IEEE Transactions on Information Theory

Limit Distribution Theory for Quantum Divergences

Journal:	*IEEE Transactions on Information Theory*
Manuscript ID	IT-24-0081
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	01-Feb-2024
Complete List of Authors:	Sreekumar, Sreejith; RWTH Aachen University, Institute for Quantum Information Berta, Mario; RWTH Aachen University, Institute for Quantum Information
Keywords:	Quantum divergences, Limit distribution, Divergence estimation, Frechet derivative, Hypothesis testing
Subject Category:	Quantum

Limit Distribution Theory for Quantum Divergences

Sreejith Sreekumar and Mario Berta

Abstract

Estimation of quantum relative entropy and its Rényi generalizations is a fundamental statistical task in quantum information theory, physics, and beyond. While several estimators of these divergences 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.

Index Terms

Quantum divergences, limit distribution, divergence estimation, Fréchet derivative, hypothesis testing

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 the aforementioned quantum divergences. Given two quantum states ρ and σ with corresponding estimators ρ_n and σ_n, respectively, and a divergence $\mathsf{D}(\rho, \sigma)$, we want to identify

S. Sreekumar (email: sreekumar@physik.rwth-aachen.de) and M. Berta are with the Institute for Quantum Information at RWTH Aachen University, Germany.

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:

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

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 divergences are 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 and dual formulations for divergences.

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 $\rho\neq\sigma$:

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

Here[1], $D[f(A)](B)$ denotes the first-order Fréchet derivative (see Section II for precise definitions) 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. We then consider two prominent quantum Rényi divergences, namely, the Petz-Rényi [11] and the minimal or sandwiched Rényi divergences of order α [13], [14]. While the former admits an analysis similar to Theorem 1, for the latter, we consider a variational form based on dual expressions and derive the limits by applying Taylor's theorem to an intermediate quantity which is easier to analyze. We also use a similar approach to derive the limit distribution for measured relative entropy estimation under a certain assumption on the uniqueness of the optimal measurement.

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

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

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

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

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 in [26]–[39]. 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 [22], [25], [30], [40]–[44], 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 [45]–[52], as well as their regularized versions [50], [53]–[63].

In the quantum setting, computational complexities of various estimators of quantum information measures have been investigated under different input models [64]–[78]. Specifically, [66] established copy complexity bounds characterizing the optimal dimension dependence for quantum Rényi entropy estimation when independent copies of the state are available for measurement. [67], [69] 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 [79], [80] was determined in [81], [82]. However, to the best of our knowledge, a limit distribution theory for quantum divergences has not been explored before. Here, we study this aspect focusing mostly on finite dimensional quantum systems (except in Section III-C where we treat quantum relative entropy between density operators on an infinite-dimensional separable Hilbert space).

B. Paper Organization

The rest of the paper is organized as follows. Section II introduces the notation and preliminary concepts required for stating our results. The main results on limit distributions of quantum divergences are presented in Section III. The applications, namely limit distributions of quantum relative entropy for tomographic estimators and performance guarantees for a multi-hypothesis testing problem, are discussed in Section IV. This is followed by concluding remarks with avenues for future research in Section V. The proofs of the main results and applications are furnished in Section VI while those of the auxiliary lemmas are provided in the Appendix.

II. Preliminaries

A. Notation

For most part, we consider a finite dimensional complex Hilbert space $\mathbb{H}_d$ of dimension d. Denote the set of linear operators from $\mathbb{H}_d$ to $\mathbb{H}_d$ by $\mathcal{L}(\mathbb{H}_d)$. Without loss of generality, we identify $\mathbb{H}_d$ and $\mathcal{L}(\mathbb{H}_d)$ with $\mathbb{C}^d$ and $\mathbb{C}^{d\times d}$, respectively. Denote the set of all $d \times d$ Hermitian, positive semi-definite, positive definite, and unitary operators by $\mathcal{H}_d$, $\mathcal{P}_d$, $\mathcal{P}_d^+$ and $\mathcal{U}_d$, respectively. Let $\mathcal{S}_d$ denote the set of density operators, i.e., the set of elements of $\mathcal{P}_d$ with unit trace, and $\mathcal{S}_d^+$ be its subset with strictly positive eigenvalues. We use $[A,B] := AB - BA$ to represent the commutator of two operators A and B. $\mathrm{Tr}\,[\cdot]$ and $\|\cdot\|_p$ for $p \geq 1$ signifies the trace operation and Schatten p-norm, respectively. The notation $\leq$ denotes the Löwner partial order in the context of operators, i.e., for $A, B \in \mathcal{H}_d$, $A \leq B$ means that $B - A \in \mathcal{P}_d$. $\mathbb{1}_{\mathcal{X}}$ denotes indicator of a set $\mathcal{X}$ and I denotes the identity operator on $\mathbb{H}_d$. For linear operators A, B, $A \ll B$ designates that the support of A is contained in that of B, and $A \ll\gg B$ means that $B \ll A \ll B$. A^{-1} stands for the generalized (Moore–Penrose) inverse of an operator A.

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 (taking values 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 $\mathcal{S}_d$. Lastly, we use $a \lesssim_x b$ to denote that $a \leq c_x b$ for some constant $c_x > 0$ which depends only on x, and adopt the conventions $0/0 = 0$, $c/0 = \infty$ for $c > 0$ and $\infty \cdot 0 = 0$.

B. Fréchet Differentiability

Definition 1 (Fréchet differentiability, see e.g. [83]) *For an open set $\mathcal{X} \subseteq \mathbb{R}$, let $f : \mathcal{X} \to \bar{\mathbb{R}}$ and $A \in \mathcal{H}_d$ with* $\mathsf{spec}(A) \subseteq \mathcal{X}$. *Then, f is called (Fréchet) differentiable at A if there exists a linear map $D[f(A)] : \mathcal{H}_d \to \mathcal{H}_d$ such that for all $H \in \mathcal{H}_d$ such that* $\mathsf{spec}(A+H) \subseteq \mathcal{X}$*,*

$$\|f(A+H) - f(A) - D[f(A)](H)\| = o(\|H\|). \tag{1}$$

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 $\mathcal{H}_d$ into $\mathcal{L}(\mathcal{H}_d)$ 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 $D^2[f(A)] : \mathcal{H}_d \times \mathcal{H}_d \to \mathcal{H}_d$.

If f is differentiable at A, then

$$D[f(A)](H) = \frac{d}{dt} f(A+tH)\big|_{t=0}, \ \forall\ H \in \mathcal{H}_d.$$

The chain rule holds: the composition of two differentiable maps f and g is differentiable and $D[g \circ f(A)] = D[g \circ f(A)]D[f(A)]$. Also, we have the product rule: for two differentiable maps f and g and $h = fg$,

$$D[h(A)](H) = f(A)D[g(A)](H) + D[f(A)](H)g(A).$$

Finally, we will frequently use that

$$D[A^{-1}](H) = -A^{-1}HA^{-1}. \tag{2}$$

C. Quantum Information Measures

The von Neumann entropy of a density operator $\rho \in \mathcal{S}_d$ is

$$\mathsf{H}(\rho) := -\operatorname{Tr}[\rho \log \rho],$$

For density operators $\rho, \sigma \in \mathcal{S}_d$, the quantum relative entropy [9] is

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

From the above two definitions, it follows that

$$\mathsf{H}(\rho) = \log d - \mathsf{D}(\rho\|\pi_d), \tag{3}$$

where $\pi_d = I/d$ is the maximally mixed state. By some abuse of notation, the classical relative entropy or Kullback-Leibler (KL) divergence [84] between two distributions P, Q on a discrete alphabet $\mathcal{X}$ is

$$\mathsf{D}(P\|Q) := \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)},$$

if $P \ll Q$, and ∞ otherwise.

