Limit Distribution for Quantum Relative Entropy

Sreejith Sreekumar and Mario Berta
Institute for Quantum Information, RWTH Aachen University, Germany.
Email: {sreekumar, berta}@physik.rwth-aachen.de

***Abstract*—Estimation of quantum relative entropy is a fundamental statistical task in quantum information theory, physics, and beyond. While several estimators of the same have been proposed in the literature along with their computational complexities explored, a limit distribution theory which characterizes the asymptotic fluctuations of the estimation error is still premature. As our main contribution, we characterize these asymptotic distributions in terms of Fréchet derivatives of elementary operator-valued functions. We achieve this by leveraging an operator version of Taylor's theorem and identifying the regularity conditions needed. As an application of our results, we consider an estimator of quantum relative entropy based on Pauli tomography of quantum states and show that the resulting asymptotic distribution is a centered normal, with its variance characterized in terms of the Pauli operators and states. We utilize the knowledge of the aforementioned limit distribution to obtain asymptotic performance guarantees for a multi-hypothesis testing problem.**

I. Introduction

Estimation of a quantum state, also known as quantum state tomography, is an important problem in quantum information theory, physics, and quantum machine learning, see e.g., [1]–[8]. In several applications, however, the quantity of interest may not be the entire state, but only a functional of it. Quantum divergences such as quantum relative entropy [9] and its Rényi generalizations [10]–[14] form an important class of such functionals. They play a central role in quantum information theory both in terms of characterizing fundamental limits as well as applications, e.g., see the books [15]–[17]. For instance, the quantum relative entropy characterizes the error-exponent in asymmetric binary quantum hypothesis testing [18] and the Petz-Rényi divergence quantifies the exponent in quantum Chernoff bounds [19], [20]. Owing to their significance, several estimators of these measures have been proposed recently in the literature and their performance investigated in terms of benchmarks such as copy and query complexity (see *Related work* section below). However, a limit distribution theory which characterizes the asymptotic distribution of estimation error is largely unexplored.

Here, we seek a limit distribution theory for quantum relative entropy [9] between quantum states ρ and σ, defined as

$$\mathsf{D}\left(\rho\|\sigma\right) := \begin{cases} \mathrm{Tr}\left[\rho\big(\log\rho - \log\sigma\big)\right], & \text{if } \rho \ll \sigma, \\ \infty, & \text{otherwise.} \end{cases} \quad (1)$$

Given estimators ρ_n and σ_n of ρ and σ, respectively, we want to identify the scaling rate r_n (or convergence rate r_n^{-1}) and the limiting variable Z such that the following convergence in distribution (weak convergence) holds:

$$r_n\big(\mathsf{D}\left(\rho_n\|\sigma_n\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} Z.$$

Of interest is also the scenario where only one state, say ρ or σ, is estimated and the other is known. Characterization of such limit distributions have several potential applications in quantum statistics and machine learning such as constructing confidence intervals for quantum hypothesis testing, asymptotic analysis of quantum algorithms, and quantum statistics (see [21]–[23] for some classical applications).

While limit distributions fully quantify the asymptotic performance, deriving such results for estimators of quantum relative entropy is challenging on account of two reasons. Firstly, limit distributions need not always exist, as is well-known for relative entropy in the classical setting. Secondly, the non-commutative framework of quantum theory makes the analysis more involved. For tackling the first challenge, we use an operator version of Taylor's expansion with remainder and ascertain primitive conditions for the existence of limits. The technical core of our contribution entails determining conditions that allow interchange of limiting operations on trace functionals of Fréchet derivatives that appear in such an expansion. For handling issues arising due to non-commutativity, we use appropriate integral expressions for operator functions.

Applying the aforementioned method to quantum relative entropy, we establish the following convergence in distribution (Theorem 1) when $r_n(\rho_n-\rho) \xrightarrow{w} L_1$ and $r_n(\sigma_n-\sigma) \xrightarrow{w} L_2$ for $\rho \neq \sigma$ under appropriate regularity conditions:

$$r_n\big(\mathsf{D}\left(\rho_n\|\sigma_n\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} \mathrm{Tr}\left[L_1(\log\rho - \log\sigma) - \rho D[\log\sigma](L_2)\right].$$

Here[1], $D[f(A)](B)$ denotes the first-order Fréchet derivative (see Definition 1 below) of an operator-valued function f at operator A in the direction of operator B, and L_1 (resp. L_2) denotes the weak limit of the estimator ρ_n (resp. σ_n), appropriately centered and scaled. Analogous to the classical case, a faster convergence rate is achieved when $\rho = \sigma$ with the limit characterized in terms of second-order derivatives.

As an application of our limit distribution results, we characterize the asymptotic distribution of an estimator of quantum relative entropy based on Pauli tomography of quantum states ρ, σ. Specifically, we show that

$$\sqrt{n}\big(\mathsf{D}\left(\hat{\rho}_n\|\hat{\sigma}_n\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} W,$$

[1] Throughout, we consider logarithms to the base e.

where W is a centered Gaussian variable with a variance that depends on the states and Pauli operators. We then use this result to obtain performance guarantees for a multi-hypothesis testing problem for determining the quantum relative entropy between an unknown state ρ and a known state σ. Assuming that identical copies of the unknown state are available for measurement, we first perform tomography to obtain an estimate of ρ and then use the knowledge of the Gaussian limit to design a test statistic (decision rule) that achieves any desired error level for appropriately chosen thresholds. The test statistic achieves the same performance even when the number of hypotheses scales at a sufficiently slow rate with the number of measurements. Such tests have potential applications to auditing of quantum differential privacy [24], as considered in [22], [25] for the classical case.

