estimator.

$$\mathrm{Var}\left(p^\#_{(\alpha)}(\lambda)\right) = \mathbb{E}\left[p^\#_{(\alpha)}(\lambda)^2\right] - \mathbb{E}\left[p^\#_{(\alpha)}(\lambda)\right]^2.$$

The second term is evaluated from the means of the $p^\#_\alpha(\lambda)$ polynomials, which we know. For the first term, we need to bound the expectation of the products of such polynomials. In fact, there is a general result [49, Proposition 4.5], [14, Corollary 3.8.8] that states that for any μ_1, μ_2,

$$p^\#_{\mu_1}(\lambda) \cdot p^\#_{\mu_2}(\lambda) = p^\#_{\mu_1 \cup \mu_2}(\lambda) + \text{linear combination of } p^\#\text{'s for partitions of size at most } |\mu_1 \cup \mu_2| - 1.$$

For our particular case, both the partitions μ_1 and μ_2 are (α), and specializing [14, Corollary 3.8.8] shows that

$$p^\#_{(\alpha)}(\lambda) \cdot p^\#_{(\alpha)}(\lambda) = p^\#_{(\alpha)\cup(\alpha)}(\lambda) + \sum_{\mu \in \mathcal{S}} C_\mu p^\#_\mu(\lambda),$$

where each coefficient C_μ is at most $(\alpha!)^2 < \exp(O(\alpha \log \alpha))$, and $\mathcal{S}$ is the set of all partitions μ that can be obtained through the following procedure:

1) Let j be an integer in the set $\{0, \dots, \alpha - 1\}$.
2) Let σ_1 be a permutation over $[\alpha + j]$ that has a cycle over the elements $\{1, \dots, \alpha\}$, and all the remaining elements are fixed points (the set $\{\alpha + 1, \dots, \alpha + j\}$ for $j \geq 1$).
3) Let σ_2 be a permutation over $[\alpha + j]$ that has a cycle over the elements $\{j+1, \dots, j+\alpha\}$, and all the remaining elements are fixed points (the set $\{1, \dots, j\}$ for $j \geq 1$).
4) Let μ be the cycle structure of $\sigma_1 \circ \sigma_2$.

The set of partitions that can be obtained through the above procedure for a *fixed* $j \in \{0, \dots, \alpha - 1\}$ will be denoted by $\mathcal{S}_j$. Now consider,

$$\begin{aligned}\mathrm{Var}\left(p^\#_{(\alpha)}(\lambda)\right) &= \mathbb{E}\left[p^\#_{(\alpha)}(\lambda)^2\right] - \mathbb{E}\left[p^\#_{(\alpha)}(\lambda)\right]^2\\ &= \mathbb{E}\left[p^\#_{(\alpha,\alpha)}(\lambda) + \sum_{\mu \in \mathcal{S}} C_\mu p^\#_\mu(\lambda)\right] - \mathbb{E}\left[p^\#_{(\alpha)}(\lambda)\right]^2\\ &= n^{\underline{2\alpha}} M_{(\alpha,\alpha)}(\boldsymbol{\eta}) + \sum_{\mu \in \mathcal{S}} C_\mu n^{\underline{|\mu|}} M_\mu(\boldsymbol{\eta})\\ &\quad - \left(n^{\underline{\alpha}} M_\alpha(\boldsymbol{\eta})\right)^2\\ &= \left(n^{\underline{2\alpha}} - \left(n^{\underline{\alpha}}\right)^2\right) M_\alpha(\boldsymbol{\eta})^2 + \sum_{\mu \in \mathcal{S}} C_\mu \cdot n^{\underline{|\mu|}} M_\mu(\boldsymbol{\eta}),\end{aligned}$$

where we have used that $M_\alpha(\boldsymbol{\eta})^2 = M_{(\alpha,\alpha)}(\boldsymbol{\eta})$. To bound $M_\mu(\boldsymbol{\eta})$ for $\mu \in \mathcal{S}$, we use the following two lemmas. Lemma 6 is proved in Appendix A and Lemma 7 is proved in Appendix B. Recall that for a partition μ, $\ell(\mu)$ denotes the length of the partition.

Lemma 6: For all $j \in \{0, \dots, \alpha - 1\}$ and $\mu \in \mathcal{S}_j$, $\ell(\mu) \leq \alpha - j$.

Definition 2: Let $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ be partitions of the same integer r. Then $\boldsymbol{\mu}$ is said to majorize $\boldsymbol{\mu}'$, denoted $\boldsymbol{\mu} \trianglerighteq \boldsymbol{\mu}'$, if for all $j \geq 1$, $\sum_{i=1}^{j} \boldsymbol{\mu}_i \geq \sum_{i=1}^{j} \boldsymbol{\mu}'_i$.

Lemma 7: Let $\boldsymbol{\mu} \trianglerighteq \boldsymbol{\mu}'$. Then for any distribution $\boldsymbol{\eta}$, $M_{\boldsymbol{\mu}}(\boldsymbol{\eta}) \geq M_{\boldsymbol{\mu}'}(\boldsymbol{\eta})$.

Noting that $n^{\underline{2\alpha}} < (n^{\underline{\alpha}})^2$, we obtain

$$\begin{aligned}\mathrm{Var}\left(p^\#_{(\alpha)}(\lambda)\right) &< \sum_{\mu \in \mathcal{S}} C_\mu \cdot n^{\underline{|\mu|}} M_\mu(\boldsymbol{\eta})\\ &\leq c_\alpha \sum_{j=0}^{\alpha-1} \sum_{\mu \in \mathcal{S}_j} n^{\underline{|\mu|}} M_\mu(\boldsymbol{\eta}) \quad (13)\\ &\leq c_\alpha \sum_{j=0}^{\alpha-1} \sum_{\mu \in \mathcal{S}_j} n^{\underline{|\mu|}} M_{\alpha+j-\ell(\mu)+1}(\boldsymbol{\eta}) \quad (14)\\ &\leq c_\alpha \sum_{j=0}^{\alpha-1} \sum_{\mu \in \mathcal{S}_j} n^{\underline{\alpha+j}} M_{2j+1}(\boldsymbol{\eta}) \quad (15)\\ &\leq c_\alpha n^\alpha \sum_{j=0}^{\alpha-1} \sum_{\mu \in \mathcal{S}_j} n^j M_{2j+1}(\boldsymbol{\eta})\\ &\leq c_\alpha n^\alpha |\mathcal{S}| \max\left\{1, n^{\alpha-1} M_{2\alpha-1}(\boldsymbol{\eta})\right\}, \quad (16)\end{aligned}$$

where (13) follows from the fact that $C_\mu \leq (\alpha!)^2 := c_\alpha$, (14) follows from Lemma 7 and the fact that $[(\alpha + j - l(\mu) + 1) \cup 1^{l(\mu)-1}] \trianglerighteq \mu$, (15) follows from Lemma 6 and the fact that $M_r(\boldsymbol{\eta})$ is a non-increasing function in r for fixed $\boldsymbol{\eta}$, and (16) follows from the fact that for $j \in \{1, \dots, \alpha - 2\}$,

$$\begin{aligned}n^{j-1} M_{2j-1}(\boldsymbol{\eta}) &\leq n^j M_{2j+1}(\boldsymbol{\eta})\\ &\Rightarrow n^j M_{2j+1}(\boldsymbol{\eta}) \leq n^{j+1} M_{2j+3}(\boldsymbol{\eta}).\end{aligned}$$

Note that the above implication follows from Lemma 7 applied to the partitions $(2j + 1, 2j + 1)$ and $(2j - 1, 2j + 3)$:

$$\left(n^j M_{2j+1}(\boldsymbol{\eta})\right)^2 \leq n^{j-1} M_{2j-1}(\boldsymbol{\eta}) \cdot n^{j+1} M_{2j+3}(\boldsymbol{\eta}).$$

Finally, note that $|\mathcal{S}|$ depends only on α. Hence, the lemma follows by setting $C_\alpha = c_\alpha |\mathcal{S}|$.

B. Converse

Note that there are two terms in the copy complexity in Theorem 1. The first is $d^{1-1/\alpha}/\varepsilon^2$, which is a lower bound in the classical setting [19] and thus a lower bound in our setting.

We use Le Cam's method to lower bound the copy complexity by the second term. We define a hypothesis testing problem next.

1) Two Point Testing: Given density matrices ρ and σ with spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, respectively, and an integer n,

- Let X be a uniform random variable over $\{0, 1\}$.
- If $X = 0$, generate a Young tableau $\boldsymbol{\lambda} \vdash n \sim SW_{\boldsymbol{\eta}}$.
- If $X = 1$, generate a Young tableau $\boldsymbol{\lambda} \vdash n \sim SW_{\boldsymbol{\nu}}$.
- Given $\boldsymbol{\lambda}$, predict X with $\hat{X}$.

Let $P_e = \min_{\hat{X}} \Pr\left(\hat{X} \neq X\right)$. From basic hypothesis testing results, we can deduce that

$$P_e = \frac{1}{2} - \frac{1}{2} d_{TV}\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right).$$

We construct two spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$, such that $\mathcal{S}_\alpha(\boldsymbol{\eta}) - \mathcal{S}_\alpha(\boldsymbol{\nu}) = \Theta(\varepsilon)$, and

$$\frac{1}{2} d_{TV}\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right) < 0.05,$$

unless $n = \Omega(d^{2-2/\alpha}/\varepsilon^{2/\alpha})$. This proves that unless n is large enough, there is no classifier that can test between the spectrums $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ with probability greater than 2/3, implying our lower bound.