For $\alpha \in (0,1) \cup (1,\infty)$, the Petz-Rényi divergence [11] between $\rho, \sigma \in \mathcal{S}_d$ is

$$\bar{\mathsf{D}}_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\operatorname{Tr}\left[\rho^\alpha\sigma^{1-\alpha}\right], & \text{if } \rho \ll \sigma \text{ or } \rho \not\perp \sigma \text{ for } \alpha \in (0,1), \\ \infty, & \text{otherwise.} \end{cases} \tag{4}$$

$\bar{\mathsf{D}}_\alpha(\rho\|\sigma)$ satisfies the data-processing inequality for $\alpha \in (0,2]$. For $\alpha \in (0,1) \cup (1,\infty)$, the sandwiched Rényi divergence [13], [14] between $\rho, \sigma \in \mathcal{S}_d$ is

$$\tilde{\mathsf{D}}_\alpha(\rho\|\sigma) := \begin{cases} \frac{\alpha}{\alpha-1}\log\left\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha, & \text{if } \rho \ll \sigma \text{ or } \rho \not\perp \sigma \text{ for } \alpha \in (0,1), \\ \infty, & \text{otherwise.} \end{cases} \tag{5}$$

$\tilde{\mathsf{D}}_\alpha(\rho\|\sigma)$ satisfies data-processing inequality for $\alpha \geq 1/2$. Also, note that $\tilde{\mathsf{D}}_{1/2}(\rho\|\sigma) = -\log F(\rho,\sigma)$, where $F(\rho,\sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2 = \big(\operatorname{Tr}\left[\left|\sqrt{\rho}\sqrt{\sigma}\right|\right]\big)^2$ denotes the fidelity [85], [86] between ρ and σ. The max-divergence between states ρ and σ is

$$\mathsf{D}_{\max}(\rho\|\sigma) := \lim_{\alpha\to\infty}\tilde{\mathsf{D}}_\alpha(\rho\|\sigma) = \inf\left\{\lambda : \rho \leq e^\lambda\sigma\right\}. \tag{6}$$

This divergence is the unique quantum generalization of the classical Rényi divergence of infinite order that satisfies the data-processing inequality. For further details about the aforementioned information measures, see the books [15], [16].

In the next section, we derive limit distribution for estimators of the aforementioned information measures.

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* (see Section II-A for definitions). Since all norms are equivalent in finite dimensions, the choice of trace norm does not incur any loss of generality. 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.

A. Quantum Relative Entropy

Theorem 1 (Limit distribution for quantum relative entropy) *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

$$r_n\big(\mathsf{D}\left(\rho_n\|\sigma_n\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} \mathrm{Tr}\left[L_1(\log\rho - \log\sigma) - \rho D[\log\sigma](L_2)\right]. \tag{7}$$

(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

$$r_n^2\mathsf{D}\left(\rho_n\|\sigma_n\right) \xrightarrow{w} \mathrm{Tr}\left[L_1 D[\log\rho](L_1 - L_2)\right] + \mathrm{Tr}\left[\frac{\rho}{2}D^2[\log\rho](L_1 - L_2, L_1 - L_2)\right]. \tag{8}$$

The proof of Theorem 1 is presented in Section VI-B. The key idea 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 (see Section VI) 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.

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 (7) and (8).*

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

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

$$r_n\big(\mathsf{D}\left(\rho_n\|\sigma_n\right) - \mathsf{D}\left(\rho\|\sigma\right)\big) \xrightarrow{w} \mathrm{Tr}\left[L_1(\log\rho - \log\sigma) - L_2\rho\sigma^{-1}\right]. \tag{9}$$

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

$$r_n^2\mathsf{D}\left(\rho_n\|\sigma_n\right) \xrightarrow{w} \frac{1}{2}\,\mathrm{Tr}\left[\left(L_1 - L_2\right)^2\rho^{-1}\right]. \tag{10}$$

Equations (9) and (10) recovers [22, Theorem 2] specialized to the discrete setting with finite support, which is the classical analogue of Theorem 1 pertaining to KL divergence. A class of divergences intermediate between the classical and quantum relative entropy are the measured relative entropies [18], [87]–[89], which equals the largest KL divergence between the output probability distributions induced by a set of measurements (quantum to classical channel) on two quantum states. In Appendix B, we characterize the distributional limits for an estimator of this quantity under a uniqueness assumption on the optimal measurement.

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

Corollary 2 (Limit distribution for von Neumann entropy) *Let* $\rho_n \ll \rho$. *If* $r_n(\rho_n - \rho) \xrightarrow{w} L$, *then*

$$r_n\big(\mathsf{H}(\rho_n) - \mathsf{H}(\rho)\big) \xrightarrow{w} -\operatorname{Tr}[L \log \rho]. \tag{11}$$

Proof. The claim follows from (3) and (7) with $L_1 = L$ and $L_2 = 0$ by noting that the regularity conditions in Part (ii) of Theorem 1 are satisfied with $\sigma_n = \sigma = \pi_d$ for all $n \in \mathbb{N}$. □

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 [79], [80] that outputs a Young's diagram as its estimate. In [82, 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 2 and the discussion above, it then follows that the asymptotic distribution of the EYD algorithm based estimator of $\mathsf{H}(\rho)$ is governed by (11) with $r_n = n^{1/2}$, $L = \operatorname{diag}(B)$ and $\rho = \lambda$, where $\operatorname{diag}(\cdot)$ denotes the operation of representing a vector as the diagonal elements of a matrix.

B. Quantum Rényi Divergence

In contrast to the classical case, there is no single definition of quantum Rényi divergence known that satisfies all natural properties desired of a quantum information measure for all α (see [16]). Among the infinite possibilities, the most important ones with direct operational significance are the Petz-Rényi and sandwiched Rényi divergences, which we consider below.

Petz-Rényi divergence: We first derive distributional limits for Petz-Rényi divergence estimators.

Theorem 2 (Limit distribution for Petz-Rényi divergence) *Let* $\alpha \in (0,1) \cup (1,2]$ *and* $\bar{\alpha} = 1 - \alpha$. *Suppose* $\rho_n \ll \sigma_n \ll \sigma$ *and* $\rho_n \ll \rho \ll \sigma$. *Then, the following hold:*

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

$$r_n\big(\bar{\mathsf{D}}_\alpha(\rho_n \| \sigma_n) - \bar{\mathsf{D}}_\alpha(\rho \| \sigma)\big) \xrightarrow{w} \frac{\operatorname{Tr}[\sigma^{\bar{\alpha}} D[\rho^\alpha](L_1)] + \operatorname{Tr}[\rho^\alpha D[\sigma^{\bar{\alpha}}](L_2)]}{(\alpha - 1)\operatorname{Tr}[\rho^\alpha \sigma^{\bar{\alpha}}]}. \tag{12}$$

(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*

$$r_n^2 \bar{\mathsf{D}}_\alpha(\rho_n \| \sigma_n) \xrightarrow{w} \frac{\operatorname{Tr}\left[\rho^{\bar{\alpha}} D^2[\rho^\alpha](L_1, L_1) + \rho^\alpha D^2[\rho^{\bar{\alpha}}](L_2, L_2) + 2D[\rho^\alpha](L_1) D[\rho^{\bar{\alpha}}](L_2)\right]}{2(\alpha - 1)}. \tag{13}$$

The proof of Theorem 2 is given in Section VI-C, and utilizes a similar approach as Theorem 1. As in the case of quantum relative entropy, the limits and the scaling rate differ in the null and alternative. Observe that we consider Petz-Rényi divergence of order less than two since it does not satisfy the data-processing inequality above this value. In the commutative case, the expressions in (13) and (12) simplify further, and in particular, we deduce the limit distributions for estimators of classical Rényi divergences as stated below (see [23] for a more general result when the dimension scales with n).

Corollary 3 (Commutative case) *If all operators in Theorem 2 commute, then*

$$r_n\big(\bar{\mathsf{D}}_\alpha\left(\rho_n\|\sigma_n\right)-\bar{\mathsf{D}}_\alpha\left(\rho\|\sigma\right)\big) \xrightarrow{w} \frac{\alpha\operatorname{Tr}\left[L_1\sigma^{\bar{\alpha}}\rho^{-\bar{\alpha}}\right]+\bar{\alpha}\operatorname{Tr}\left[L_2\rho^{\alpha}\sigma^{-\alpha}\right]}{(\alpha-1)\operatorname{Tr}\left[\rho^{\alpha}\sigma^{\bar{\alpha}}\right]}. \tag{14}$$

Moreover, if $\rho=\sigma$, then

$$r_n^2\bar{\mathsf{D}}_\alpha\left(\rho_n\|\sigma_n\right) \xrightarrow{w} \frac{\operatorname{Tr}\left[(L_1-L_2)^2\rho^{-1}\right]}{2(\alpha-1)}. \tag{15}$$

Sandwiched Rényi divergence: We next consider the minimal or sandwiched Rényi divergence of order α.