A. Related Work

Statistical analysis of estimators of classical information measures and divergences has been an active area of research over the past few decades. The relevant literature pertains broadly to showing consistency, quantifying convergence rates of estimators (or equivalently sample complexity), and characterizing their limiting distributions. Consistency and/or convergence rates for various estimators of f-divergences, which subsumes entropy and mutual information as special cases, have been studied, see e.g., [26]–[33]. Limit distributions for several f-divergence estimators such as those based on kernel density estimates, k-nearest neighbour methods, and plug-in methods have been established recently, e.g., [22], [25], [29], [32], [34], while corresponding results for Rényi divergences have been studied in [23]. Limit distribution theory has also been explored extensively in the optimal transport literature for the class of Wasserstein distances.

In the quantum setting, computational complexities of various estimators of quantum information measures have been investigated under different input models, see e.g., [35]–[42]. Specifically, [36] established copy complexity bounds characterizing the optimal dimension dependence for quantum Rényi entropy estimation when independent copies of the state are available for measurement. [37], [38] considered entropy estimation under a quantum query model, which assumes access to an oracle that prepares the input quantum state. For limit distributional results in the quantum setting, the asymptotic distribution for spectrum estimation of a quantum state based on the empirical Young's diagram (EYD) algorithm [43], [44] has been established in [45], [46]. Local asymptotic normality results for quantum experiments based on a family of density operators indexed by a parameter have also been studied, see e.g., [47], [48]. However, to the best of our knowledge, a limit distribution theory for quantum relative entropy has not been explored before.

II. Preliminaries

A. Notation

Throughout, $\mathbb{H}$ is a separable complex Hilbert space, which can be identified with $\ell^2(\mathbb{N})$, the space of absolute square-summable infinite sequences with complex elements. $\mathcal{L}(\mathbb{H})$ denotes the set of linear operators (henceforth, referred to simply as operators) from $\mathbb{H}$ to $\mathbb{H}$. $\mathrm{Tr}\,[\cdot]$ and $\|\cdot\|_1$ signifies the trace functional and the trace-norm (Schatten 1-norm), respectively. $\mathcal{H}$ and $\mathcal{P}$ stands for the set of Hermitian and positive semi-definite operators acting on $\mathbb{H}$, respectively, while $\mathcal{H}_1$ denotes the subset of $\mathcal{H}$ with finite trace-norm. $\mathcal{S}$ designates the set of density operators (or quantum states), i.e., the set of elements of $\mathcal{P}$ with unit trace, and $\mathcal{S}^+$ denotes its subset with strictly positive eigenvalues. The notation $\leq$ represents the Löwner partial order in the context of operators, i.e., for $A, B \in \mathcal{H}$, $A \leq B$ means that $B - A \in \mathcal{P}$. $\mathbb{1}_{\mathcal{X}}$ denotes indicator of a set $\mathcal{X}$ and I denotes the identity operator on $\mathbb{H}$. For operators A, B, $A \ll B$ signifies that the support of A is contained in that of B. A^{-1} stands for the generalized (Moore–Penrose) inverse of an operator A. When $\mathbb{H}$ is finite dimensional, we use a subscript d to indicate the dimension wherever relevant, e.g., $\mathcal{H}_d$ for the set of Hermitian operators.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a sufficiently rich probability space on which all random elements are defined. A sequence of random elements $(X_n)_{n\in\mathbb{N}}$ taking values in a topological space $\mathfrak{S}$ converges weakly to a random element X (in $\mathfrak{S}$) if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all bounded continuous functions $f : \mathfrak{S} \to \mathbb{R}$. This is denoted by $X_n \xrightarrow{w} X$. A random density operator is a Borel-measurable mapping from Ω to $\mathcal{S}$.

B. Fréchet differentiability

The notion of Fréchet differentiability of an operator-valued function on the space of Hermitian operators with bounded trace-norm will play an essential role in our results below. Let $\mathsf{spec}(A)$ denote the spectrum of an operator A.

Definition 1 (Fréchet differentiability, see e.g. [49]) *For an open set $\mathcal{X} \subseteq \mathbb{R}$, let $f : \mathcal{X} \to \bar{\mathbb{R}}$ and $A \in \mathcal{H}_1$ with $\mathsf{spec}(A) \subseteq \mathcal{X}$. Then, f is called (Fréchet) differentiable at A if there exists a linear map $D[f(A)] : \mathcal{H} \to \mathcal{H}$ such that for all $H \in \mathcal{H}_1$ such that $\mathsf{spec}(A + H) \subseteq \mathcal{X}$,*

$$\|f(A + H) - f(A) - D[f(A)](H)\| = o(\|H\|_1).$$

Then, $D[f(A)]$ is called the (Fréchet) derivative of f at A and $D[f(A)](H)$ is the directional derivative of f at A in the direction H. The derivative of f induces a map from $\mathcal{H}$ into $\mathcal{L}(\mathcal{H})$ given by $D[f] : A \to D[f(A)]$. If this map is also differentiable at A, then f is said to be twice differentiable at A with the corresponding second-order derivative given by a bilinear map $D^2[f(A)] : \mathcal{H} \times \mathcal{H} \to \mathcal{H}$.