Note that the second term in the complexity expression of Theorem 1 dominates when $\varepsilon > 1/\sqrt{d}$. We henceforth assume in the remainder of this section that $\varepsilon > 1/\sqrt{d}$.

Consider the following two spectrums:

$$\boldsymbol{\eta} = \left(\frac{1 + (\varepsilon d)^{1/\alpha}}{d}, \frac{1 - \frac{(\varepsilon d)^{1/\alpha}}{d-1}}{d}, \ldots, \frac{1 - \frac{(\varepsilon d)^{1/\alpha}}{d-1}}{d}\right) \quad (17)$$

$$\boldsymbol{\nu} = \left(\frac{1}{d}, \ldots, \frac{1}{d}\right). \quad (18)$$

Note that for any $d > 2$, assuming that[2] $\varepsilon < \log d$, we have

$$(\varepsilon d)^{1/\alpha} < d - 1. \quad (19)$$

Thus $\boldsymbol{\eta}$ is a valid distribution. $\boldsymbol{\nu}$ is simply the maximally-mixed state.

Lemma 8: Suppose $\varepsilon > 1/\sqrt{d}$ and $d > (3\alpha)^{\frac{2\alpha}{\alpha-1}}$. Then

$$|\mathcal{S}_\alpha(\boldsymbol{\nu}) - \mathcal{S}_\alpha(\boldsymbol{\eta})| \geq \frac{1}{\alpha - 1} \log\left(1 + \frac{2\varepsilon}{3}\right).$$

Proof: Computing the moments of $\boldsymbol{\eta}$, we have

$$M_\alpha(\boldsymbol{\eta}) = \frac{1}{d^\alpha}\left(\left(1 + (\varepsilon d)^{\frac{1}{\alpha}}\right)^\alpha + (d-1)\left(1 - \frac{(\varepsilon d)^{\frac{1}{\alpha}}}{d-1}\right)^\alpha\right).$$

For $\alpha \geq 1$ and $x \geq 0$, note that $(1+x)^\alpha > 1+x^\alpha$, and if $x \leq 1$, $(1-x)^\alpha > 1 - \alpha x$. Using these two inequalities above with $x = (\varepsilon d)^{1/\alpha}$ in the first term and with $x = (\varepsilon d)^{1/\alpha}/(d-1)$ in the second term (and using (19)), we obtain

$$\begin{aligned} M_\alpha(\boldsymbol{\eta}) &= \frac{1}{d^\alpha}\left(\left(1 + (\varepsilon d)^{\frac{1}{\alpha}}\right)^\alpha + (d-1)\left(1 - \frac{(\varepsilon d)^{\frac{1}{\alpha}}}{d-1}\right)^\alpha\right) \\ &\geq \frac{1}{d^\alpha}\left(1 + \varepsilon d + (d-1)\cdot\left(1 - \frac{\alpha(\varepsilon d)^{1/\alpha}}{d-1}\right)\right) \\ &\geq \frac{1}{d^\alpha}\left(d + \varepsilon d - \alpha(\varepsilon d)^{1/\alpha}\right) \\ &\geq \frac{d}{d^\alpha}\left(1 + \frac{2}{3}\varepsilon\right) \\ &= M_\alpha(\boldsymbol{\nu}) \cdot \left(1 + \frac{2}{3}\varepsilon\right), \end{aligned}$$

whenever $d > \frac{(3\alpha)^{\frac{\alpha}{\alpha-1}}}{\varepsilon}$, which is implied by the conditions $\varepsilon > 1/\sqrt{d}$ and $d > (3\alpha)^{\frac{2\alpha}{\alpha-1}}$. ■

Lemma 9: Any algorithm that can test between $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ with probability at least 2/3 requires at least $\Omega\left(\frac{d^{2-2/\alpha}}{\varepsilon^{2/\alpha}}\right)$ copies.

Proof: We prove that $d_{TV}\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right) < 0.05$. Bounding the total variation distance is hard to handle, and therefore other distance measures are used to bound the total variation distance. By Lemma 2, we know that

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

The objective is to bound the χ^2 distance between the SW distributions for the two states with n copies. We use the following formula, derived in [14, Corollary 6.2.4]. The result in this form was obtained from related results on Schur functions [50].

Lemma 10: Let $x_1, \ldots, x_d$ be such that $\sum x_i = 0$, and $x_i \geq -1$. Let $\boldsymbol{\eta}$ be the spectrum with $\boldsymbol{\eta}_i = (1+x_i)/d$, and $\boldsymbol{\nu}$ be the spectrum of the maximally mixed state, namely $\boldsymbol{\nu}_i = 1/d$. Then,

$$\chi^2\left(SW_{\boldsymbol{\eta}}, SW_{\boldsymbol{\nu}}\right) = \sum_{\mu:1\leq\ell(\mu)\leq d} \frac{s_\mu(x)^2}{d^{\overline{\mu}} d^{|\mu|}} n^{\underline{|\mu|}},$$

where for a partition μ, $d^{\overline{\mu}}$ is defined below.

Definition 3: Let μ be a partition. Index each box in the Young tableaux for μ with an entry (i, j), where i the row number and j is the column number of the box. For each box $\square$ in the tableaux, let $c(\square) = j - i$ be the content of $\square$. Then for a real number $z \in \mathbb{R}$,

$$z^{\overline{\mu}} = \prod_{\square}(z + c(\square)).$$

We will use the following bound on these falling powers of partitions for our lower bound.

Lemma 11: Let μ be a partition such that $\ell(\mu) \leq d$, where $\ell(\mu)$ is the number of non-zero entries of μ (which is also the number of non-empty rows in the Young tableaux). Then

$$d^{\overline{\mu}} \geq \left(\frac{d}{e}\right)^{|\mu|}.$$

This result is proved in Appendix C, and we now prove our result using this lemma.

The distribution $\boldsymbol{\nu}$ corresponds to the spectrum defined in (18), and we choose the x_i's to make the spectrum $\boldsymbol{\eta}$ equal to (17). In particular, let $x_1 = (\varepsilon d)^{1/\alpha}$, and $x_i = -\frac{(\varepsilon d)^{1/\alpha}}{d-1}$ for $i = 2, \ldots, d$. Let $y_1 = 1$, and $y_i = -1/(d-1)$ for $i = 2, \ldots, d$. Then,

$$x = (\varepsilon d)^{1/\alpha} \cdot \left(1, \frac{-1}{d-1}, \ldots, \frac{-1}{d-1}\right) = (\varepsilon d)^{1/\alpha} \cdot y.$$

Recall that the Schur polynomial $s_\mu(x)$ is a homogeneous symmetric polynomial of degree $|\mu|$. This implies,

$$s_\mu(x) = (\varepsilon d)^{\frac{|\mu|}{\alpha}} s_\mu(y). \quad (20)$$

[2] If $\varepsilon \geq \log d$, then $\hat{S}_\alpha = 0$ is a valid estimate and the problem becomes trivial.

Let y_+ be the vector of absolute values of y, namely

$$y_+ = \left(1, \frac{1}{d-1}, \dots, \frac{1}{d-1}\right).$$

Then, $\left|s_\mu(y)\right| \le \left|s_\mu(y_+)\right|$. Using the fact that $d^{\overline{\mu}} \ge (d/e)^{|\mu|}$ and $n^{\underline{m}} \le n^m$,

$$\begin{aligned}\chi^2\left(SW_{\boldsymbol\eta}, SW_{\boldsymbol\nu}\right) &= \sum_{\mu:1\le\ell(\mu)\le d} \frac{s_\mu(x)^2}{d^{\overline{|\mu|}} d^{|\mu|}} n^{\underline{|\mu|}} \\ &= \sum_{\substack{\mu:1\le\ell(\mu)\le d\\ 1<|\mu|\le n}} \frac{s_\mu(x)^2}{d^{\overline{|\mu|}} d^{|\mu|}} n^{\underline{|\mu|}} \\ &\le \sum_{\substack{\mu:1\le\ell(\mu)\le d\\ 1<|\mu|\le n}} s_\mu(y_+)^2 \cdot \left(\frac{n(\varepsilon d)^{2/\alpha}}{(d/e)\cdot d}\right)^{|\mu|} \\ &= \sum_{\substack{\mu:1\le\ell(\mu)\le d\\ 1<|\mu|\le n}} s_\mu(y_+)^2 \cdot \left(\frac{e\cdot n\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^{|\mu|} \\ &\le \sum_{m=2}^{n} \left(\left(\frac{en\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^m \cdot \left(\sum_{\mu:|\mu|=m} s_\mu(y_+)^2\right)\right),\end{aligned}$$

where the second equality follows from the fact that $n^{\underline{|\mu|}} = 0$ for $|\mu| > n$ and the fact that the term with $|\mu| = 1$ vanishes since $\sum_{i=1}^d x_i = 0$. Let $p(m)$ denote the partition number of m, i.e., the number of unordered partitions of m. Bounds on the growth of partition numbers are well established [39]. We only need the following loose bound that holds for all m

$$p(m) < e^{3\sqrt{m}}.$$

This gives

$$\begin{aligned}&\chi^2\left(SW_{\boldsymbol\eta}, SW_{\boldsymbol\nu}\right) \\ &\quad\le \sum_{m=2}^{n}\left(e^{3\sqrt{m}}\left(\frac{en\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^m \cdot \max_{\mu:|\mu|=m} s_\mu(y_+)^2\right). \qquad (21)\end{aligned}$$

The entries of y_+ have the following structure. The first entry is 1, and all other entries are $1/(d-1)$. This allows us to use the "branching rule" of Schur polynomials. The general form can be found in [38, eq. 5.10]. A special case appears in the following form in [51, eq. 1.4].