Theorem 3 (Limit distribution for sandwiched Rényi divergence) *Let $\alpha\in[0.5,1)\cup(1,\infty)$ and $\bar{\alpha}=1-\alpha$. Suppose $\rho_n\ll\sigma_n\ll\sigma$ and $\rho_n\ll\rho\ll\sigma$. If $\big(r_n(\rho_n-\rho),r_n(\sigma_n-\sigma)\big)\xrightarrow{w}(L_1,L_2)$, then*

$$\begin{aligned}&r_n\big(\tilde{\mathsf{D}}_\alpha\left(\rho_n\|\sigma_n\right)-\tilde{\mathsf{D}}_\alpha\left(\rho\|\sigma\right)\big)\\&\xrightarrow{w}\frac{\alpha}{\alpha-1}\frac{\operatorname{Tr}\Big[\Big(D[\rho^{\frac{1}{2}}](L_1)\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}D[\sigma^{\frac{\bar{\alpha}}{\alpha}}](L_2)\rho^{\frac{1}{2}}\Big)\big(\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\big)^{\alpha-1}\Big]}{\big\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\big\|_\alpha^\alpha}.\end{aligned} \tag{16}$$

If all operators above commute, then (16) simplifies to (14).

The proof of Theorem 3 is given in Section VI-D. Different from the approach used in the previous results, we compute this limit by recasting the term $\|\rho^{1/2}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{1/2}\|_\alpha$ in (5) as a maximization using dual expressions for Schatten norms. However, establishing the desired limit with the new expression is more involved on account of the additional maximization involved. To this end, we consider upper and lower bounds (without the maximization) and show that they coincide with the expression in (16) asymptotically, thus establishing the claim. We mention here that the right-hand side (RHS) of (16) vanishes when $\rho=\sigma$, showing that correct scaling rate in the null setting for a non-degenerate limit should be r_n^2. However, the above technique does not lead to a proof of this claim due to non-matching upper and lower limits.

As a corollary of Theorem 3, we characterize the limit distributions for fidelity and max-divergence.

Corollary 4 (Fidelity and max-divergence) *Let $\rho_n\ll\sigma_n\ll\sigma$ and $\rho_n\ll\rho\ll\sigma$. If $\big(r_n(\rho_n-\rho),r_n(\sigma_n-\sigma)\big)\xrightarrow{w}(L_1,L_2)$, then*

$$\begin{aligned}&r_n\big(F(\rho_n,\sigma_n)-F(\rho,\sigma)\big)\\&\xrightarrow{w}\big(F(\rho,\sigma)\big)^{\frac{1}{2}}\operatorname{Tr}\Big[\Big(D[\rho^{\frac{1}{2}}](L_1)\sigma\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}L_2\rho^{\frac{1}{2}}\Big)\big(\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}\big)^{-\frac{1}{2}}\Big],\end{aligned}$$

and

$$\begin{aligned}&r_n\big(\mathsf{D}_{\max}(\rho_n\|\sigma_n)-\mathsf{D}_{\max}(\rho\|\sigma)\big)\\&\xrightarrow{w}e^{-\mathsf{D}_{\max}(\rho\|\sigma)}\operatorname{Tr}\Big[\Big(D[\rho^{\frac{1}{2}}](L_1)\sigma^{-1}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{-1}D[\rho^{\frac{1}{2}}](L_1)-\rho^{\frac{1}{2}}\sigma^{-1}L_2\sigma^{-1}\rho^{\frac{1}{2}}\Big)\Pi_{\max}\Big],\end{aligned}$$

where $\Pi_{\max}$ denotes the eigenprojection corresponding to the maximal eigenvalue of $\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}$.

C. Generalization to Infinite-dimensional Quantum Systems

Here, we consider a generalization of Theorem 1 to infinite-dimensional quantum systems when the underlying Hilbert space is separable. The appropriate notion of Fréchet differentiability relevant for our purposes is that of an operator-valued function on the space of Hermitian operators with bounded trace-norm, i.e., A and H in Definition 1 are required to have finite trace-norm, and $o(\|H\|)$ is replaced by $o(\|H\|_1)$ in (1). Defining the Frétchet derivative in this manner, the following result (see Section VI-F for proof) provides sufficient conditions under which limit distribution for quantum relative entropy exists.

Theorem 4 (Quantum relative entropy: Infinite dimensional case) *Let $\rho_n \ll \sigma_n \ll \sigma$ and $\rho_n \ll \rho \ll \sigma$ be such that $\mathsf{D}\left(\rho_n\|\sigma_n\right) < \infty$, $\mathsf{D}\left(\rho\|\sigma\right) < \infty$, and there exists a constant c satisfying $\mathbb{P}\big(\left\|\rho_n\sigma_n^{-1}\right\|_\infty > c\big) \rightarrow 0$. 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 (7) holds.

(ii) (Two-sample null) If $\rho=\sigma$ and $\big(r_n(\rho_n-\rho), r_n(\sigma_n-\rho)\big) \xrightarrow{w} (L_1, L_2)$ in trace norm, then (8) holds.

We briefly discuss the regularity assumptions in the above theorem. In the infinite dimensional case, $\mathsf{D}\left(\rho\|\sigma\right)$ 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 $\mathbb{P}\big(\left\|\rho_n\sigma_n^{-1}\right\|_\infty > c\big) \rightarrow 0$ 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, employing the bra-ket notation from quantum theory, take $\rho=\sigma=|0\rangle\langle 0|$, and

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

so that $\left\|\rho_n\sigma_n^{-1}\right\|_\infty$ 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 $\mathsf{D}\left(\rho_n\|\sigma_n\right)$ diverges. Hence, the limit $\sqrt{n}\mathsf{D}\left(\rho_n\|\sigma_n\right)$ does not exist and Theorem 4 does not hold.

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. The relevant multi-hypothesis testing problem can be formulated as[2]

$$H_i : \epsilon_i < \mathsf{D}\left(\rho_i\|\sigma_i\right) \leq \epsilon_{i+1}, \tag{17}$$

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 $\mathcal{T}_n =$

[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.

$\{M_n(i), i \in \mathcal{I}\}$ 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 Appendix B 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 $t_{i,n} < \mathsf{D}(\hat{\rho}_n \| \hat{\sigma}_n) \leq t_{i+1,n}$, 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 (17).

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:

$$R_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad R_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

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

$$\rho = \frac{1}{d}\left(I + \sum_{j=1}^{d^2-1} s_j(\rho)\gamma_j\right), \tag{18}$$

with $s_j(\rho) = \mathrm{Tr}[\rho\gamma_j]$. Note that γ_j, for $1 \leq j \leq d^2 - 1$, are traceless and have eigenvalues ± 1. Moreover, any operator of the form (18), 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

$$\hat{\rho}_n = \mathbb{1}_{\|\hat{s}^{(n)}(\rho)\|_2 \leq 1} \hat{S}_n(\rho) + \frac{\mathbb{1}_{\|\hat{s}^{(n)}(\rho)\|_2 > 1}}{\left\|\hat{s}^{(n)}(\rho)\right\|_2} \hat{S}_n(\rho), \tag{19a}$$

and

$$\hat{\sigma}_n = \frac{I}{nd} + \left(1 - \frac{1}{n}\right)\left(\mathbb{1}_{\|\hat{s}^{(n)}(\sigma)\|_2 \leq 1} \hat{S}_n(\sigma) + \frac{\mathbb{1}_{\|\hat{s}^{(n)}(\sigma)\|_2 > 1}}{\left\|\hat{s}^{(n)}(\sigma)\right\|_2} \hat{S}_n(\sigma)\right), \tag{19b}$$

respectively, where

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

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

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

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*

$$\sqrt{n}\big(\mathsf{D}\left(\hat{\rho}_n \| \sigma\right) - \mathsf{D}\left(\rho \| \sigma\right)\big) \xrightarrow{w} W_1 \sim N\big(0, v_1^2(\rho,\sigma)\big), \tag{20a}$$

$$\sqrt{n}\big(\mathsf{D}\left(\hat{\rho}_n \| \hat{\sigma}_n\right) - \mathsf{D}\left(\rho \| \sigma\right)\big) \xrightarrow{w} W_2 \sim N\big(0, v_2^2(\rho,\sigma)\big), \tag{20b}$$

where

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

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

The proof of Proposition 1 is given in Section VI-G and follows by an application of Theorem 1. The main ingredient of the proof is to show that $\left(\sqrt{n}(\hat{\rho}_n - \rho), \sqrt{n}(\hat{\sigma}_n - \sigma)\right) \xrightarrow{w} (L_\rho, L_\sigma)$, where $L_\rho := \sum_{j=1}^{d^2-1} \gamma_j Z_j(\rho)$ and

$Z_j(\rho) \sim N\big(0, 4s_j^+(\rho)s_j^-(\rho)/d^2\big)$. The claim then follows from (7) by noting that all relevant regularity conditions are satisfied.