III. Main Results

Let ρ_n and σ_n be random density operators such that $\rho_n \xrightarrow{w} \rho$ and $\sigma_n \xrightarrow{w} \sigma$ in *trace norm*. Further, let $(r_n)_{n\in\mathbb{N}}$ denote a diverging positive sequence. In the following, *null* and *alternative* refers to the scenarios $\rho = \sigma$ and $\rho \neq \sigma$, respectively, while *two-sample* signifies that both ρ and σ are estimated. The following result provides sufficient conditions under which limit distribution for quantum relative entropy exists.

Theorem 1 (Limit distribution for quantum relative entropy) *Let $\rho_n \ll \sigma_n \ll \sigma$ and $\rho_n \ll \rho \ll \sigma$ be such that $\mathsf{D}(\rho_n\|\sigma_n) < \infty$, $\mathsf{D}(\rho\|\sigma) < \infty$, and there exists a constant c satisfying $\mathbb{P}\big(\|\rho_n\sigma_n^{-1}\|_\infty > c\big) \to 0$. Then, the following hold:*

(i) (Two-sample alternative) If $\big(r_n(\rho_n-\rho), r_n(\sigma_n-\sigma)\big) \xrightarrow{w} (L_1, L_2)$ in trace norm, then

$$r_n\big(\mathsf{D}(\rho_n\|\sigma_n) - \mathsf{D}(\rho\|\sigma)\big) \xrightarrow{w} \mathrm{Tr}\,[L_1(\log\rho - \log\sigma) - \rho D[\log\sigma](L_2)]. \quad (2)$$

(ii) (Two-sample null) If $\rho = \sigma$ and $\big(r_n(\rho_n - \rho), r_n(\sigma_n - \rho)\big) \xrightarrow{w} (L_1, L_2)$ in trace norm, then

$$r_n^2\mathsf{D}(\rho_n\|\sigma_n) \xrightarrow{w} \mathrm{Tr}\left[\frac{\rho}{2}D^2[\log\rho](L_1 - L_2, L_1 - L_2)\right] + \mathrm{Tr}\,[L_1 D[\log\rho](L_1 - L_2)]. \quad (3)$$

Due to space limitations, we omit the proof of Theorem 1 (and other results below) and refer the reader to an extended version [50]. The key idea in the proof relies on applying an operator version of Taylor's theorem to the function, $(x, y) \mapsto x(\log x - \log y)$, and showing that the remainder terms (e.g., second and higher order terms in the alternative case) vanish under the conditions stated in the theorem. At a technical level, the arguments use uniform Bochner-integrability of the remainder terms (guaranteed under the assumptions) to justify interchange of limits, trace, and integrals. We note that the regularity conditions in Theorem 1 are same as that of [22, Theorem 1] specialized to the discrete setting. Also, observe that analogous to the classical case, the limits depend on whether $\rho = \sigma$ (null) or $\rho \neq \sigma$ (alternative), and that the convergence rate is faster in the former.

We briefly discuss the regularity assumptions in the above theorem. In the infinite dimensional case, $\mathsf{D}(\rho\|\sigma)$ can be unbounded even if the support conditions $\rho \ll \sigma$ are satisfied. This necessitates the finiteness assumption on the quantum relative entropies above. The condition $\mathbb{P}\big(\|\rho_n\sigma_n^{-1}\|_\infty > c\big) \to 0$ imposes a stochastic boundedness assumption on the operator $\rho_n\sigma_n^{-1}$ and is a natural condition for the existence of distributional limits even in the classical case (see [22, Theorem 2 and Remark 1]). To see this, take $\rho = \sigma = |0\rangle\langle 0|$, and

$$\rho_n = \big(1 - n^{-1}\big)|0\rangle\langle 0| + n^{-1}|n\rangle\langle n|,$$
$$\sigma_n = \big(1 - e^{-n^2}\big)|0\rangle\langle 0| + e^{-n^2}|n\rangle\langle n|,$$

where $|i\rangle$ denotes the i^{th} basis element of $\mathbb{H}$ (in an arbitrary orthonormal basis of $\mathbb{H}$) and $|i\rangle\langle i|$ denotes the unique operator which takes $|i\rangle$ to $|i\rangle$ and $|j\rangle$, for $j \neq i$, to the zero element. Note that $\|\rho_n\sigma_n^{-1}\|_\infty$ diverges. Observe that $\sqrt{n}(\rho_n - \rho) \xrightarrow{w} 0$ and $\sqrt{n}(\sigma_n - \sigma) \xrightarrow{w} 0$, where 0 denotes the zero operator. However, it is easily seen by a straightforward computation that $\mathsf{D}(\rho_n\|\sigma_n)$ diverges. Hence, the limit $\sqrt{n}\mathsf{D}(\rho_n\|\sigma_n)$ does not exist and Theorem 1 does not hold.

In the finite dimensional case, Theorem 1 leads to the following corollary.

Corollary 1 (Finite dimensional case) *Let $\rho_n \ll \sigma_n \ll \sigma$ and $\rho_n \ll \rho \ll \sigma$. The following hold:*

(i) (Two-sample alternative) If $\big(r_n(\rho_n-\rho), r_n(\sigma_n-\sigma)\big) \xrightarrow{w} (L_1, L_2)$, then (2) holds.

(ii) (Two-sample null) If $\rho = \sigma$ and $\big(r_n(\rho_n - \rho), r_n(\sigma_n - \rho)\big) \xrightarrow{w} (L_1, L_2)$, then (3) holds.