Lemma 12: The Schur polynomial $s_\mu(z)$ can be decomposed as:

$$s_\mu(z) = \sum_{\lambda\prec\mu} (z_1)^{|\mu|-|\lambda|} s_\lambda\left(z_2^d\right), \qquad (22)$$

where z_2^d denotes the second through last components of z and the summation is over all partitions λ such that $\mu_1 \ge \lambda_1 \ge \mu_2 \ge \lambda_2 \ge \mu_3 \ge \dots$.

Applying this with $z = y_+$,

$$s_\mu(y_+) = \sum_{\lambda\prec\mu}\left(\frac{1}{d-1}\right)^{|\lambda|} s_\lambda\left(1^{d-1}\right). \qquad (23)$$

From (2), we see that $s_\lambda\left(1^{d-1}\right)$ is the number of semistandard Young tableaux with shape λ and entries from $[d-1]$. We can trivially bound $s_\lambda\left(1^{d-1}\right) \le (d-1)^{|\lambda|}$, the total number of ways of filling the Young tableaux with entries from $[d-1]$, *without any regard to ordering*.

We need one final definition.

Definition 4: For a partition μ, let $\text{prec}(\mu)$ be the number of partitions λ such that $\lambda \prec \mu$, where $\prec$ is as defined in Lemma 12.

Lemma 13: Let $|\mu| = m$. Then

$$\text{prec}(\mu) = \prod_{i=1}^{\infty}(\mu_i - \mu_{i+1} + 1) < m^{\sqrt{2m}}.$$

Proof: The equality is due to a simple counting argument. For the inequality, let $\mu_{i_1} > \mu_{i_2} > \dots > \mu_{i_k} \ge 1$ be the distinct elements in μ. If $k = 1$, the inequality is easy to show, so assume that $k > 1$. Then, $k(k+1)/2 \le \mu_{i_1} + \dots + \mu_{i_k} \le m$, implying that $k < \sqrt{2m}$. Moreover, $\mu_{i_1} - \mu_{i_k} \le |\mu| - 1 = m - 1$ since $\mu_{i_k} > 1$. Then

$$\text{prec}(\mu) \le \prod_{j=1}^{k}\left(1 + \mu_{i_j} - \mu_{i_{j+1}}\right) \le m^k < m^{\sqrt{2m}}.$$

■

Therefore,

$$\begin{aligned}s_\mu(y_+) &= \sum_{\lambda\prec\mu}\left(\frac{1}{d-1}\right)^{|\lambda|} s_\lambda\left(1^{d-1}\right) \\ &\le \sum_{\lambda\prec\mu}\left(\frac{1}{d-1}\right)^{|\lambda|}(d-1)^{|\lambda|} \\ &= \text{prec}(\mu) \le |\mu|^{\sqrt{2|\mu|}}. \qquad (24)\end{aligned}$$

Plugging (24) in (21),

$$\begin{aligned}\chi^2\left(SW_{\boldsymbol\eta}, SW_{\boldsymbol\nu}\right) &\le \sum_{m=2}^{n}\left(e^{3\sqrt{m}}\left(\frac{en\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^m \cdot \max_{\mu:|\mu|=m} s_\mu(y_+)^2\right) \\ &\le \sum_{m=2}^{n}\left(e^{3\sqrt{m}}\left(\frac{en\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^m \cdot m^{2\sqrt{2m}}\right) \\ &\le \sum_{m=2}^{n}\left((em)^{3\sqrt{m}}\left(\frac{en\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^m\right).\end{aligned}$$

Finally note that $m^{\sqrt{m}} < 2\cdot 2^m$ for all $m > 1$. Therefore,

$$\chi^2\left(SW_{\boldsymbol\eta}, SW_{\boldsymbol\nu}\right) \le \sum_{m=2}^{n} 8\left(\frac{(2e)^4 n\varepsilon^{2/\alpha}}{d^{2-2/\alpha}}\right)^m.$$

Therefore, unless $n \ge \Omega\left(\frac{d^{2-\frac{2}{\alpha}}}{\varepsilon^{\frac{2}{\alpha}}}\right)$, the χ^2 distance is small, proving the result. ■

V. von Neumann Entropy

A. Empirical Entropy Upper Bound

Analogous to the classical setting, the empirical distribution is $\widehat{\boldsymbol\eta}_i \overset{\text{def}}{=} \frac{\boldsymbol\lambda_i}{n}$ and the empirical estimate of $S(\rho)$ is

$$\widehat{S(\boldsymbol\lambda)} \overset{\text{def}}{=} \sum_{i=1}^{d}\frac{\boldsymbol\lambda_i}{n}\log\frac{n}{\boldsymbol\lambda_i} = \sum_{i=1}^{d}\widehat{\boldsymbol\eta}_i\log\frac{1}{\widehat{\boldsymbol\eta}_i}.$$

We prove the following bound on the mean squared error of this estimator.

Theorem 2: The empirical entropy estimate satisfies:

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

An immediate corollary is the following sample complexity bound.

Corollary 1:

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

Proof: By Markov's inequality on Theorem 2, there is a constant C such that with probability at least 0.9,

$$\begin{aligned}\left|\left(\widehat{S(\rho)} - S(\rho)\right)\right| &< C\sqrt{\frac{d^4}{n^2} + \frac{d^2}{n} + \frac{\log^2 n}{n}} \\ &< C\left(\frac{d^2}{n} + \frac{d}{\sqrt{n}} + \frac{\log n}{\sqrt{n}}\right).\end{aligned}$$

Bounding each term to at most $\varepsilon/3C$ gives the sample complexity bound. ■

Proof of Theorem 2: The mean-squared error of an estimator $\hat{X}$ of a parameter x can be decomposed as

$$\mathbb{E}\left[\left(x - \hat{X}\right)^2\right] = \mathbb{E}\left[\left(x - \mathbb{E}\left[\hat{X}\right]\right)^2\right] + \mathbb{E}\left[\left(\hat{X} - \mathbb{E}\left[\hat{X}\right]\right)^2\right],$$

where the first term is the squared bias, and the second term is the variance. In particular,

$$\begin{aligned}&\mathbb{E}\left[\left(\widehat{S(\boldsymbol{\lambda})} - S(\rho)\right)^2\right] \\ &= \left(S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right)^2 + \mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right). \qquad (25)\end{aligned}$$

The theorem follows by plugging the following two bounds on the bias and variance, respectively, into (25).

Lemma 14:

$$\left|S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right| \le \frac{d^2}{n} + 9\frac{d}{\sqrt{n}}.$$

Lemma 15:

$$\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) = O\left(\frac{\log^2 n}{n}\right).$$

■

1) Bounding the Bias (Proof of Lemma 14): The bias of the empirical estimate can be bounded as:

$$\begin{aligned}\left|S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right| &= \left|\mathbb{E}\left[\sum_{i=1}^{d}\left(\boldsymbol{\eta}_i \log\frac{1}{\boldsymbol{\eta}_i} - \widehat{\boldsymbol{\eta}}_i \log\frac{1}{\widehat{\boldsymbol{\eta}}_i}\right)\right]\right| \\ &\le \left|\mathbb{E}\left[\sum_{i=1}^{d}\left(\boldsymbol{\eta}_i - \widehat{\boldsymbol{\eta}}_i\right)\log\frac{1}{\boldsymbol{\eta}_i}\right]\right| \\ &\quad + \left|\mathbb{E}\left[\sum_{i=1}^{d}\left(\widehat{\boldsymbol{\eta}}_i \log\frac{\widehat{\boldsymbol{\eta}}_i}{\boldsymbol{\eta}_i}\right)\right]\right| \qquad (26) \\ &\le \left|\sum_{i=1}^{d}\left(\boldsymbol{\eta}_i - \mathbb{E}\left[\widehat{\boldsymbol{\eta}}_i\right]\right)\log\frac{1}{\boldsymbol{\eta}_i}\right| \\ &\quad + \mathbb{E}\left[\sum_{i=1}^{d}\frac{\left(\widehat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i\right)^2}{\boldsymbol{\eta}_i}\right], \qquad (27)\end{aligned}$$

where (26) is by the triangle inequality, and (27) is from Lemma (2). The second term is the expected χ^2-distance of the empirical distance and the underlying distribution. Reference [23, Th. 4.7] states that

$$\mathbb{E}\left[\sum_{i=1}^{d}\frac{\left(\widehat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i\right)^2}{\boldsymbol{\eta}_i}\right] \le \frac{d^2}{n},$$

which bounds the second term of (27). To bound the first term, we again use the following result from [23] that bounds the expected value of $\widehat{\boldsymbol{\eta}}_i$ around $\boldsymbol{\eta}_i$.