B. *Performance Guarantees for 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 (17). 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 (17) 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 (17) for $\sigma_i = \sigma$ therein. Let $\hat{D}_n = \mathsf{D}\left(\hat{\rho}_n \| \sigma\right)$ with $\hat{\rho}_n$ given in (19a). Then, the test statistic*

$$T_n = \sum_{i \in \mathcal{I}} i \mathbb{1}_{\hat{D}_n \in \mathcal{L}_{i,n}(c)} \text{ with } \mathcal{L}_{i,n}(c) := (\epsilon_i + cn^{-1/2}, \epsilon_{i+1} + cn^{-1/2}),$$

asymptotically achieves a level τ provided

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

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

The proof of Proposition 2 (see Section VI-H) follows by an application of Proposition 1 and Portmanteau theorem [90, 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 ρ_i, σ with $i \in \mathcal{I}$.

Remark 2 (Growing number of hypotheses) *An inspection of the proof of Proposition 2 reveals that it 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 important quantum divergences such as quantum relative entropy, its Rényi generalizations, and measured 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. These functions simplify in the commuting case and coincide with previously known expressions in the classical case. 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. Any progress in this direction would be extremely useful to understand the non-asymptotic behavior of such estimators. Also, accounting for the case of infinite dimensional quantum systems would be a natural extension. In Section III-C, we treated the case of quantum relative entropy for density operators on a separable Hilbert space. We believe that our approach can be extended to more general scenarios with appropriate technical modifications to ensure uniform integrability of terms that appear in a Taylor's expansion. However, this is beyond the scope of the current article. Of interest further is to understand the asymptotic and non-asymptotic behavior of other classes of estimators such as those based on variational methods, for which the techniques used here may not be directly applicable. Lastly, it would also be beneficial to study the statistical behaviour of other quantum divergences not considered here such as quantum χ^2 divergence [91] and geometric Rényi divergence [92], [93].

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).

VI. Proofs

A. Technical Concepts and Auxiliary Results

We need the concept of Bochner-integrability [94] in the proofs of our main results, which we briefly mention. Let $(\mathfrak{X}, \Sigma, \mu)$ be a measure space and $\mathfrak{B}$ be a Banach space. A function $f : \mathfrak{X} \to \mathfrak{B}$ is said to be Bochner-integrable if there exists a sequence of simple functions g_n such that $g_n \to f$, μ-a.e., and

$$\lim_{n\to\infty} \int_{\mathfrak{X}} \|f - g_n\|_{\mathfrak{B}} \, d\mu = 0,$$

where $\|\cdot\|_{\mathfrak{B}}$ denotes the Banach space norm. A Bochner-measurable function f is Bochner-integrable iff $\int_{\mathfrak{X}} \|f\|_{\mathfrak{B}} \, d\mu < \infty$. Moreover, if f is Bochner-integrable, then

$$\left\| \int_{\mathfrak{X}} f d\mu \right\|_{\mathfrak{B}} \leq \int_{\mathfrak{X}} \|f\|_{\mathfrak{B}} \, d\mu.$$

The following technical lemma will be handy for our purposes. Its proof is given in Appendix A.

Lemma 1 (Properties of linear operators) *The following hold:*

(i) Suppose A and B be linear operators such that $A \ll B$. Then, $AB \ll A \ll B$.

(ii) Let A, B be square linear operators and P be a projection (i.e., a square operator such that $P^2 = P$) such that $A \ll P, B$. Then, AB, ABP, APB and $APBP$ have the same eigenvalues. In particular, $\mathrm{Tr}\,[AB] = \mathrm{Tr}\,[ABP] = \mathrm{Tr}\,[APB] = \mathrm{Tr}\,[APBP]$.

(iii) Suppose $B \leq A \leq C$. Then, $\|A\|_1 \leq \|B\|_1 + \|C\|_1$.

We next proceed with the proofs of the main results.

B. Proof of Theorem 1

For $A_1, A_2 > 0$, let

$$f(A_1, A_2) = A_1(\log A_1 - \log A_2).$$

Consider the integral representation

$$\begin{aligned}\log A &= \int_0^\infty \left(\frac{1}{(\tau+1)I} - \frac{1}{\tau I + A}\right) d\tau \\ &= \int_0^1 \left(\frac{1}{(\tau+1)I} - \frac{1}{\tau I + A}\right) d\tau + \int_1^\infty \left(\frac{1}{(\tau+1)I} - \frac{1}{\tau I + A}\right) d\tau, \end{aligned} \tag{21}$$

for $A > 0$. We have $D\left[(\tau I + A)^{-1}\right](H) = -(\tau I + A)^{-1} H (\tau I + A)^{-1}$ by (2) and

$$D[\log A](H) = \int_0^\infty (\tau I + A)^{-1} H (\tau I + A)^{-1} d\tau.$$

Applying the above, we obtain via the chain rule and product rule for Fréchet derivatives that

$$\begin{aligned}
D^{(1,0)}[f(A_1, A_2)](H) &= A_1 \int_0^\infty (\tau I + A_1)^{-1} H (\tau I + A_1)^{-1} d\tau + H (\log A_1 - \log A_2), \\
D^{(0,1)}[f(A_1, A_2)](H) &= -A_1 \int_0^\infty (\tau I + A_2)^{-1} H (\tau I + A_2)^{-1} d\tau, \\
D^{(2,0)}[f(A_1, A_2)](H_1, H_2) &= H_1 \int_0^\infty (\tau I + A_1)^{-1} H_2 (\tau I + A_1)^{-1} d\tau \\
&\quad + H_2 \int_0^\infty (\tau I + A_1)^{-1} H_1 (\tau I + A_1)^{-1} d\tau \\
&\quad - A_1 \int_0^\infty (\tau I + A_1)^{-1} H_1 (\tau I + A_1)^{-1} H_2 (\tau I + A_1)^{-1} d\tau \\
&\quad - A_1 \int_0^\infty (\tau I + A_1)^{-1} H_2 (\tau I + A_1)^{-1} H_1 (\tau I + A_1)^{-1} d\tau, \\
D^{(0,2)}[f(A_1, A_2)](H_1, H_2) &= A_1 \int_0^\infty (\tau I + A_2)^{-1} H_1 (\tau I + A_2)^{-1} H_2 (\tau I + A_2)^{-1} d\tau \\
&\quad + A_1 \int_0^\infty (\tau I + A_2)^{-1} H_2 (\tau I + A_2)^{-1} H_1 (\tau I + A_2)^{-1} d\tau, \\
D^{(1,1_+)}[f(A_1, A_2)](H_1, H_2) &= -H_1 \int_0^\infty (\tau I + A_2)^{-1} H_2 (\tau I + A_2)^{-1} d\tau, \\
D^{(1_+,1)}[f(A_1, A_2)](H_1, H_2) &= -H_2 \int_0^\infty (\tau I + A_2)^{-1} H_1 (\tau I + A_2)^{-1} d\tau,
\end{aligned}$$

where the notation $D^{(1,1_+)}$ means that the order of differentiation is with respect to first coordinate followed by the second, and vice versa for $D^{(1_+,1)}$.