Since all norms are equivalent in finite dimensions, the choice of the norm in the space of density operators with respect to which weak convergence is considered does not matter. Also, note that the stochastic boundedness assumption used in Theorem 1 is not required.

Remark 1 (One-sample null and alternative) *The one-sample case refers to the setting when $\sigma_n = \rho$ (null) or $\sigma_n = \sigma$ (alternative) for all $n \in \mathbb{N}$, i.e., when only ρ is approximated by ρ_n. In this case, the respective limits can be obtained by letting $L_2 = 0$ in (2) and (3).*

Simplified expressions for the limit variables in Theorem 1 exist when all relevant density operators commute, as stated in the following corollary.

Corollary 2 (Commutative case) *If all operators in Theorem 1 commute, then*

$$r_n\big(\mathsf{D}(\rho_n\|\sigma_n) - \mathsf{D}(\rho\|\sigma)\big) \xrightarrow{w} \mathrm{Tr}\left[L_1(\log\rho - \log\sigma) - L_2\rho\sigma^{-1}\right]. \quad (4)$$

Additionally, when $\rho = \sigma$, then

$$r_n^2\mathsf{D}(\rho_n\|\sigma_n) \xrightarrow{w} \frac{1}{2}\mathrm{Tr}\left[\big(L_1 - L_2\big)^2\rho^{-1}\right]. \quad (5)$$

Equations (4) and (5) recovers [22, Theorem 2] specialized to the discrete setting with finite support, which is the classical analogue of Theorem 1 pertaining to KL divergence.

Specializing Theorem 1 to von Neumann entropy leads to the following result.

Corollary 3 (Limit distribution for von Neumann entropy) *Let $\rho_n \ll \rho$. If $r_n(\rho_n - \rho) \xrightarrow{w} L$ in trace-norm, then*

$$r_n\big(\mathsf{H}(\rho_n) - \mathsf{H}(\rho)\big) \xrightarrow{w} -\mathrm{Tr}\,[L\log\rho]. \quad (6)$$

The above claim follows from (2) with $L_1 = L$ and $L_2 = 0$ by noting that the regularity conditions in Part (ii) of Theorem 1 are satisfied with $\sigma_n = \sigma = I$ for all $n \in \mathbb{N}$, and using the relation $\mathsf{H}(\rho) = -\mathsf{D}(\rho\|I)$, where I denotes the identity operator. Since I is not a quantum state, here we used that the definition of quantum relative entropy in (1) extends to the case where σ is replaced by I without affecting any of the claims.

It is well-known that $\mathsf{H}(\rho) = \mathsf{H}(\lambda)$, where $\lambda \in \mathcal{H}_d$ denotes the diagonal operator comprising of the eigenvalues of ρ arranged in non-increasing order. In other words, $\mathsf{H}(\rho)$ equals the Shannon entropy of the probability distribution composed of the eigenvalues of ρ. An unbiased estimator of the spectrum of a quantum state is given by the EYD algorithm [43], [44] that outputs a Young's diagram as its estimate. In [46, Theorem 3.1], the limit distribution of this estimator with a scaling rate $n^{1/2}$ is characterized in terms of a d-dimensional Brownian functional $B = (B_1, \ldots, B_d)$. From Corollary 3 and the discussion above, it then follows that the asymptotic distribution of the EYD algorithm based estimator of $\mathsf{H}(\rho)$ is governed by (6) with $r_n = n^{1/2}$, $L = \mathrm{diag}(B)$ and $\rho = \lambda$, where $\mathrm{diag}(\cdot)$ denotes the operation of representing a vector as the diagonal elements of a matrix.

IV. Application

Limit theorems for classical divergences have several applications in statistics, computational science and biology such as constructing confidence intervals for hypothesis testing [21], auditing of differential privacy [22], and biological data analysis [23]. Here, we consider an application of Theorem 1 in establishing performance guarantees for the problem of testing for the quantum relative entropy between unknown states in the finite dimensional setting. The relevant multi-hypothesis testing problem can be formulated as[2]

$$H_i : \epsilon_i < \mathsf{D}\left(\rho_i \| \sigma_i\right) \leq \epsilon_{i+1}, \tag{7}$$

where $\epsilon_i \geq 0$ satisfy $\epsilon_{i+1} > \epsilon_i$ for $i \in \mathcal{I} = \{0, \ldots, m-1\}$. We are interested in the setting where approximately nd^2 identical copies of the unknown states are available for the tester. The goal then is to design a test $\mathcal{T}_n = \{M_n(i), i \in \mathcal{I}\}$ with $M_n(i) \geq 0$ for all i, and $\sum_{i \in \mathcal{I}} M_n(i) = I$ that achieves a specified performance, i.e., an m-outcome positive operator-valued measure (POVM) with index set $\mathcal{I}$ (see e.g., [16] for further details). Denoting the original hypothesis by H and the test outcome by $\hat{H}$, the performance of $\mathcal{T}_n$ is quantified by the error probabilities

$$\alpha_{i,n}(\mathcal{T}_n, \rho_i^{\otimes n}) := \mathbb{P}(\hat{H} \neq i | H = i) = \mathrm{Tr}\left[\rho_i^{\otimes n} \sum_{j \neq i} M_n(j)\right].$$

A test $\mathcal{T}_n$ is said to achieve level τ if $\alpha_{i,n}(\mathcal{T}_n, \rho_i^{\otimes n}) \leq \tau$ for every $i \in \mathcal{I}$. A sequence of tests $(\mathcal{T}_n)_{n \in \mathbb{N}}$ is asymptotically said to achieve level τ if $\limsup_{n \to \infty} \alpha_{i,n}(\mathcal{T}_n, \rho_i^{\otimes n}) \leq \tau$ for every $i \in \mathcal{I}$.