Lemma 16 [23, Th. 1.4]:

$$\left|\boldsymbol{\eta}_i - \mathbb{E}\left[\left(\widehat{\boldsymbol{\eta}}_i\right)\right]\right| \le 2\sqrt{\frac{\min\{1, \boldsymbol{\eta}_i d\}}{n}}.$$

Let $c_1, \ldots, c_d$ be the constants such that $\boldsymbol{\eta}_i - \mathbb{E}\left[\widehat{\boldsymbol{\eta}}_i\right] = c_i\sqrt{\frac{d\boldsymbol{\eta}_i}{n}}$, then by Lemma 16, $|c_i| \le 2$. Since $\sum_{i=1}^{d}\boldsymbol{\eta}_i = \sum_{i=1}^{d}\widehat{\boldsymbol{\eta}}_i = 1$,

$$\sqrt{\frac{d}{n}}\left(\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\right) = \sum_{i=1}^{d}\left(\boldsymbol{\eta}_i - \mathbb{E}\left[\widehat{\boldsymbol{\eta}}_i\right]\right) = 0,$$

implying that $\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i} = 0$. Therefore,

$$\sum_{i=1}^{d}\left(\boldsymbol{\eta}_i - \mathbb{E}\left[\widehat{\boldsymbol{\eta}}_i\right]\right)\log\frac{1}{\boldsymbol{\eta}_i} = \sqrt{\frac{d}{n}}\left(\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\log\frac{1}{\boldsymbol{\eta}_i}\right). \qquad (28)$$

Since $\sqrt{\frac{d}{n}}$ is a constant, to bound the first term of (27) it will suffice to upper bound the following maximization problem:

$$\textbf{P1:} \quad \text{maximize} \quad \left|\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\log\frac{1}{\boldsymbol{\eta}_i}\right|$$

$$\text{subject to} \quad |c_i| \le 2, \ \text{and} \ \sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i} = 0.$$

By the triangle inequality,

$$\left|\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\log\frac{1}{\boldsymbol{\eta}_i}\right| \le \left|\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\log\frac{1}{c_i^2\boldsymbol{\eta}_i}\right| + \left|\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\log c_i^2\right| \qquad (29)$$

We bound the terms individually. We first consider the second term. Since $|c_i| \le 2$, the largest value of $|c_i \log c_i^2|$ is $2\log 4$. Therefore,

$$\left|\sum_{i=1}^{d} c_i\sqrt{\boldsymbol{\eta}_i}\log c_i^2\right| \le 2\log 4 \cdot \left(\sum_{i=1}^{d}\sqrt{\boldsymbol{\eta}_i}\right) \le (2\log 4)\cdot\sqrt{d},$$

where we use that $\sum_{i=1}^{d}\sqrt{\boldsymbol{\eta}_i} < \sqrt{d}$ by concavity of square root.

Let $x_i = c_i\sqrt{\boldsymbol{\eta}_i}$, then $\sum_i x_i = 0$, and since $\sum \boldsymbol{\eta}_i = 1$, $\sum_i x_i^2 \le 4$. Therefore, to bound the first term of (29), it will suffice to solve **P2** below.

$$\textbf{P2:} \quad \text{maximize} \ \sum_{i=1}^{d} x_i \log\frac{1}{x_i^2} \qquad (30)$$

$$\text{subject to } \sum_{i=1}^{d} x_i = 0, \text{ and } \sum_{i=1}^{d} x_i^2 \leq 4. \tag{31}$$

We show in Appendix D that

Lemma 17: The maximum value of the optimization problem **P2** is at most $\frac{16}{e}\sqrt{d}$.

Plugging this in (29), the maximum of **P1** is at most $(16/e + 2\log 4)\sqrt{d}$. Therefore,

$$\left|\sum_{i=1}^{d}(\boldsymbol{\eta}_i - \mathbb{E}[\widehat{\boldsymbol{\eta}_i}])\log\frac{1}{\boldsymbol{\eta}_i}\right| \leq \left(\frac{16}{e} + 2\log 4\right)\frac{d}{\sqrt{n}} \leq \frac{9d}{\sqrt{n}}.$$

Plugging this in turn into (27) yields

$$\left|S(\rho) - \mathbb{E}\left[\widehat{S(\boldsymbol{\lambda})}\right]\right| \leq \frac{d^2}{n} + \frac{9d}{\sqrt{n}},$$

thus bounding the bias.

2) Proof of Lemma 15: We will use the bounded difference variance bound [52, Corollary 3.2]. In particular, we consider the LIS interpretation of weak Schur sampling. Let $X^n \in [d]^n$, and let $\boldsymbol{\lambda}$ be the shape of its young tableaux through the LIS interpretation. Let $\boldsymbol{\lambda}'$ be the shape of the Young tableaux corresponding to a sequence with Hamming distance at most one from X^n. Let $S(\boldsymbol{\lambda})$, and $S(\boldsymbol{\lambda}')$ denote their respective empirical von Neumann entropy. The next lemma, proved in Appendix E, states that changing one of the n symbols has a *small effect* on the empirical entropy.

Lemma 18: Let $\boldsymbol{\lambda}$, and $\boldsymbol{\lambda}'$ be two Young tableaux shapes obtained from the LIS of two length-n samples that differ in at most one symbol. If $n \geq 27$, then

$$\left|\widehat{S(\boldsymbol{\lambda})} - \widehat{S(\boldsymbol{\lambda}')}\right| \leq \frac{50\log n}{n}.$$

We invoke the bounded difference inequality [52, Corollary 3.2] along with Lemma 18. The empirical entropy estimate changes by at most $50\log n/n$ when one symbol is changed. Therefore, the variance is at most

$$\mathrm{Var}\left(\widehat{S(\boldsymbol{\lambda})}\right) \leq \frac{1}{4}n \cdot \left(\frac{50\log n}{n}\right)^2 \leq \frac{625\log^2 n}{n}.$$

VI. Non-Integral α

A. $\alpha > 1$

We prove the following sample complexity bound for estimating $\mathcal{S}_\alpha(\rho)$ for $\alpha > 1$. Throughout this section, we let $\boldsymbol{\eta}$ denote the true (sorted) spectrum and $\boldsymbol{\lambda}/n$ denote its empirical estimate.

Theorem 3: For $\alpha > 1$, the empirical estimator of $\mathcal{S}_\alpha(\rho)$ outputs a $\pm\varepsilon$ estimate with $O\left(\frac{d^2}{\varepsilon^2}\right)$ copies of ρ with probability at least 0.9.

Proof: Recall that $n = \sum \boldsymbol{\lambda}_i$. Define

$$M_\alpha(\boldsymbol{\lambda}) \stackrel{\text{def}}{=} \sum_{i=1}^{d}\left(\frac{\boldsymbol{\lambda}_i}{n}\right)^\alpha.$$

We show that for large enough n, $M_\alpha(\boldsymbol{\lambda})$ is within a small multiplicative factor of $M_\alpha(\boldsymbol{\eta})$. The following result, proved in [44, Appendix G], shows each term $(\boldsymbol{\lambda}_i/n)^\alpha$ concentrates around $\boldsymbol{\eta}_i^\alpha$.

Lemma 19: Let $\beta > 1$. Then there is a constant C_β such that

$$\mathbb{E}\left[\left|\boldsymbol{\lambda}_i{}^\beta - (\boldsymbol{\eta}_i n)^\beta\right|\right] < C_\beta \cdot \left(n^{\beta/2} + \sqrt{n}(\boldsymbol{\eta}_i n)^{\beta-1}\right).$$

Then,

$$\begin{aligned}\mathbb{E}[|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})|] &= \mathbb{E}\left[\left|\sum_{i=1}^{d}\left(\left(\frac{\boldsymbol{\lambda}_i}{n}\right)^\alpha - \boldsymbol{\eta}_i^\alpha\right)\right|\right] \\ &\leq \frac{1}{n^\alpha}\sum_{i=1}^{d}\mathbb{E}\left[\left|\boldsymbol{\lambda}_i{}^\alpha - (\boldsymbol{\eta}_i n)^\alpha\right|\right] \\ &\leq \frac{C_\alpha}{n^\alpha}\sum_{i=1}^{d}\left(n^{\alpha/2} + \sqrt{n}(\eta_i n)^{\alpha-1}\right) \quad (32) \\ &= C_\alpha\left(\frac{d}{n^{\alpha/2}} + \frac{M_{\alpha-1}(\boldsymbol{\eta})}{\sqrt{n}}\right), \quad (33)\end{aligned}$$

where (32) uses Lemma 19.