Note that $f(A_1, A_2)$ is continuously twice differentiable function from $\mathcal{P}_d^+ \times \mathcal{P}_d^+$ to $\mathcal{L}(\mathbb{H}_d)$. Hence, applying the operator version of multivariate Taylor's theorem (see e.g. [83]), we obtain for $A_1, A_2,\ B_1, B_2 > 0$ that

$$\begin{aligned}f(B_1, B_2) &= f(A_1, A_2) + D^{(1,0)}[f(A_1, A_2)](B_1 - A_1) + D^{(0,1)}[f(A_1, A_2)](B_2 - A_2) \\ &\quad + \int_0^1 (1-t) D^{(2,0)}[f((1-t)A_1 + tB_1, (1-t)A_2 + tB_2)](B_1 - A_1, B_1 - A_1)\, dt\end{aligned}$$

$$
\begin{aligned}
&+ \int_0^1 (1-t) D^{(0,2)}[f((1-t)A_1 + tB_1, (1-t)A_2 + tB_2)](B_2 - A_2, B_2 - A_2)\, dt \\
&+ \int_0^1 (1-t) D^{(1,1_+)}[f((1-t)A_1 + tB_1, (1-t)A_2 + tB_2)](B_1 - A_1, B_2 - A_2)\, dt \\
&+ \int_0^1 (1-t) D^{(1_+,1)}[f((1-t)A_1 + tB_1, (1-t)A_2 + tB_2)](B_2 - A_2, B_1 - A_1)\, dt. \qquad (22)
\end{aligned}
$$

Two-sample null: Assume first that $\rho_n, \sigma_n, \rho > 0$. Setting $B_1 = \rho_n$, $B_2 = \sigma_n$ and $A_1 = A_2 = \rho$ in (22), and defining

$$
v(\rho_n, \rho, \tau, t) := \big(\tau I + (1-t)\rho + t\rho_n\big)^{-1},
$$

we have

$$
\begin{aligned}
&f(\rho_n, \sigma_n) \\
&= \rho \int_0^\infty (\tau I + \rho)^{-1} (\rho_n - \rho) (\tau I + \rho)^{-1} d\tau - \rho \int_0^\infty (\tau I + \rho)^{-1} (\sigma_n - \rho) (\tau I + \rho)^{-1} d\tau \\
&\quad + 2 \int_0^1 (1-t)(\rho_n - \rho) \int_0^\infty v(\rho_n, \rho, \tau, t)(\rho_n - \rho) v(\rho_n, \rho, \tau, t) d\tau dt \\
&\quad - 2 \int_0^1 (1-t) \left((1-t)\rho + t\rho_n\right) \int_0^\infty v(\rho_n, \rho, \tau, t)(\rho_n - \rho) v(\rho_n, \rho, \tau, t)(\rho_n - \rho) v(\rho_n, \rho, \tau, t) d\tau dt \\
&\quad + 2 \int_0^1 (1-t)\big((1-t)\rho + t\rho_n\big) \int_0^\infty v(\sigma_n, \rho, \tau, t)(\sigma_n - \rho) v(\sigma_n, \rho, \tau, t)(\sigma_n - \rho) v(\sigma_n, \rho, \tau, t) d\tau dt \\
&\quad - 2 \int_0^1 (1-t)(\rho_n - \rho) \int_0^\infty v(\sigma_n, \rho, \tau, t)(\sigma_n - \rho) v(\sigma_n, \rho, \tau, t) d\tau dt. \qquad (23)
\end{aligned}
$$

To extend the validity of the above equation to $0 \leq \rho_n \ll \sigma_n \geq 0$, we consider $f\big(\tilde{\rho}_n(\epsilon), \tilde{\sigma}_n(\epsilon)\big)$ with

$$
\begin{aligned}
\tilde{\rho}_n(\epsilon) &:= \rho_n + \epsilon \pi_d, \\
\tilde{\sigma}_n(\epsilon) &:= \sigma_n + \epsilon \pi_d,
\end{aligned}
$$

for $\epsilon > 0$, and take limit $\epsilon \to 0$. Then, the desired expression follows since $f\big(\tilde{\rho}_n(\epsilon), \tilde{\sigma}_n(\epsilon)\big)$ is continuous in ϵ for $\rho_n \ll \sigma_n$, and uniform integrability conditions which allows for interchange of limits and integral hold for $\epsilon \leq 1$. We illustrate the latter condition for some of the terms above. For the first term, by using Hölder's inequality for Schatten-norms and $\|A\|_p \geq \|A\|_q$ for any linear operator A and $1 \leq p \leq q \leq \infty$, we have

$$
\begin{aligned}
\left\| \int_0^\infty (\tau I + \rho)^{-1} (\tilde{\rho}_n(\epsilon) - \rho) (\tau I + \rho)^{-1} d\tau \right\|_1 &\leq \|\tilde{\rho}_n(\epsilon) - \rho\|_1 \int_0^\infty \left\| (\tau I + \rho)^{-1} \right\|_\infty^2 d\tau \\
&\lesssim_\rho \|\rho_n - \rho\|_1 + 1, \qquad (24a)
\end{aligned}
$$

$$
\begin{aligned}
&\left\| (1-t)(\tilde{\rho}_n(\epsilon) - \rho) \int_0^\infty v\big(\tilde{\rho}_n(\epsilon), \rho, \tau, t\big)(\tilde{\rho}_n(\epsilon) - \rho) v\big(\tilde{\rho}_n(\epsilon), \rho, \tau, t\big) d\tau \right\|_1 \\
&\quad \leq (1-t) \|\tilde{\rho}_n(\epsilon) - \rho\|_1^2 \int_0^\infty \left\| (\tau I + (1-t)\rho)^{-1} \right\|_\infty^2 d\tau \qquad (24b) \\
&\quad \lesssim_\rho (1-t) \|\tilde{\rho}_n(\epsilon) - \rho\|_1^2 (1-t)^{-1} \\
&\quad \lesssim \|\rho_n - \rho\|_1^2 + 1, \qquad (24c)
\end{aligned}
$$

$$
\big\| (1-t)\big((1-t)\rho + t\tilde{\rho}_n(\epsilon)\big) \int_0^\infty v\big(\tilde{\rho}_n(\epsilon), \rho, \tau, t\big)(\tilde{\rho}_n(\epsilon) - \rho) v\big(\tilde{\rho}_n(\epsilon), \rho, \tau, t\big)
$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad (\tilde{\rho}_n(\epsilon)-\rho)v\big(\tilde{\rho}_n(\epsilon),\rho,\tau,t\big)d\tau\big\|_1\\
&\le (1-t)\left\|\tilde{\rho}_n(\epsilon)-\rho\right\|_1^2\int_0^\infty \left\|(\tau I+(1-t)\rho)^{-1}\right\|_\infty^2 d\tau\\
&\lesssim_\rho \left\|\rho_n-\rho\right\|_1^2+1. \qquad (24d)
\end{aligned}
$$

Similar bounds hold for the remaining terms via analogous steps.

Let $g_n := r_n^2 \mathsf{D}\left(\rho_n\|\sigma_n\right)$. Multiplying by r_n^2 and taking trace in (23), we obtain

$$
\begin{aligned}
g_n = {} & 2\int_0^1 (1-t)\,\mathrm{Tr}\left[r_n(\rho_n-\rho)\int_0^\infty v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\right]dt\\
&-2\int_0^1(1-t)\,\mathrm{Tr}\Bigg[\left((1-t)\rho+t\rho_n\right)\int_0^\infty v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)\\
&\qquad\qquad\qquad v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\Bigg]dt\\
&+2\int_0^1(1-t)\,\mathrm{Tr}\Bigg[\left((1-t)\rho+t\rho_n\right)\int_0^\infty v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)\\
&\qquad\qquad\qquad v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)d\tau\Bigg]dt\\
&-2\int_0^1(1-t)\,\mathrm{Tr}\left[r_n(\rho_n-\rho)\int_0^\infty v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)d\tau\right]dt, \qquad (25)
\end{aligned}
$$

where we used that the first two terms above vanish. To see this for the first term, note that

$$
\begin{aligned}
&\int_0^\infty \left\|\rho\left(\tau I+\rho\right)^{-1}\left(\rho_n-\rho\right)\left(\tau I+\rho\right)^{-1}\right\|_1 d\tau\\
&=\int_0^1 \left\|\rho\left(\tau I+\rho\right)^{-1}\left(\rho_n-\rho\right)\left(\tau I+\rho\right)^{-1}\right\|_1 d\tau+\int_1^\infty \left\|\rho\left(\tau I+\rho\right)^{-1}\left(\rho_n-\rho\right)\left(\tau I+\rho\right)^{-1}\right\|_1 d\tau\\
&\le \left\|(\rho_n-\rho)\rho^{-1}\right\|_1+\left\|\rho(\rho_n-\rho)\right\|_1\int_1^\infty \tau^{-2}d\tau\\
&\lesssim \left\|\rho^{-1}\right\|_\infty+\int_1^\infty \tau^{-2}d\tau\\
&<\infty. \qquad (26)
\end{aligned}
$$

Similarly,

$$
\int_0^\infty \left\|\rho(\rho_n-\rho)\left(\tau I+\rho\right)^{-2}\right\|_1 d\tau<\infty. \qquad (27)
$$