A pertinent approach to realize a hypothesis test is to first perform tomography of the states to obtain estimates, $\hat{\rho}_n, \hat{\sigma}_n$, and then compute the relative entropy between them. A standard class of tests (motivated from the Neyman-Pearson theorem) then decides in favor of H_i if $t_{i,n} < \mathsf{D}\left(\hat{\rho}_n \| \hat{\sigma}_n\right) \leq t_{i+1,n}$, where $t_{i,n}$ for $0 \leq i \leq m$ are critical values chosen according to the desired level $\tau_i \in (0,1]$ for i^{th} error probability. Each such test (statistic) T_n induces a POVM indexed by $\mathcal{I}$, denoted by $\mathcal{T}_n^{\mathrm{tom}}\big(\{t_i\}_{i \in \mathcal{I}}\big)$, for which we will use the shorthand $\mathcal{T}_n^{\mathrm{tom}}$. Let $\alpha_{i,n}\big(\mathcal{T}_n^{\mathrm{tom}}, \rho_i^{\otimes n}\big) = \mathbb{P}(T_n \neq i | H = i)$ denote the error probability for the test statistic T_n given the i^{th} hypothesis is true.

To obtain concrete performance guarantees for the aforementioned hypothesis test, we consider a specific tomographic estimator for density operators based on Pauli measurements. This can be considered as a quantum analogue of the classical plug-in estimator based on empirical probability distributions. We first describe the estimator and characterize its limiting distribution, which will then be used to construct the test statistic for (7).

A. Tomographic Estimator of Quantum States

Let $d = 2^N$ for some integer N, and $\{\gamma_j\}_{j=0}^{d^2-1}$ denote the set of multi-qubit (N-qubit) Pauli operators constructed as the N-fold tensor product of standard Pauli operators acting on a qubit. Specifically, $\gamma_j = \bigotimes_{i=1}^N \gamma_{j,i}$ with $\gamma_{j,i} \in \{R_k\}_{k=0}^3$, where $\{R_k\}_{k=0}^3$ denotes the single-qubit Pauli basis with the following representations in the standard basis:

$$R_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$

$$R_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We may take $\gamma_0 = I$. The multi-qubit Pauli operators are Hermitian and form an orthogonal operator basis for the real vector space $\mathcal{H}_d$ with respect to the Hilbert-Schmidt inner product. Consequently, any multi-qubit density operator ρ can be written as

$$\rho = \frac{I}{d} + \frac{1}{d} \sum_{j=1}^{d^2-1} s_j(\rho) \gamma_j, \tag{8}$$

with $s_j(\rho) = \mathrm{Tr}\left[\rho \gamma_j\right]$. Note that γ_j, for $1 \leq j \leq d^2 - 1$, are traceless and have eigenvalues ± 1. Moreover, any operator of the form (8), with $s_j(\rho)$ replaced by s_j such that $\|s\|_2 \leq 1$, is a valid density operator, where $s = (s_1, \ldots, s_{d^2-1})$. In particular, $\|s\|_2 = 1$ corresponds to *pure* states, while $\|s\|_2 < 1$ corresponds to *mixed* states, where pure and mixed states refer to a state ρ such that $\mathrm{Tr}\left[\rho^2\right] = 1$ and $\mathrm{Tr}\left[\rho^2\right] < 1$, respectively. Let Λ_j^+ (resp. Λ_j^-) and P_j^+ (resp. P_j^-) denote the set of outcomes and eigenspace corresponding to the eigenvalue $+1$ (resp. -1) of γ_j, respectively. Then

$$s_j(\rho) = s_j^+(\rho) - s_j^-(\rho),$$

where $s_j^+(\rho) := \mathrm{Tr}\left[\rho P_j^+\right]$ and $s_j^-(\rho) := \mathrm{Tr}\left[\rho P_j^-\right] = 1 - s_j^+(\rho)$.

Assume that identical copies of ρ and σ are available as desired, on which measurements using Pauli operators can be performed and the outcomes recorded. Let $O_k(j, \rho)$ denote the k^{th} measurement outcome using γ_j on ρ. A tomographic estimator of ρ and σ is then given by

$$\hat{\rho}_n = \mathbb{1}_{\|\hat{s}^{(n)}(\rho)\|_2 \leq 1} \hat{S}_n(\rho) + \frac{\mathbb{1}_{\|\hat{s}^{(n)}(\rho)\|_2 > 1}}{\left\|\hat{s}^{(n)}(\rho)\right\|_2} \hat{S}_n(\rho), \tag{9a}$$

and

$$\hat{\sigma}_n = \frac{I}{nd} + \left(1 - \frac{1}{n}\right) \Bigg(\mathbb{1}_{\|\hat{s}^{(n)}(\sigma)\|_2 \leq 1} \hat{S}_n(\sigma) + \frac{\mathbb{1}_{\|\hat{s}^{(n)}(\sigma)\|_2 > 1}}{\left\|\hat{s}^{(n)}(\sigma)\right\|_2} \hat{S}_n(\sigma) \Bigg),$$

[2]Note that here (ρ_i, σ_i), $1 \leq i \leq m$, denote pairs of quantum states and not random density operators as was used until now.