By Lemma [19, Lemma 1], for $\alpha > 1$, $M_\alpha(\boldsymbol{\eta}) \geq d^{1-\alpha}$, and $M_{\alpha-1}(\boldsymbol{\eta}) \leq dM_\alpha(\boldsymbol{\eta})$. Substituting in (33),

$$\begin{aligned}\mathbb{E}[|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})|] &\leq C_\alpha\left(\frac{d}{n^{\alpha/2}} + \frac{M_{\alpha-1}(\boldsymbol{\eta})}{\sqrt{n}}\right) \\ &\leq C_\alpha\left(\frac{d^\alpha M_\alpha(\boldsymbol{\eta})}{n^{\alpha/2}} + \frac{dM_\alpha(\boldsymbol{\eta})}{\sqrt{n}}\right) \\ &\leq C_\alpha\left(\frac{d^\alpha}{n^{\alpha/2}} + \frac{d}{\sqrt{n}}\right)M_\alpha(\boldsymbol{\eta}).\end{aligned}$$

By Markov's Inequality,

$$\begin{aligned}&\Pr\left(|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})| > \varepsilon M_\alpha(\boldsymbol{\eta})\right) \\ &\leq \frac{\mathbb{E}[|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})|]}{\varepsilon M_\alpha(\boldsymbol{\eta})} \leq \frac{C_\alpha}{\varepsilon}\left(\frac{d^\alpha}{n^{\alpha/2}} + \frac{d}{\sqrt{n}}\right).\end{aligned}$$

When $n > Cd^2\left(\frac{1}{\varepsilon^2} + \frac{1}{\varepsilon^{2/\alpha}}\right)$, the result follows. Since $\alpha > 1$, the first term dominates. ∎

B. $\alpha < 1$

In this section, we will prove the following:

Theorem 4: The empirical estimator of $\mathcal{S}_\alpha(\rho)$ is a $\pm\varepsilon$ estimate with $O\left((d/\varepsilon)^{2/\alpha}\right)$ copies.

Similar to the case of large α, we need the following result, which is proved in [44, Appendix H].

Lemma 20: Let $\beta < 1$. Then there is a constant C_β such that

$$\mathbb{E}\left[\left|\boldsymbol{\lambda}_i{}^\beta - (\eta_i n)^\beta\right|\right] < C_\beta \cdot n^{\beta/2}.$$

We now prove the copy complexity bound assuming this result.

Proof of Theorem 4: Recall that $n = \sum \boldsymbol{\lambda}_i$. Define,

$$M_\alpha(\boldsymbol{\lambda}) \stackrel{\text{def}}{=} \sum_{i=1}^{d}\left(\frac{\boldsymbol{\lambda}_i}{n}\right)^\alpha.$$

Then by the triangle inequality,

$$\mathbb{E}[|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})|] \leq \frac{1}{n^\alpha}\sum_{i=1}^{d}\mathbb{E}\left[\left|\boldsymbol{\lambda}_i{}^\alpha - (\boldsymbol{\eta}_i n)^\alpha\right|\right]$$

$$\leq \frac{C_\alpha}{n^\alpha} \sum_{i=1}^{d} n^{\alpha/2} \tag{34}$$

$$= C_\alpha \frac{d}{n^{\alpha/2}}, \tag{35}$$

where (34) follows from Lemma 20. For $\alpha < 1$, $M_\alpha(\boldsymbol{\eta}) \geq 1$. Substituting in (35),

$$\mathbb{E}[|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})|] \leq C_\alpha \frac{d}{n^{\alpha/2}} \leq C_\alpha \frac{d}{n^{\alpha/2}} M_\alpha(\boldsymbol{\eta}).$$

By Markov's Inequality,

$$\begin{aligned} &\Pr(|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})| > \varepsilon M_\alpha(\boldsymbol{\eta})) \\ &\leq \frac{\mathbb{E}[|M_\alpha(\boldsymbol{\lambda}) - M_\alpha(\boldsymbol{\eta})|]}{\varepsilon M_\alpha(\boldsymbol{\eta})} \leq \frac{C_\alpha}{\varepsilon}\left(\frac{d}{n^{\alpha/2}}\right). \end{aligned}$$

Therefore, when $n > C\left(\frac{d}{\varepsilon}\right)^{2/\alpha}$, the result follows. ∎

Appendix A
Proof of Lemma 6

Fix $j \in \{0, \ldots, \alpha - 1\}$. First we prove that each cycle in $\sigma = \sigma_1 \circ \sigma_2$ contains an element in $\{j+1, \ldots, \alpha\}$. Let

$$S = \{j+1, \ldots, \alpha\}, \quad F_1 = \{\alpha+1, \ldots, \alpha+j\}, \\ \text{and } F_2 = \{1, \ldots, j\}.$$

F_1 is fixed under σ_1, and F_2 is fixed under σ_2. Now consider any cycle in σ, and pick an element k in the cycle. If $k \in S$, then the claim is true. Otherwise:

Case 1: $k \in F_2$.

Let n_k be the largest integer such that $\sigma_1(k), \sigma_1^2(k), \ldots, \sigma_1^{n_k}(k) \in F_2$. (If $\sigma_1(k) \notin F_2$, define $n_k = 0$.) Note that, since σ_1 performs a cycle on $F_2 \cup S$, there must exist m such that $\sigma_1^m(k) \in S$. Hence, n_k is finite. Then,

$$\begin{aligned} \sigma^{n_k+1}(k) &= (\sigma_1 \circ \sigma_2)^{n_k+1}(k) \overset{\text{(a)}}{=} (\sigma_1 \circ \sigma_2)^{n_k}(\sigma_1(k)) \\ &\overset{\text{(b)}}{=} \sigma_1^{n_k+1}(k) \overset{\text{(c)}}{\in} S, \end{aligned}$$

where (a) follows from the definition of n_k and the fact that points in F_2 are fixed under σ_2, (b) follows similarly (by induction), and (c) follows from the definition of σ_1 and n_k.

Case 2: $k \in F_1$.

Let n_k be the largest integer such that $\sigma_2(k), \sigma_2^2(k), \ldots, \sigma_2^{n_k}(k) \in F_1$. (If $\sigma_2(k) \notin F_1$, define $n_k = 0$.) Note that, since σ_2 performs a cycle on $F_1 \cup S$, there must exist m such that $\sigma_2^m(k) \in S$. Hence, n_k is finite. Then,

$$\begin{aligned} \sigma^{n_k+1}(k) = (\sigma_1 \circ \sigma_2)^{n_k+1}(k) &\overset{\text{(a)}}{=} (\sigma_1 \circ \sigma_2)^{n_k}(\sigma_2(k)) \\ &\overset{\text{(b)}}{=} \sigma_1 \circ \sigma_2^{n_k+1}(k), \end{aligned}$$

where (a) follows from the definition of n_k and the fact that points in F_1 are fixed under σ_1, and (b) follows similarly (by induction). Now, by definition of σ_2 and n_k, $\sigma_2^{n_k+1}(k) \in S$. Then, by definition of σ_1, $\sigma_1 \circ \sigma_2^{n_k+1}(k) \in F_2 \cup S$. If $\sigma_1 \circ \sigma_2^{n_k+1}(k) \in S$, then the claim is true. Finally, if $\sigma_1 \circ \sigma_2^{n_k+1}(k) \in F_2$, this falls back to case 1 which has been resolved.

Since $|\{j+1, \ldots, \alpha\}| = \alpha - j$ it follows that $\ell(\mu) \leq \alpha - j$.

Appendix B
Proof of Lemma 7

Let $\ell = \ell(\mu_1) = \ell(\mu_2)$, $\mu_1 = (x_1, \ldots, x_\ell)$, and $\mu_2 = (y_1, \ldots, y_\ell)$. Then

$$\begin{aligned} M_{\mu_1}(\boldsymbol{\eta}) &= \prod_{i=1}^{\ell} M_{x_i}(\boldsymbol{\eta}) = \prod_{i=1}^{\ell} \sum_{j=1}^{d} \eta_j^{x_i} \\ &= \sum_{j_1, \ldots, j_\ell \in [d]^\ell} \eta_{j_1}^{x_1} \ldots \eta_{j_\ell}^{x_\ell}. \end{aligned}$$

We define an equivalence relation on $[d]^\ell$ as follows: $(j_1, \ldots, j_\ell) \sim (\hat{j}_1, \ldots, \hat{j}_\ell)$ if there exists a permutation σ on $[\ell]$ such that $\sigma(j_1, \ldots, j_\ell) = (\hat{j}_1, \ldots, \hat{j}_\ell)$. We denote by $\mathcal{E}$ the set of equivalence classes created by this relation, and for each $E \in \mathcal{E}$ we pick a representative element and denote it by $(j_1, \ldots, j_\ell)_E$. For each E, define $g_E{:}E \to [\ell!]$ as

$$\begin{aligned} g_E(j_1 \ldots j_\ell) = \left|\left\{\sigma : \sigma\left((j_1 \ldots j_\ell)_E\right) = (j_1 \ldots j_\ell)\right\}\right|, \\ (j_1 \ldots j_\ell) \in E. \end{aligned}$$

Now note that, for each E, $g_E(.)$ is a constant function. Indeed, if $(j_1, \ldots, j_\ell)$ and $(\hat{j}_1, \ldots, \hat{j}_\ell)$ belong to E, then there exists σ_1 such that $\sigma_1(j_1, \ldots, j_\ell) = (\hat{j}_1, \ldots, \hat{j}_\ell)$. Therefore if $\sigma((j_1, \ldots, j_\ell)_E) = (j_1, \ldots, j_\ell)$, then $\sigma_1 \circ \sigma((j_1, \ldots, j_\ell)_E) = (\hat{j}_1, \ldots, j_\ell)$. Similarly, if $\sigma((j_1, \ldots, j_\ell)_E) = (\hat{j}_1, \ldots, \hat{j}_\ell)$, then $\sigma_1^{(-1)} \circ \sigma((j_1, \ldots, j_\ell)_E) = (j_1, \ldots, j_\ell)$. So define $g{:}\mathcal{E} \to [\ell!]$ as $g(E) = g_E((j_1, \ldots, j_\ell)_E)$. Now,

$$\begin{aligned} M_{\mu_1}(\boldsymbol{\eta}) &= \sum_{j_1, \ldots, j_\ell \in [d]^\ell} \eta_{j_1}^{x_1} \ldots \eta_{j_\ell}^{x_\ell} \\ &= \sum_{E \in \mathcal{E}} \sum_{(j_1, \ldots, j_\ell) \in E} \eta_{j_1}^{x_1} \ldots \eta_{j_\ell}^{x_\ell} \\ &= \sum_{E \in \mathcal{E}} \frac{1}{g(E)} \sum_{\sigma} \eta_{\sigma(j_{1,E})}^{x_1} \ldots \eta_{\sigma(j_{\ell,E})}^{x_\ell} \\ &\geq \sum_{E \in \mathcal{E}} \frac{1}{g(E)} \sum_{\sigma} \eta_{\sigma(j_{1,E})}^{y_1} \ldots \eta_{\sigma(j_{\ell,E})}^{y_\ell} = M_{\mu_2}(\boldsymbol{\eta}), \end{aligned}$$

where the inequality follows from Muirhead's theorem [53], [54, p. 125].