Hence, $\rho\left(\tau I+\rho\right)^{-1}\left(\rho_n-\rho\right)\left(\tau I+\rho\right)^{-1}$ and $\rho(\rho_n-\rho)\left(\tau I+\rho\right)^{-2}$ are Bochner-integrable functions (with respect to Lebesgue measure on $(0,\infty)$). Then, we have

$$
\begin{aligned}
\mathrm{Tr}\left[\int_0^\infty \rho\left(\tau I+\rho\right)^{-1}\left(\rho_n-\rho\right)\left(\tau I+\rho\right)^{-1}d\tau\right] &\overset{(a)}{=} \int_0^\infty \mathrm{Tr}\left[\rho\left(\tau I+\rho\right)^{-1}\left(\rho_n-\rho\right)\left(\tau I+\rho\right)^{-1}\right]d\tau\\
&\overset{(b)}{=} \int_0^\infty \mathrm{Tr}\left[\rho(\rho_n-\rho)\left(\tau I+\rho\right)^{-2}\right]d\tau\\
&\overset{(c)}{=} \mathrm{Tr}\left[\rho(\rho_n-\rho)\int_0^\infty \left(\tau I+\rho\right)^{-2}d\tau\right]\\
&= \mathrm{Tr}\left[\rho(\rho_n-\rho)\rho^{-1}\right]\\
&= \mathrm{Tr}\left[\rho_n-\rho\right]\overset{(d)}{=}0, \qquad (28)
\end{aligned}
$$

where

(a) follows from (26) by the fact that $\mathrm{Tr}[\cdot]$ is continuous linear (hence bounded) functional on the normed space $\mathcal{L}(\mathbb{H}_d)$ (with $\|\cdot\|_1$ norm);

(b) uses that $[\rho, (\tau I + \rho)^{-1}] = 0$ and the cylic property of $\mathrm{Tr}[\cdot]$;

(c) follows via the same argument as in (a) using (27);

(d) is because $\mathrm{Tr}[\rho_n] = \mathrm{Tr}[\rho] = 1$ on account of ρ_n and ρ being density operators.

Likewise, it can be shown that

$$\mathrm{Tr}\left[\int_0^\infty \rho\,(\tau I+\rho)^{-1}(\sigma_n-\rho)(\tau I+\rho)^{-1}\,d\tau\right]=0. \tag{29}$$

We next analyze the limit of the expression in (25). To prove the weak convergence of g_n to its desired limit, it suffices to show that for every subsequence of $(g_n)_{n\in\mathbb{N}}$, there exists a further subsequence along which the sequence converges to a unique weak limit. We refer to this as the *subsequence argument*. Let

$$p_n(r_n) := (1-t)r_n(\rho_n-\rho)\int_0^\infty v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau, \tag{30a}$$

$$q_n(r_n) := (1-t)\big((1-t)\rho+t\rho_n\big)\int_0^\infty v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau, \tag{30b}$$

$$\tilde{p}_n(r_n) := (1-t)\big((1-t)\rho+t\rho_n\big)\int_0^\infty v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)d\tau, \tag{30c}$$

$$\tilde{q}_n(r_n) := (1-t)r_n(\rho_n-\rho)\int_0^\infty v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)d\tau. \tag{30d}$$

Then, the following bounds hold:

$$\|p_n(r_n)\|_1 \le \|r_n(\rho_n-\rho)\|_1^2\int_0^\infty \|v(\sigma_n,\rho,\tau,t)\|_\infty^2\,d\tau, \tag{31a}$$

$$\|q_n(r_n)\|_1 \le \|r_n(\rho_n-\rho)\|_1^2\int_0^\infty \|v(\sigma_n,\rho,\tau,t)\|_\infty^3\,d\tau, \tag{31b}$$

$$\|\tilde{p}_n(r_n)\|_1 \le \|r_n(\sigma_n-\rho)\|_1^2\int_0^\infty \|v(\sigma_n,\rho,\tau,t)\|_\infty^3\,d\tau, \tag{31c}$$

$$\|\tilde{q}_n(r_n)\|_1 \le \|r_n(\rho_n-\rho)\|_1\|r_n(\sigma_n-\rho)\|_1\int_0^\infty \|v(\sigma_n,\rho,\tau,t)\|_\infty^2\,d\tau. \tag{31d}$$

To show the aforementioned claim of unique weak limit, for any subsequence $(n_k)_{k\in\mathbb{N}}$, consider a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big(r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\big)\to(L_1,L_2)$ in $\|\cdot\|_1$ almost surely (a.s.). This is possible by Skorokhods representation theorem (see e.g. [95]) due to separability of $\mathcal{L}(\mathbb{H}_d)$. Then, since $\sigma_{n_{k_j}}\to\rho$, we have that $(1-t)\rho+t\sigma_{n_{k_j}}\ge c\rho$ for a constant $0<c<1$ (which depends on the realization $\sigma_{n_{k_j}}$) and sufficiently large j. This implies that the integrals in (31) are finite. For instance, the integral in (31a) is $O(1)$ as τ approaches zero and $O(\tau^{-2})$ as τ tends to ∞. Hence, we obtain

$$\left\|p_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right\|_1^2, \tag{32a}$$

$$\left\|q_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right\|_1^2, \tag{32b}$$

$$\left\|\tilde{p}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\right\|_1^2, \tag{32c}$$

$$\left\|\tilde{q}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right\|_1\left\|r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\right\|_1. \tag{32d}$$

Next, recall that $\left((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\right) \rightarrow (L_1, L_2)$ (a.s.) in the space of linear operators with bounded trace norm implies that $\left(r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right)_{j\in\mathbb{N}}$ and $\left(r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\right)_{j\in\mathbb{N}}$ are uniformly Bochner-integrable sequences (a.s.). This combined with (31) and (32) then shows that $\left(p_{n_{k_j}}(r_{n_{k_j}})\right)_{j\in\mathbb{N}}$, $\left(q_{n_{k_j}}(r_{n_{k_j}})\right)_{j\in\mathbb{N}}$, $(\tilde{p}_{n_{k_j}})_{j\in\mathbb{N}}$ and $(\tilde{q}_{n_{k_j}})_{j\in\mathbb{N}}$ are uniformly Bochner-integrable sequences. Taking limits $j\rightarrow\infty$, interchanging limits and integral in (25), and noting that $\rho_{n_{k_j}}\rightarrow\rho$, $\sigma_{n_{k_j}}\rightarrow\rho$ a.s. and $\|\cdot\|_1$, yields

$$
\begin{aligned}
&g_{n_{k_j}}\\
&\rightarrow \mathrm{Tr}\left[L_1\int_0^\infty (\tau I+\rho)^{-1}L_1(\tau I+\rho)^{-1}d\tau-\rho\int_0^\infty(\tau I+\rho)^{-1}L_1(\tau I+\rho)^{-1}L_1(\tau I+\rho)^{-1}d\tau\right]\\
&\quad+\mathrm{Tr}\left[\rho\int_0^\infty(\tau I+\rho)^{-1}L_2(\tau I+\rho)^{-1}L_2(\tau I+\rho)^{-1}d\tau-L_1\int_0^\infty(\tau I+\rho)^{-1}L_2(\tau I+\rho)^{-1}d\tau\right]\\
&\quad=\mathrm{Tr}\left[L_1D[\log\rho](L_1-L_2)\right]+\mathrm{Tr}\left[\frac{\rho}{2}D^2[\log\rho](L_1-L_2,L_1-L_2)\right].
\end{aligned}
$$

Hence, every subsequence $(g_{n_k})_{k\in\mathbb{N}}$ has a further subsequence $(g_{n_{k_j}})_{j\in\mathbb{N}}$ with the same unique limit which implies (8).