respectively, where

$$\hat{S}_n(\rho) := \frac{I}{d} + \frac{1}{d}\sum_{j=1}^{d^2-1} \hat{s}_j^{(n)}(\rho)\gamma_j,$$

$$\hat{s}_j^{(n)}(\rho) := \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{O_k(j,\rho)\in\Lambda_j^+} - \mathbb{1}_{O_k(j,\rho)\in\Lambda_j^-}, \quad j \neq 0,$$

$$\hat{s}^{(n)}(\rho) := \big(\hat{s}_1^{(n)}(\rho), \ldots, \hat{s}_{d^2-1}^{(n)}(\rho)\big).$$

It follows from the above discussion that $\hat{\rho}_n, \hat{\sigma}_n \in \mathcal{S}_d$ for all $n \in \mathbb{N}$. Note that the extra term (negligible asymptotically) I/nd ensures that $\hat{\sigma}_n > 0$ so that $\hat{\rho}_n \ll \hat{\sigma}_n$ and $\mathsf{D}\left(\hat{\rho}_n\|\hat{\sigma}_n\right)$ is finite. Also, observe that to construct $\hat{\rho}_n$ and $\hat{\sigma}_n$, we need $n(d^2-1)$ independent copies, each of ρ and σ, available for measurement.

Let $N(c, v^2)$ denote the one-dimensional normal distribution with mean c and variance v^2. The following result shows that the limit distribution for estimators of quantum relative entropy based on Pauli tomography is Gaussian.

Proposition 1 (Limit distribution for tomographic estimator) *Let $\rho, \sigma > 0$. Then*

$$\sqrt{n}\big(\mathsf{D}\left(\hat{\rho}_n\|\sigma\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} W_1 \sim N\big(0, v_1^2(\rho,\sigma)\big),$$
$$\sqrt{n}\big(\mathsf{D}\left(\hat{\rho}_n\|\hat{\sigma}_n\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} W_2 \sim N\big(0, v_2^2(\rho,\sigma)\big),$$

where

$$v_1^2(\rho,\sigma) := \sum_{j=1}^{d^2-1} \frac{4 s_j^+(\rho) s_j^-(\rho)}{d^2} \operatorname{Tr}\left[\gamma_j(\log\rho - \log\sigma)\right]^2,$$

$$v_2^2(\rho,\sigma) := v_1^2(\rho,\sigma) + \sum_{j=1}^{d^2-1} \frac{4 s_j^+(\sigma) s_j^-(\sigma)}{d^2} \operatorname{Tr}\left[\rho D[\log\sigma](\gamma_j)\right]^2.$$

The proof of Proposition 1 follows by an application of Corollary 1. The main ingredient of the proof is to show that $\big(\sqrt{n}(\hat{\rho}_n - \rho), \sqrt{n}(\hat{\sigma}_n - \sigma)\big) \xrightarrow{w} (L_\rho, L_\sigma)$, where $L_\rho := \sum_{j=1}^{d^2-1} \gamma_j Z_j(\rho)$ and $Z_j(\rho) \sim N\big(0, 4 s_j^+(\rho) s_j^-(\rho)/d^2\big)$. The claim then follows from (2) by noting that all relevant regularity conditions are satisfied. In particular, the assumption $\rho, \sigma > 0$ ensures that the support conditions are satisfied.

B. Performance Guarantees for Quantum Multi-hypothesis Testing

For simplicity of presentation, we will assume that $\sigma_i = \sigma$ for all $i \in \mathcal{I}$ with σ known for the test in (7). Such a scenario arises, for instance, when testing for the mixedness of an unknown state ρ with respect to the maximally mixed state, $\sigma = \pi_d$. Also, for $\tau \in [0,1]$, let

$$Q^{-1}(\tau) = \inf\left\{z \in \mathbb{R} : (2\pi)^{-1/2}\int_z^\infty e^{-u^2/2} du \leq \tau\right\},$$

be the inverse complimentary cumulative distribution function of the standard normal distribution $N(0,1)$. The following proposition provides a test statistic for the multi-hypothesis testing problem in (7) by utilizing the limit distribution for quantum relative entropy and characterizes its error probabilities.

Proposition 2 (Performance of multi-hypothesis testing) *Let $\tau \in (0,1]$, and $\rho_i, \sigma > 0$ for $i \in \mathcal{I}$ satisfy the hypotheses in (7) for $\sigma_i = \sigma$ therein. Let $\hat{D}_n = \mathsf{D}\left(\hat{\rho}_n\|\sigma\right)$ with $\hat{\rho}_n$ given in (9a). Then, the test statistic*

$$T_n = \sum_{i\in\mathcal{I}} i \mathbb{1}_{\hat{D}_n \in \mathcal{L}_{i,n}(c)},$$

with $\mathcal{L}_{i,n}(c) := (\epsilon_i + cn^{-1/2}, \epsilon_{i+1} + cn^{-1/2})$, asymptotically achieves a level τ provided

$$c \geq 2d\, Q^{-1}(\tau)|\log b|,$$

where b denotes the minimum of the eigenvalues of σ and ρ_i over all $i \in \mathcal{I}$.

The proof of Proposition 2 follows by an application of Proposition 1 and Portmanteau theorem [51, Theorem 2.1]. The threshold c achieving a desired asymptotic level τ is determined by utilizing the knowledge that $\sqrt{n}\big(\hat{D}_n - \mathsf{D}\left(\rho_i\|\sigma\right)\big)$ converges in distribution to a centered normal under hypothesis i, whose variance $v_1^2(\rho_i, \sigma)$ can be uniformly bounded for all $i \in \mathcal{I}$.