Appendix C
Proof of Lemma 11

Let $|\mu| = qd + r$, where $0 < r < d$, and q are non-negative integers. We will show that of all μ with $|\mu| = qd + r$ and $\ell(\mu) \leq d$, the tableaux with q columns with d boxes, and one last column with r boxes, minimizes $d^{\overline{\mu}}$. Toward this end consider a tableaux with at least two non-empty columns that have less than d boxes in them. Then we can move a box from the last row with length μ_1 to the end of the first column that does not have length equal to d. This operation moves a box to the left and below, thereby decreasing the value of $c(\square)$.

We now assume that the partition μ has q columns with d boxes and one column with r boxes. For this partition μ, by Definition 3,

$$d^{\overline{\mu}} = (d+q)^{\underline{r}} \prod_{j=0}^{q-1} (d+j)^{\underline{d}} \geq (d!)^q \cdot (d)^{\underline{r}}. \tag{36}$$

We will show that for any integer $0 \le t \le d$,

$$d^{\underline{t}} \ge \left(\frac{d}{e}\right)^t. \tag{37}$$

Plugging this bound in (36), and noting that $d! = d^{\underline{d}}$, we obtain,

$$d^{\overline{\mu}} \ge \left(\left(\frac{d}{e}\right)^d\right)^q \cdot \left(\frac{d}{e}\right)^r = \left(\frac{d}{e}\right)^{|\mu|} \tag{38}$$

We now prove (37). Let

$$f(t) = \frac{d^{\underline{t}}}{\left(\frac{d}{e}\right)^t}.$$

Then for any $t \le d$,

$$\frac{f(t+1)}{f(t)} = \frac{(d-t)}{(d/e)},$$

and this ratio is monotonically decreasing with t. Therefore, the smallest value of $f(t)$ occurs at either $t = 0$ or $t = d$. At $t = 0$, (37) is true since both sides are 1, and at $t = d$, we need to show that $d! > (d/e)^d$, which follows from Stirling's approximation.

Appendix D
Proof of Lemma 17

We will first show that at the maxima, there can be at most three distinct values that the x_i's can take, of which at most one is positive.

Consider $x_i > 0$ and $x_j > 0$. Then, by the concavity of logarithm, if we replace both by $(x_i+x_j)/2$ the objective value increases. The constraints, on the other hand, are still valid.

We now consider the negative values. Writing the Lagrangian,

$$\begin{aligned}\mathcal{L}(x_1, \dots, x_d, \gamma_1, \gamma_2) = \sum_{i=1}^{d} \left(x_i \log \frac{1}{x_i^2}\right) + \gamma_1 \left(4 - \sum_i x_i^2\right) \\ + \gamma_2 \left(\sum_i x_i\right).\end{aligned}$$

Differentiating with respect to x_i,

$$\frac{\partial \mathcal{L}(x_1, \dots, x_d, \gamma_1, \gamma_2)}{\partial x_i} = \log \frac{1}{x_i^2} - 2 - 2\gamma_1 x_i + \gamma_2 = 0.$$

The function $-2\log(-x) - 2\gamma_1 x - 2 + \gamma_2$ is strictly convex on $(-\infty, 0)$, and therefore has at most two roots.

Therefore, there are at most three distinct values that x_i's can take, and at most one of them is positive. Let $y_1 > 0 > -y_2 > -y_3$ be these values, and let d_1, d_2, d_3 be the multiplicities of these. Therefore, the optimization problem can be written as:

$$\begin{aligned}\mathbf{P3}: \text{maximize} \quad & d_1 y_1 \log \frac{1}{y_1^2} - d_2 y_2 \log \frac{1}{y_2^2} - d_3 y_3 \log \frac{1}{y_3^2} \\ \text{subject to} \quad & d_1 y_1 - d_2 y_2 - d_3 y_3 = 0, \\ & d_1 y_1^2 + d_2 y_2^2 + d_3 y_3^2 \le 4, \text{ and} \\ & d_1 + d_2 + d_3 \le d.\end{aligned}$$

Substituting $d_1 y_1 = d_2 y_2 + d_3 y_3$ the objective becomes

$$\begin{aligned}(d_2 y_2 + d_3 y_3) \log \frac{1}{y_1^2} - d_2 y_2 \log \frac{1}{y_2^2} - d_3 y_3 \log \frac{1}{y_3^2} \\ = d_2 y_2 \log \frac{y_2^2}{y_1^2} + d_3 y_3 \log \frac{y_3^2}{y_1^2}.\end{aligned}$$

Since $d_1 y_1 \ge d_2 y_2$, we have $y_2/y_1 \le d_1/d_2$, and

$$\begin{aligned}d_2 y_2 \log \frac{y_2^2}{y_1^2} &\le 2 d_2 y_2 \log \frac{d_1}{d_2} \\ &= 2\sqrt{d_1} \cdot \left(\sqrt{\frac{d_2}{d_1}} \log \frac{d_1}{d_2}\right) \cdot \left(\sqrt{d_2} y_2\right).\end{aligned}$$

Since $d_2 y_2^2 \le 4$, $\sqrt{d_2} y_2 < 2$. Moreover, for any $z > 0$,

$$z \log \frac{1}{z^2} = 2z \log \frac{1}{z} \le \frac{2}{e}.$$

This shows that

$$d_2 y_2 \log \frac{y_2^2}{y_1^2} \le \frac{8}{e}\sqrt{d_1} \le \frac{8}{e}\sqrt{d}.$$

By a similar argument,

$$d_3 y_3 \log \frac{y_3^2}{y_1^2} \le \frac{8}{e}\sqrt{d}.$$

Summing up the two terms bounds the objective of **P3**, and plugging in (28), we obtain

$$\left|\sum_{i=1}^{d} \left(\boldsymbol{\eta}_i - \mathbb{E}\left[\widehat{\boldsymbol{\eta}}_i\right]\right) \log \frac{1}{\boldsymbol{\eta}_i}\right| \le \frac{16}{e}\sqrt{d} \cdot \sqrt{\frac{d}{n}} = \frac{16}{e} \frac{d}{\sqrt{n}}.$$

Appendix E
Proof of Lemma 18

Proof: In the classical setting, changing one element can change at most two probabilities of the empirical distribution, using which one can bound the variance of the empirical entropy estimator. However, in our case, changing one symbol can change the length of more than one row of the Young tableaux. Reference [23, Prop. 2.2] showed that the cumulative row sums are bounded (see also [55]). In particular, *for any* $j = 1, \dots, d$,

$$\left|\sum_{i=1}^{j} \boldsymbol{\lambda}_i - \sum_{i=1}^{j} \boldsymbol{\lambda}'_i\right| \le 1.$$

Suppose $\Delta_i \overset{\text{def}}{=} \boldsymbol{\lambda}_i' - \boldsymbol{\lambda}_i$, then for all $j = 1, \dots, d$,

$$-1 \le \sum_{i=1}^{j} \Delta_i \le 1. \tag{39}$$

This also implies that for each i, $-2 \le \Delta_i \le 2$. This proves a bounded difference condition on $\boldsymbol{\lambda}_i$, which can be used to prove its concentration using McDiarmid's inequality [52, Th. 6.2].