If $[\rho_n,\rho]=0$, then $[r_n(\rho_n-\rho),\rho]=0$. Consider a subsequence $r_{n_j}(\rho_{n_j}-\rho)\rightarrow L_1$ a.s. Since the commutator is a continuous linear functional of its individual arguments, we have

$$
[L_1,\rho]=\lim_{j\rightarrow\infty}[r_{n_j}(\rho_{n_j}-\rho),\rho]=0,\ \text{a.s.} \tag{33}
$$

Hence, L_1 and ρ commutes. The proof of $[L_1,L_2]=[L_2,\rho]=0$ under the conditions $[\rho_n,\rho]=[\sigma_n,\rho]=[\sigma_n,\rho_n]=0$ follow similarly. Under this scenario, the expression in the RHS of (8) simplifies to

$$
\begin{aligned}
&\mathrm{Tr}\left[L_1^2\int_0^\infty(\tau I+\rho)^{-2}d\tau\right]-\mathrm{Tr}\left[L_1^2\rho\int_0^\infty(\tau I+\rho)^{-3}d\tau\right]+\mathrm{Tr}\left[L_2^2\rho\int_0^\infty(\tau I+\rho)^{-3}d\tau\right]\\
&\quad-\mathrm{Tr}\left[L_1L_2\int_0^\infty(\tau I+\rho)^{-2}d\tau\right]\\
&=\frac{1}{2}\mathrm{Tr}\left[L_1^2\rho^{-1}\right]+\frac{1}{2}\mathrm{Tr}\left[L_2^2\rho^{-1}\right]-\mathrm{Tr}\left[L_1L_2\rho^{-1}\right]\\
&=\frac{1}{2}\mathrm{Tr}\left[(L_1-L_2)^2\rho^{-1}\right].
\end{aligned}
$$

Finally, consider the case $\rho\geq 0$. Note that the left-hand side (LHS) and RHS of (8) is invariant to restricting the space to the support of ρ. To see this, let $U=\sum_{i=1}^r|i\rangle\langle u_i|$, where r is the rank of ρ and $|u_i\rangle$ are the eigenvectors corresponding to the non-zero eigenvalues of ρ and $\{|i\rangle\}_{i=1}^r$ denotes the computational basis of $\mathbb{H}^r$. Setting $\tilde{\rho}=U\rho U^\dagger$, $\tilde{\rho}_n=U\rho_nU^\dagger$, $\tilde{\sigma}_n=U\sigma_nU^\dagger$, $P:=U^\dagger U$, and noting that $\rho_n\ll\sigma_n\ll\rho\ll\gg P$, it follows from Lemma 1(ii) that

$$
\begin{aligned}
\mathsf{D}(\rho_n\|\sigma_n)&=\mathrm{Tr}\left[\rho_n\log\rho_n\right]-\mathrm{Tr}\left[\rho_n\log\sigma_n\right]\\
&=\mathrm{Tr}\left[\rho_nP\log\rho_nP\right]-\mathrm{Tr}\left[\rho_nP\log\sigma_nP\right]\\
&=\mathrm{Tr}\left[U\rho_nU^\dagger U\log\rho_nU^\dagger\right]-\mathrm{Tr}\left[U\rho_nU^\dagger U\log\sigma_nU^\dagger\right]\\
&=\mathrm{Tr}\left[\tilde{\rho}_n\log\tilde{\rho}_n\right]-\mathrm{Tr}\left[\tilde{\rho}_n\log\tilde{\sigma}_n\right].
\end{aligned}
$$

Note that $\tilde{\rho} > 0$ and $\tilde{\rho}_n, \tilde{\sigma}_n \geq 0$ are density operators. Next, by Portmanteau's theorem, the support of L_1 and L_2 is contained in that of ρ since $\rho_n - \rho \ll \rho, \sigma_n - \rho \ll \rho$ due to $\rho_n, \sigma_n \ll \rho$. Hence, similar to the above, the RHS is also invariant to replacing all operators by their sandwiched versions obtained by left multiplying with U and right multiplying with $U^\dagger$. Finally, observe that $\big(r_n(\rho_n - \rho), r_n(\sigma_n - \rho)\big) \xrightarrow{w} (L_1, L_2)$ implies $\big(r_n(\tilde{\rho}_n - \tilde{\rho}), r_n(\tilde{\sigma}_n - \tilde{\rho})\big) \xrightarrow{w} (UL_1U^\dagger, UL_2U^\dagger)$ by an application of Slutsky's theorem [95]. Hence, the previous proof applies and the claim follows.

Two-sample alternative: Assume first that $\rho_n, \sigma_n, \rho, \sigma > 0$. Setting $B_1 = \rho_n$, $B_2 = \sigma_n$, $A_1 = \rho$ and $A_2 = \sigma \neq \rho$ in (22), we have

$$
\begin{aligned}
&f(\rho_n, \sigma_n) \\
&= f(\rho, \sigma) + \rho \int_0^\infty (\tau I + \rho)^{-1} (\rho_n - \rho) (\tau I + \rho)^{-1} d\tau - \rho \int_0^\infty (\tau I + \sigma)^{-1} (\sigma_n - \sigma)(\tau I + \sigma)^{-1} d\tau \\
&\quad + (\rho_n - \rho)(\log \rho - \log \sigma) + 2 \int_0^1 (1-t)(\rho_n - \rho) \int_0^\infty v(\rho_n, \rho, \tau, t)(\rho_n - \rho) v(\rho_n, \rho, \tau, t) d\tau dt \\
&\quad - 2 \int_0^1 (1-t)\left((1-t)\rho + t\rho_n\right) \int_0^\infty v(\rho_n, \rho, \tau, t)(\rho_n - \rho) v(\rho_n, \rho, \tau, t)(\rho_n - \rho) v(\rho_n, \rho, \tau, t) d\tau dt \\
&\quad + 2 \int_0^1 (1-t)\big((1-t)\rho + t\rho_n\big) \int_0^\infty v(\sigma_n, \sigma, \tau, t)(\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t)(\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) d\tau dt \\
&\quad - 2 \int_0^1 (1-t)(\rho_n - \rho) \int_0^\infty v(\sigma_n, \sigma, \tau, t)(\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) d\tau dt.
\end{aligned} \tag{34}
$$

This equation extends to $0 \leq \rho_n \ll \sigma_n \geq 0$ via similar arguments in Part (i). Multiplying by r_n and taking trace, we obtain

$$
\begin{aligned}
g_n &:= r_n\big(\mathsf{D}(\rho_n \| \sigma_n) - \mathsf{D}(\rho \| \sigma)\big) \\
&= \mathrm{Tr}\left[r_n(\rho_n - \rho)(\log \rho - \log \sigma) - \rho \int_0^\infty (\tau I + \sigma)^{-1} r_n(\sigma_n - \sigma)(\tau I + \sigma)^{-1} d\tau\right] \\
&\quad + \mathrm{Tr}\left[2 \int_0^1 (1-t) r_n^{\frac{1}{2}} (\rho_n - \rho) \int_0^\infty v(\rho_n, \rho, \tau, t) r_n^{\frac{1}{2}} (\rho_n - \rho) v(\rho_n, \rho, \tau, t) d\tau\right] dt \\
&\quad - 2 \int_0^1 (1-t)\, \mathrm{Tr}\Bigg[\left((1-t)\rho + t\rho_n\right) \int_0^\infty v(\rho_n, \rho, \tau, t) r_n^{\frac{1}{2}} (\rho_n - \rho) v(\rho_n, \rho, \tau, t) r_n^{\frac{1}{2}} (\rho_n - \rho) \\
&\qquad\qquad v(\rho_n, \rho, \tau, t) d\tau\Bigg] dt \\
&\quad + 2 \int_0^1 (1-t)\, \mathrm{Tr}\Bigg[\big((1-t)\rho + t\rho_n\big) \int_0^\infty v(\sigma_n, \sigma, \tau, t) r_n^{\frac{1}{2}} (\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) r_n^{\frac{1}{2}} (\sigma_n - \sigma) \\
&\qquad\qquad v(\sigma_n, \sigma, \tau, t) d\tau\Bigg] dt \\
&\quad - 2 \int_0^1 (1-t)\, \mathrm{Tr}\left[r_n^{\frac{1}{2}} (\rho_n - \rho) \int_0^\infty v(\sigma_n, \sigma, \tau, t) r_n^{\frac{1}{2}} (\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) d\tau\right] dt,
\end{aligned} \tag{35}
$$

where we used (28).