Remark 2 (Growing number of hypotheses) *It can be shown that Proposition 2 continues to hold even when the number of hypotheses scales with n, given the new hypotheses boundaries are chosen consistent with the previous ones and are well-separated, i.e.,* $\min_{i\in\mathcal{I}_n} \mathsf{D}\left(\rho_i\|\sigma\right) - \epsilon_i = \omega(n^{-1/2})$*, where $\omega(\cdot)$ denotes the asymptotic little omega notation and $\mathcal{I}_n$ is the index set of hypotheses that grows with n ($\mathcal{I}_n \subseteq \mathcal{I}_{n+1}$ for every $n \in \mathbb{N}$).*

V. Concluding Remarks

This paper studied limit distributions for a certain class of estimators of quantum relative entropy. Taking recourse to an operator version of Taylor's theorem, the limit distributions are characterized in terms of trace functionals of first or second-order Fréchet derivatives of elementary functions. We employed the derived results to show that the asymptotic distribution of an estimator of quantum relative entropy based on Pauli tomography of states is normal. We then utilized this knowledge to propose a test statistic for a multi-hypothesis testing problem and characterized its asymptotic performance.

Looking forward, several open questions remain. One pertinent question concerns the rate of convergence of the empirical distribution of the divergence to its limit in the flavor of classical Berry-Esseen theorem. Of interest further is to understand the asymptotic and non-asymptotic behavior of other classes of estimators such as those based on variational methods. Lastly, it would also be beneficial to study the statistical behaviour of other divergences such as quantum χ^2 divergence [52] and quantum Rényi divergences [11], [13], [14], [53], [54] (see [50] for further results on quantum Rényi divergences).

Acknowledgement

SS acknowledges support from the Excellence Cluster - Matter and Light for Quantum Computing (ML4Q). MB acknowledges funding from the European Research Council (ERC Grant Agreement No. 948139).

References

[1] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, "Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum," *Phys. Rev. Lett.*, vol. 70, pp. 1244–1247, Mar 1993.

[2] R. Blume-Kohout, "Optimal, reliable estimation of quantum states," *New J. of Phys.*, vol. 12, no. 4, Apr. 2010.

[3] M. Cramer, M. Plenio, S. Flammia, R. Somma, D. Gross, S. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, "Efficient quantum state tomography," *Nature Commun.*, vol. 1, p. 149, 12 2010.

[4] M. Christandl and R. Renner, "Reliable quantum state tomography," *Phys. Rev. Lett.*, vol. 109, 08 2011.

[5] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-optimal tomography of quantum states," in *Proc. 48th Annu. ACM Symp. Theory Comput.*, ser. STOC '16, New York, NY, USA, 2016, p. 913–925.

[6] J. Wright, "How to learn a quantum state," Ph.D. dissertation, Carnegie Mellon University, 2016.

[7] R. O'Donnell and J. Wright, "Efficient quantum tomography," in *Proc. 48th Annu. ACM Symp. Theory Comput.*, ser. STOC '16, New York, NY, USA, 2016, p. 899–912.

[8] H. Yuen, "An improved sample complexity lower bound for (fidelity) quantum state tomography," *Quantum*, vol. 7, p. 890, Jan. 2023.

[9] H. Umegaki, "Conditional expectations in an operator algebra IV (entropy and information)," *Kodai Math. Seminar Rep.*, vol. 14, no. 2, pp. 59–85, 1962.

[10] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. Probab.*, vol. 1. Berkely: University of California Press, 1961, pp. 547–561.

[11] D. Petz, "Quasi-entropies for states of a von Neumann algebra," *Publ. Res. Inst. Math. Sci.*, vol. 21, no. 4, pp. 787–800, Aug. 1985.

[12] ——, "Quasi-entropies for finite quantum systems," *Rep. Math. Phys.*, vol. 23, no. 1, pp. 57–65, Feb. 1986.

[13] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties," *J. Math. Phys.*, vol. 54, no. 12, 2013.

[14] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Commun. Math. Phys.*, vol. 331, pp. 593–622, 2014.

[15] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge University Press, 2017.

[16] M. Tomamichel, *Quantum Information Processing with Finite Resources: Mathematical Foundations*. Springer, 2015.

[17] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*. Springer Berlin, Heidelberg, 2016.

[18] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Commun. Math. Phys.*, vol. 143, no. 1, pp. 99 – 114, 1991.

[19] M. Nussbaum and A. Szkola, "The Chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Stat.*, vol. 37, 08 2006.

[20] Audenaert, KMR and Nussbaum, M and Szkola, A and Verstraete, Frank, "Asymptotic error rates in quantum hypothesis testing," *Commun. Math. Phys.*, vol. 279, no. 1, pp. 251–283, 2008.

[21] R. Okano and M. Imaizumi, "Inference for projection-based wasserstein distances on finite spaces," *Statistica Sinica*, vol. 34, no. 2, 2024.

[22] S. Sreekumar, Z. Goldfeld, and K. Kato, "Limit distribution theory for f-divergences," *IEEE Trans. Inf. Theory*, vol. 70, no. 2, pp. 1233–1267, 2024.

[23] M. Pietrzak, G. Rempala, M. Seweryn, and J. Wesolowski, "Limit theorems for empirical Rényi entropy and divergence with applications to molecular diversity analysis," *TEST*, vol. 25, 05 2015.