Note that $\boldsymbol{\lambda}_i$ changes by at most two when one of the inputs changes, and hence $c = 2$. This gives

$$\Pr\left(|\boldsymbol{\lambda}_i - \mathbb{E}[\boldsymbol{\lambda}_i]| > t\right) \le 2 \cdot e^{-\frac{t^2}{2n}}. \tag{40}$$

Without loss of generality assume that $\ell(\boldsymbol{\lambda}) \geq \ell(\boldsymbol{\lambda}')$, i.e., the number of rows in $\boldsymbol{\lambda}$ is at least the number of rows in $\boldsymbol{\lambda}'$. By the Taylor series, for any $-x \leq \delta \leq x$,

$$(x+\delta)\log(x+\delta) = x\log x + \delta(1+\log x) + \sum_{j=2}^{\infty} \frac{\delta^j}{(j-1)j}\frac{(-1)^j}{x^{j-1}}.$$

Let $f(x) = x\log x$. Then

$$\begin{aligned}
&\widehat{S(\boldsymbol{\lambda}')} - \widehat{S(\boldsymbol{\lambda})}\\
&= \sum_{i=1}^{\ell(\boldsymbol{\lambda})} \frac{\boldsymbol{\lambda}_i'}{n}\log\frac{n}{\boldsymbol{\lambda}_i'} - \frac{\boldsymbol{\lambda}_i}{n}\log\frac{n}{\boldsymbol{\lambda}_i}\\
&= \sum_{i=1}^{\ell(\boldsymbol{\lambda})} -f\left(\frac{\boldsymbol{\lambda}_i+\Delta_i}{n}\right) + f\left(\frac{\boldsymbol{\lambda}_i}{n}\right)\\
&= \sum_{\boldsymbol{\lambda}_i>1} -\left[\frac{\Delta_i}{n}\left(1+\log\frac{\boldsymbol{\lambda}_i}{n}\right) + \frac{1}{n}\sum_{j=2}^{\infty}\left(\frac{\Delta_i}{\boldsymbol{\lambda}_i}\right)^{j-1}\frac{\Delta_i(-1)^j}{(j-1)j}\right]\\
&\quad + \sum_{\boldsymbol{\lambda}_i=1}\left[\frac{1+\Delta_i}{n}\log\frac{n}{1+\Delta_i} - \frac{1}{n}\log n\right].
\end{aligned}$$

We now consider the terms separately, and prove the following series of (simple) claims.

1) If (39) holds, and λ_i's are non-increasing, then

$$\begin{aligned}
&\left|\sum_{\boldsymbol{\lambda}_i>1}\left[\frac{\Delta_i}{n}\left(1+\log\frac{\boldsymbol{\lambda}_i}{n}\right)\right]\right|\\
&\leq \left|\sum_{\boldsymbol{\lambda}_i>1}\frac{\Delta_i}{n}\right| + \left|\sum_{\boldsymbol{\lambda}_i>1}\frac{\Delta_i}{n}\log\frac{\boldsymbol{\lambda}_i}{n}\right|\\
&\leq \frac{1}{n} + \frac{2}{n}\log\frac{n}{2}.
\end{aligned}$$

Proof: The first term is a direct consequence of (39). The second term follows from:

Lemma 21: Let $x_1 \geq x_2 \geq \cdots \geq x_m \geq 0$ be nonnegative numbers. Further, let $\Delta_1, \ldots, \Delta_m$ satisfy $-1 \leq \sum_{i=1}^{j}\Delta_i \leq 1$ for all j. Then $|\sum_{i=1}^{m}\Delta_i x_i| \leq x_1$. ■

Proof: By induction over m. The case $m=1$ is immediate. For general m, the hypotheses in the statement imply that

$$-x_1 - \Delta_1 x_1 \leq \sum_{i=2}^{m}\Delta_i x_1 \leq x_1 - \Delta_1 x_1,$$

and the induction hypothesis implies that the lower and upper bounds on the sum are negative and positive, respectively. Thus since $0 \leq x_2 \leq x_1$, we have

$$-x_1 - \Delta_1 x_1 \leq \sum_{i=2}^{m}\Delta_i x_2 \leq x_1 - \Delta_1 x_1,$$

or, equivalently,

$$-x_1 - \Delta_1 x_1 - \Delta_2 x_2 \leq \sum_{i=3}^{m}\Delta_i x_2 \leq x_1 - \Delta_1 x_1 - \Delta_2 x_2.$$

Repeating this argument $m-2$ times yields the conclusion. ■

2) For $\lambda \in \mathbb{N}$, let $I_\lambda = \{i{:}\boldsymbol{\lambda}_i = \lambda\}$ be the set of rows with length λ. Then,

$$|\{i \in I_\lambda : \Delta_i \neq 0\}| \leq 4, \tag{41}$$

i.e., there are at most four non-zero Δ_i's for each distinct value of $\boldsymbol{\lambda}_i$.

Proof: Let $\boldsymbol{\lambda}_i = \lambda$ for $i \in \{h_1, \ldots, h_2\}$. However, since the $\boldsymbol{\lambda}_i'$'s are non-increasing, Δ_i's are non-increasing for all $i \in \{h_1, \ldots, h_2\}$. If more than four of these are non-zero, then there are at least three consecutive positive, or three consecutive negative Δ_i's. However, this would violate (39). ■

3) For $\boldsymbol{\lambda}_i \geq 2$,

$$\left|\sum_{j=2}^{\infty}\left(\frac{\Delta_i}{\boldsymbol{\lambda}_i}\right)^{j-1}\frac{\Delta_i(-1)^j}{(j-1)j}\right| \leq \frac{\Delta_i^2}{\boldsymbol{\lambda}_i}.$$

Proof: Using $|\Delta_i| \leq 2$, and $\boldsymbol{\lambda}_i \geq 2$, we have $|\Delta_i|/\boldsymbol{\lambda}_i \leq 1$. This implies that for any $j \geq 1$, $(|\Delta_i|/\boldsymbol{\lambda}_i)^j \leq |\Delta_i|/\boldsymbol{\lambda}_i$. This gives

$$\begin{aligned}
&\left|\sum_{j=2}^{\infty}\left(\frac{\Delta_i}{\boldsymbol{\lambda}_i}\right)^{j-1}\frac{\Delta_i(-1)^j}{(j-1)j}\right|\\
&\leq \sum_{j=2}^{\infty}\left(\frac{|\Delta_i|}{\boldsymbol{\lambda}_i}\right)^{j-1}\frac{|\Delta_i|}{(j-1)j}\\
&\leq \left(\frac{|\Delta_i|}{\boldsymbol{\lambda}_i}\right)\cdot\sum_{j=2}^{\infty}\frac{|\Delta_i|}{(j-1)j} = \frac{\Delta_i^2}{\boldsymbol{\lambda}_i} \leq \frac{4}{\boldsymbol{\lambda}_i}.
\end{aligned}$$

Therefore, the second term in the first summand can be bounded as follows:

$$\begin{aligned}
&\sum_{\boldsymbol{\lambda}_i>1} -\frac{1}{n}\left(\sum_{j=2}^{\infty}\left(\frac{\Delta_i}{\boldsymbol{\lambda}_i}\right)^{j-1}\frac{\Delta_i(-1)^j}{(j-1)j}\right)\\
&\leq \frac{1}{n}\sum_{\boldsymbol{\lambda}_i>1}\frac{\Delta_i^2}{\boldsymbol{\lambda}_i} \leq \frac{32\log n}{n},
\end{aligned}$$

where we used (41), $|\Delta_i| \leq 2$, and that for $n \geq 3$, $\sum_{1\leq i\leq n}\frac{1}{i} \leq 2\log n$. ■

4) The second summation satisfies

$$\left|\sum_{\boldsymbol{\lambda}_i=1}\left[\frac{1+\Delta_i}{n}\log\frac{n}{1+\Delta_i} - \frac{1}{n}\log n\right]\right| \leq \frac{8\log n}{n}$$

whenever $n \geq 27$.

Proof:

$$\begin{aligned}
&\left|\sum_{\boldsymbol{\lambda}_i=1}\left[\frac{1+\Delta_i}{n}\log\frac{n}{1+\Delta_i} - \frac{1}{n}\log n\right]\right|\\
&= \left|\sum_{\boldsymbol{\lambda}_i=1,\,\Delta_i\neq 0}\left[\frac{1+\Delta_i}{n}\log\frac{n}{1+\Delta_i} - \frac{1}{n}\log n\right]\right|
\end{aligned}$$

$$\leq \sum_{\boldsymbol{\lambda}_i=1,\,\Delta_i\neq 0}\left|\left[\frac{1+\Delta_i}{n}\log\frac{n}{1+\Delta_i}-\frac{1}{n}\log n\right]\right|$$

$$\leq 4\left|\left[\frac{3}{n}\log\frac{n}{3}-\frac{1}{n}\log n\right]\right|\leq\frac{8\log n}{n},$$

where the middle inequality holds whenever $n \geq 27$.
Using these five simple claims, we can bound the difference between $\widehat{S(\boldsymbol{\lambda})}$ and $S(\boldsymbol{\lambda}')$. ■

Acknowledgment

The authors would like to thank John Wright for sharing his results on von Neumann entropy estimation and for detailed comments on a manuscript of the paper. In particular, he pointed out a mistake in the proof of the copy complexity lower bound for integral α, which has been fixed in this version.

References

[1] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states," *Phys. Rev. Lett.*, vol. 69, pp. 2881–2884, Nov. 1992.

[2] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels," *Phys. Rev. Lett.*, vol. 70, pp. 1895–1899, Mar. 1993.

[3] M. H. Hsieh and M. M. Wilde, "Entanglement-assisted communication of classical and quantum information," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4682–4704, Sep. 2010.

[4] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. Heidelberg, Germany: Springer, 2017.

[5] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémiai Kiadó, 1981.

[7] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, no. 4, pp. 2738–2747, 1995.

[8] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.*, vol. 41, no. 12, pp. 2343–2349, 1994.

[9] H.-K. Lo, "Quantum coding theorem for mixed states," *Optics Commun.*, vol. 119, nos. 5–6, pp. 552–556, 1995.

[10] J. Cardy, *Measuring Quantum Entanglement*, Max Born Lecture, Univ. Oxford, Oxford, U.K., 2012.

[11] H. Haeffner *et al.*, "Scalable multi-particle entanglement of trapped ions," *Nature*, vol. 438, pp. 643–646, Dec. 2005.