[24] C. Hirche, C. Rouzé, and D. S. França, "Quantum differential privacy: An information theory perspective," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5771–5787, 2023.

[25] S. Sreekumar, Z. Goldfeld, and K. Kato, "Limit distribution theory for KL divergence and applications to auditing differential privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Taipei, Taiwan, 2023, pp. 2607–2612.

[26] L. Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, no. 6, p. 1191–1253, Jun. 2003.

[27] N. Leonenko, L. Pronzato, and V. Savani, "A class of Rényi information estimators for multidimensional densities," *Ann. Stat.*, vol. 36, no. 5, pp. 2153 – 2182, 2008.

[28] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Estimating divergence functionals and the likelihood ratio by convex risk minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5847–5861, Oct. 2010.

[29] K. Kandasamy, A. Krishnamurthy, B. Poczos, L. Wasserman, and J. M. Robins, "Nonparametric von Mises estimators for entropies, divergences and mutual informations," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Montréal, Canada, Dec. 2015, pp. 397–405.

[30] S. Singh and B. Póczos, "Finite-sample analysis of fixed-k nearest neighbor density functional estimators," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, Barcelona, Spain, Dec. 2016, pp. 1225–1233.

[31] T. B. Berrett, R. J. Samworth, and M. Yuan, "Efficient multivariate entropy estimation via k-nearest neighbour distances," *Ann. Stat.*, vol. 47, no. 1, pp. 288–318, Feb. 2019.

[32] K. Moon and A. Hero, "Multivariate f-divergence estimation with confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, Montréal, Canada, 2014, pp. 2420–2428.

[33] S. Sreekumar and Z. Goldfeld, "Neural estimation of statistical divergences," *J. Mach. Learn. Res.*, vol. 23, no. 126, pp. 1–75, 2022.

[34] Z. Goldfeld and K. Kato, "Limit distributions for smooth total variation and χ^2-divergence in high dimensions," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 2640–2645.

[35] T. Li and X. Wu, "Quantum query complexity of entropy estimation," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2899–2921, Mar. 2017.

[36] J. Acharya, I. Issa, N. V. Shende, and A. B. Wagner, "Estimating quantum entropy," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 2, pp. 454–468, 2020.

[37] S. Subramanian and M.-H. Hsieh, "Quantum algorithm for estimating Rényi entropies of quantum states," *Phys. Rev. A*, vol. 104, no. 2, p. 022428, Aug. 2021.

[38] T. Gur, M.-H. Hsieh, and S. Subramanian, "Sublinear quantum algorithms for estimating von neumann entropy," *Electron. Colloquium Comput. Complex.*, vol. TR21, 2021.

[39] S. Rethinasamy, R. Agarwal, K. Sharma, and M. M. Wilde, "Estimating distinguishability measures on quantum computers," *Phys. Rev. A*, vol. 108, Jul 2023.

[40] Y. Wang, B. Zhao, and X. Wang, "Quantum algorithms for estimating quantum entropies," *Phys. Rev. Appl.*, vol. 19, Apr 2023.

[41] Q. Wang, Z. Zhang, K. Chen, J. Guan, W. Fang, J. Liu, and M. Ying, "Quantum algorithm for fidelity estimation," *IEEE Trans. Inf. Theory*, vol. 69, no. 1, pp. 273–282, Mar. 2023.

[42] Z. Goldfeld, D. Patel, S. Sreekumar, and M. M. Wilde, "Quantum neural estimation of entropies," *Phys. Rev. A*, vol. 109, p. 032431, Mar 2024.

[43] 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, 05 1988.

[44] M. Keyl and R. F. Werner, "Estimating the spectrum of a density operator," *Phys. Rev. A*, vol. 64, p. 052311, Oct 2001.

[45] C. Tracy and H. Widom, "On the distributions of the lengths of the longest monotone subsequences in random words," *Probab. Theory Relat. Fields*, vol. 119, 04 1999.

[46] C. Houdré and H. Xu, "On the limiting shape of Young diagrams associated with inhomogeneous random words," in *High Dimensional Probab. VI*. Basel: Springer Basel, 2013, pp. 277–302.

[47] M. Guta and A. Jencova, "Local asymptotic normality in quantum statistics," *Communications in Mathematical Physics*, vol. 276, 12 2007.

[48] K. Yamagata, A. Fujiwara, and R. Gill, "Quantum local asymptotic normality based on a new quantum likelihood ratio," *The Annals of Statistics*, vol. 41, pp. 2197–2217, 08 2013.

[49] R. Bhatia, *Matrix Analysis*. Springer, 1997, vol. 169.

[50] S. Sreekumar and M. Berta, "Limit distribution theory for quantum divergences," *arXiv:2311.13694*, 2023.

[51] P. Billingsley, *Convergence of probability measures*, 2nd ed. New York: John Wiley & Sons Inc., 1999.

[52] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete, "The χ^2-divergence and mixing times of quantum Markov processes," *J. Math. Phys.*, vol. 51, no. 12, p. 122201, 12 2010.

[53] K. Matsumoto, "A new quantum version of f-divergence," in *Nagoya Winter Workshop: Reality and Measurement in Algebraic Quantum Theory (Springer)*, Nagoya, Japan, Mar. 2015, pp. 229–273.

[54] K. Fang and H. Fawzi, "Geometric Rényi divergence and its applications in quantum channel capacities," *Commun. Math. Phys.*, vol. 384, pp. 1615 – 1677, 2019.