[12] X.-S. Ma *et al.*, "Quantum teleportation over 143 kilometres using active feed-forward," *Nature*, vol. 489, no. 7415, pp. 269–273, 2012.

[13] A. Montanaro and R. de Wolf, "A survey of quantum property testing," *Theory Comput. Graduate Surveys*, vol. 7, pp. 1–81, Jul. 2016.

[14] J. Wright, "How to learn a quantum state," Ph.D. dissertation, Comput. Sci. Dept., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2016.

[15] 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 *Proc. Symp. Theory Comput.*, 2011, pp. 1–10.

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

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

[18] J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi, "The complexity of estimating Rényi entropy," in *Proc. ACM-SIAM Symp. Discr. Algorithms*, 2015, pp. 1–15.

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

[20] M. Obremski and M. Skorski, "Rényi entropy estimation revisited," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Heidelberg, Germany: Springer, 2017, pp. 1–15.

[21] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5628–5641, Sep. 2017.

[22] R. O'Donnell and J. Wright, "Efficient quantum tomography," in *Proc. ACM Symp. Theory Comput.*, 2016, pp. 899–912.

[23] R. O'Donnell and J. Wright, "Efficient quantum tomography II," in *Proc. Symp. Theory Comput.*, 2017, pp. 962–974.

[24] M. Hayashi, *A Group Theoretic Approach to Quantum Information*. Cham, Switzerland: Springer, 2016.

[25] R. O'Donnell and J. Wright, "Quantum spectrum testing," in *Proc. Symp. Theory Comput.*, 2015, pp. 529–538.

[26] C. Bădescu, R. O'Donnell, and J. Wright, "Quantum state certification," *arXiv preprint arXiv:1708.06002*, 2017.

[27] S. Pallister, A. Montanaro, and N. Linden. (2017). *Optimal Verification of Entangled States With Local Measurements*. [Online]. Available: https://arxiv.org/abs/1709.03353

[28] N. Yu. (2019). *Efficient Independence Testing for Quantum States*. [Online]. Available: arXiv preprint arXiv:1904.03218

[29] M. Bavarian, S. Mehraban, and J. Wright, private communication, 2016.

[30] S. Bravyi, A. W. Harrow, and A. Hassidim, "Quantum algorithms for testing properties of distributions," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3971–3981, Jun. 2011.

[31] S. Chakraborty, E. Fischer, A. Matsliah, and R. De Wolf. (2010). *New Results on Quantum Property Testing*. [Online]. Available: https://arxiv.org/abs/1005.0523

[32] I. S. Sardharwalla, S. Strelchuk, and R. Jozsa, "Quantum conditional query complexity," *Quantum Inf. Comput.*, vol. 17, nos. 7–8, pp. 541–567, 2017.

[33] T. Li and X. Wu. (2017). *Quantum Query Complexity of Entropy Estimation*. [Online]. Available: https://arxiv.org/abs/arXiv:1710.06025

[34] S. Subramanian and M.-H. Hsieh. (2019). *Quantum Algorithm for Estimating Rényi Entropies of Quantum States*. [Online]. Available: https://arxiv.org/abs/1908.05251

[35] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, "Learning the quantum algorithm for state overlap," *New J. Phys.*, vol. 20, no. 11, 2018, Art. no. 113022.

[36] S. Johri, D. S. Steiger, and M. Troyer, "Entanglement spectroscopy on a quantum computer," *Phys. Rev. B, Condens. Matter*, vol. 96, no. 19, 2017, Art. no. 195136.

[37] Y. Subaşı, L. Cincio, and P. J. Coles, "Entanglement spectroscopy with a depth-two quantum circuit," *J. Phys. A Math. Theor.*, vol. 52, no. 4, 2019, Art. no. 044001.

[38] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*. Oxford, U.K.: Oxford Univ. Press, 1998.

[39] G. H. Hardy and S. Ramanujan, "Asymptotic formulæ in combinatory analysis," *Proc. London Math. Soc.*, vol. 2, no. 1, pp. 75–115, 1918.

[40] R. Alicki, S. Rudnicki, and S. Sadowski, "Symmetry properties of product states for the system of *N n*-level atoms," *J. Math. Phys.*, vol. 29, no. 5, pp. 1158–1162, 1988.

[41] M. Keyl and R. F. Werner, "Estimating the spectrum of a density operator," *Phys. Rev. A*, vol. 64, no. 5, 2001, Art. no. 052311.

[42] M. Hayashi and K. Matsumoto, "Estimating the spectrum of a density operator," *Phys. Rev. A*, vol. 66, no. 2, 2002, Art. no. 022311.

[43] M. Christandl and G. Mitchison, "The spectra of quantum states and the Kronecker coefficients of the symmetric group," *Commun. Math. Phys.*, vol. 261, no. 3, pp. 789–797, 2006.

[44] J. Acharya, I. Issa, N. V. Shende, and A. B. Wagner. (2017). *Measuring Quantum Entropy*. [Online]. Available: https://arxiv.org/abs/1711.00814

[45] R. P. Stanley, *Enumerative Combinatorics*, vol. 2. New York, NY, USA: Cambridge Univ. Press, 1999.

[46] C. T. Hepler, "On the complexity of computing characters of finite groups," M.S. thesis, Dept. Comput. Sci., Univ. Calgary, Calgary, AB, Canada, 1994.

[47] R. O'Donnell and J. Wright, "Guest column: A primer on the statistics of longest increasing subsequences and quantum states (shortened version)," *SIGACT News*, vol. 48, no. 3, pp. 37–59, Sep. 2017.

[48] P.-L. Méliot. (2010). *Kerov's Central Limit Theorem for Schur–Weyl Measures of Parameter 1/2*. [Online]. Available: https://arxiv.org/abs/1009.4034

[49] V. Ivanov and G. Olshanski, "Kerov's central limit theorem for the Plancherel measure on Young diagrams," in *Symmetric Functions 2001: Surveys of Developments and Perspectives*, vol. 74. Dordrecht, The Netherlands: Springer, 2002, pp. 93–151.

[50] A. Okounkov and G. Olshanski, "Shifted Schur functions," *St. Petersburg Math. J.* vol. 9, pp. 239–300, May 1996.

[51] A. Lascoux and S. O. Warnaar, “Branching rules for symmetric functions and sln basic hypergeometric series,” *Adv. Appl. Math.*, vol. 46, nos. 1–4, pp. 424–456, 2011.

[52] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, U.K.: Oxford Univ. Press, 2013.

[53] R. F. Muirhead, “Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters,” *Proc. Edinburgh Math. Soc.*, vol. 21, pp. 144–162, Feb. 1902.

[54] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, vol. 143, 2nd ed. New York, NY, USA: Springer, 2011.

[55] N. Bhatnagar and N. Linial, “On the Lipschitz constant of the RSK correspondence,” *J. Combinatorial Theory A*, vol. 119, no. 1, pp. 63–82, 2012.

Jayadev Acharya (Member, IEEE) received the Bachelor of Technology degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, in 2007, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, San Diego, in 2009 and 2014, respectively. He is an Assistant Professor with the School of Electrical and Computer Engineering, Cornell University. He was a Postdoctoral Associate in electrical engineering and computer science with MIT from 2014 to 2016.

Ibrahim Issa (Member, IEEE) received the Ph.D. degree from the School of Electrical Engineering, Cornell University, Ithaca, NY, USA, in 2017, where his thesis on information leakage earned the Outstanding ECE Ph.D. Thesis Research Award. He joined the Electrical and Computer Engineering Department, American University of Beirut in January 2019, as an Assistant Professor. He was with the Laboratory for Information in Networked Systems, Swiss Federal Institute of Technology, Lausanne, as a Postdoctoral Researcher from August 2017 to December 2018. His research interests include privacy and security, information theory, machine learning, and quantum information theory.

Nirmal V. Shende received the B.Tech. degree in electronics and communication engineering from the Visvesvaraya National Institute of Technology, Nagpur, India, the M.E. degree in telecommunication engineering from the Indian Institute of Science, Bengaluru, India, and the Ph.D. degree from the School of Electrical and Computer Engineering, Cornell University, Ithaca, USA, where he was a recipient of the Irwin M. and Joan K. Jacobs fellowship. He is currently with Marvell Semiconductor, Santa Clara, CA, USA. His research interests are in areas of information theory, quantum information theory, and wireless communication.

Aaron B. Wagner (Fellow, IEEE) received the B.S. degree in electrical engineering from the University of Michigan, Ann Arbor, in 1999, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 2002 and 2005, respectively.

From 2005 to 2006 academic year, he was a Postdoctoral Research Associate with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign and a Visiting Assistant Professor with the School of Electrical and Computer Engineering, Cornell University. Since 2006, he has been with the School of Electrical and Computer Engineering, Cornell University, where he is currently a Professor. He has received the NSF CAREER Award, the David J. Sakrison Memorial Prize from the U.C. Berkeley EECS Department, the Bernard Friedman Memorial Prize in Applied Mathematics from the U.C. Berkeley Department of Mathematics, the James L. Massey Research and Teaching Award for Young Scholars from the IEEE Information Theory Society, and teaching awards at the department, college, and university level at Cornell University. He was a Distinguished Lecturer for the IEEE Information Theory Society from 2018 to 2019.