Let

$$
\bar{p}_n(r_n) := (1-t)\big((1-t)\rho + t\rho_n\big) \int_0^\infty v(\sigma_n, \sigma, \tau, t) r_n^{\frac{1}{2}} (\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) r_n^{\frac{1}{2}} (\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) d\tau,
$$

$$\bar{q}_n(r_n) := (1-t) r_n^{\frac{1}{2}} (\rho_n - \rho) \int_0^\infty v(\sigma_n, \sigma, \tau, t) r_n^{\frac{1}{2}} (\sigma_n - \sigma) v(\sigma_n, \sigma, \tau, t) d\tau. \tag{36}$$

Then, via steps akin to (31), we have

$$\|\bar{p}_n(r_n)\|_1 \leq \left\| r_n^{\frac{1}{2}} (\sigma_n - \sigma) \right\|_1^2 \int_0^\infty \|v(\sigma_n, \sigma, \tau, t)\|_\infty^3 \, d\tau,$$

$$\|\bar{q}_n(r_n)\|_1 \leq \left\| r_n^{\frac{1}{2}} (\rho_n - \rho) \right\|_1 \left\| r_n^{\frac{1}{2}} (\sigma_n - \sigma) \right\|_1 \int_0^\infty \|v(\sigma_n, \sigma, \tau, t)\|_\infty^2 \, d\tau.$$

As in Part (i), for any subsequence $(n_k)_{k\in\mathbb{N}}$, consider a further subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}} - \rho), r_{n_{k_j}}(\sigma_{n_{k_j}} - \sigma)\big) \to (L_1, L_2)$ in $\|\cdot\|_1$ a.s. Noting that there exists a constant $0 < c < 1$ such that $(1-t)\rho + t\sigma_{n_{k_j}} \geq c\sigma$ for sufficiently large j, we have

$$\left\| \bar{p}_{n_{k_j}}(r_{n_{k_j}}) \right\|_1 \lesssim_\sigma \left\| r_{n_{k_j}}^{\frac{1}{2}} (\sigma_{n_{k_j}} - \sigma) \right\|_1^2,$$

$$\left\| \bar{q}_{n_{k_j}}(r_{n_{k_j}}) \right\|_1 \lesssim_\sigma \left\| r_{n_{k_j}}^{\frac{1}{2}} (\rho_{n_{k_j}} - \rho) \right\|_1 \left\| r_{n_{k_j}}^{\frac{1}{2}} (\sigma_{n_{k_j}} - \sigma) \right\|_1.$$

The above equations and (32) subsequently implies that $\big(p_{n_{k_j}}(r_{n_{k_j}})\big)_{j\in\mathbb{N}}$, $\big(q_{n_{k_j}}(r_{n_{k_j}})\big)_{j\in\mathbb{N}}$, $\big(\bar{p}_{n_{k_j}}(r_{n_{k_j}})\big)_{j\in\mathbb{N}}$ and $\big(\bar{q}_{n_{k_j}}(r_{n_{k_j}})\big)_{j\in\mathbb{N}}$ are uniformly Bochner-integrable. Moreover, $\big((r_{n_{k_j}}^{1/2}(\rho_{n_{k_j}} - \rho), r_{n_{k_j}}^{1/2}(\sigma_{n_{k_j}} - \sigma)\big) \to (0, 0)$. Taking limits $j \to \infty$ and interchanging limits and integral yields

$$\begin{aligned} g_{n_{k_j}} &\to \mathrm{Tr}\left[L_1 (\log\rho - \log\sigma) - \rho \int_0^\infty (\tau I + \sigma)^{-1} L_2 (\tau I + \sigma)^{-1} d\tau \right] \\ &= \mathrm{Tr}\left[L_1 (\log\rho - \log\sigma) \right] - \mathrm{Tr}\left[\rho D[\log\sigma](L_2) \right]. \end{aligned}$$

This implies (9) via the subsequence argument mentioned in Part (i).

The claim that $[L_1, L_2] = [L_1, \rho] = [L_1, \sigma] = [L_2, \rho] = [L_2, \sigma] = 0$ provided $[\rho_n, \rho] = [\sigma_n, \sigma] = [\rho, \sigma] = [\sigma_n, \rho_n] = [\sigma_n, \rho] = [\sigma, \rho_n] = 0$ is similar to (33). In this case, the above limit simplifies as

$$\mathrm{Tr}\left[\rho \int_0^\infty (\tau I + \sigma)^{-1} L_2 (\tau I + \sigma)^{-1} d\tau \right] = \mathrm{Tr}\left[L_2 \rho \sigma^{-1} \right].$$

Finally, consider the case $0 \leq \rho \ll \sigma \geq 0$. Let $U = \sum_{i=1}^r |i\rangle\langle u_i|$ and P be as defined in Part (i). Then, setting $\tilde{\rho} = U\rho U^\dagger$, $\tilde{\rho}_n = U\rho_n U^\dagger$, $\tilde{\sigma} = U\sigma U^\dagger$, $\tilde{\sigma}_n = U\sigma_n U^\dagger$ and noting that $\rho_n \ll \sigma_n \ll \rho \ll\gg P \ll \sigma$, we have by applying Lemma 1(ii) that

$$\begin{aligned} \mathsf{D}(\rho_n\|\sigma_n) - \mathsf{D}(\rho\|\sigma) &= \mathrm{Tr}[\rho_n \log\rho_n] - \mathrm{Tr}[\rho_n \log\sigma_n] - \mathrm{Tr}[\rho\log\rho] + \mathrm{Tr}[\rho\log\sigma] \\ &= \mathrm{Tr}[\rho_n P \log\rho_n P] - \mathrm{Tr}[\rho_n P\log\sigma_n P] - \mathrm{Tr}[\rho P\log\rho P] + \mathrm{Tr}[\rho P\log\sigma P] \\ &= \mathrm{Tr}[\tilde{\rho}_n \log\tilde{\rho}_n] - \mathrm{Tr}[\tilde{\rho}_n \log\tilde{\sigma}_n] - \mathrm{Tr}[\tilde{\rho}\log\tilde{\rho}] + \mathrm{Tr}[\tilde{\rho}\log\tilde{\sigma}]. \end{aligned}$$

Note that $\tilde{\rho} > 0$ and $\tilde{\rho}_n$ are density operators, while $\tilde{\sigma} \gg \tilde{\rho}$ and $\tilde{\sigma}_n \gg \tilde{\rho}_n$ have positive eigenvalues whose sum is upper bounded by 1 (hence, not a density operator in general). However, the last expression equals $\mathsf{D}(\tilde{\rho}_n\|\tilde{\sigma}_n) - \mathsf{D}(\tilde{\rho}\|\tilde{\sigma})$ which is well-defined. Similarly, the RHS is invariant to left multiplying all operators by U and right multiplying by $U^\dagger$. Hence, the LHS and RHS of (7) is invariant to restricting to support of ρ. The proof is completed by noting that $\big(r_n(\rho_n - \rho), r_n(\sigma_n - \sigma)\big) \xrightarrow{w} (L_1, L_2)$ implies $\big(r_n(\tilde{\rho}_n - \tilde{\rho}), r_n(\tilde{\sigma}_n - \tilde{\sigma})\big) \xrightarrow{w} (UL_1U^\dagger, UL_2U^\dagger)$, and the previous proof applies since $\tilde{\rho}, \tilde{\sigma} > 0$.

C. Proof of Theorem 2

In the following, we will assume without loss of generality that $\rho, \sigma > 0$. The proofs for extending to the general case $\rho \ll \sigma$ follows via similar arguments as given in the proof of Theorem 1 by invoking Lemma 1 and using the fact that the support of $D[\rho^\alpha](L_1)$ for $\alpha \in (0,2]$ is contained in the support of ρ since $L_1 \ll \rho$ a.s. due to $\rho_n \ll \rho$. We first prove Part (ii).

Two-sample alternative: Consider the case $\alpha \in (0,1)$, $\bar{\alpha} := 1-\alpha$, and $Q_\alpha(\rho,\sigma) := \mathrm{Tr}\left[\rho^\alpha \sigma^{\bar{\alpha}}\right]$. We will initially show that

$$\begin{aligned} r_n\big(Q_\alpha(\rho_n,\sigma_n) - Q_\alpha(\rho,\sigma)\big) &\xrightarrow{w} \mathrm{Tr}\left[\sigma^{\bar{\alpha}} D[\rho^\alpha](L_1)\right] + \mathrm{Tr}\left[\rho^\alpha D[\sigma^{\bar{\alpha}}](L_2)\right] \\ &= c_\alpha \,\mathrm{Tr}\left[\int_0^\infty \tau^\alpha (\tau I+\rho)^{-1} L_1 (\tau I+\rho)^{-1} d\tau \sigma^{\bar{\alpha}}\right] \\ &\quad + c_{\bar{\alpha}} \,\mathrm{Tr}\left[\rho^\alpha \int_0^\infty \tau^{\bar{\alpha}} (\tau I+\sigma)^{-1} L_2 (\tau I+\sigma)^{-1} d\tau\right], \end{aligned}$$

where $c_\alpha = \pi/\sin(\pi\alpha)$. Then, applying the functional delta method [95, Theorem 3.9.4] with $\phi(x) = \log x/(\alpha-1)$ at $x = Q_\alpha(\rho,\sigma)$ leads to

$$r_n\big(\bar{\mathsf{D}}_\alpha(\rho_n\|\sigma_n) - \bar{\mathsf{D}}_\alpha(\rho\|\sigma)\big) \xrightarrow{w} \frac{\mathrm{Tr}\left[\sigma^{\bar{\alpha}} D[\rho^\alpha](L_1)\right] + \mathrm{Tr}\left[\rho^\alpha D[\sigma^{1-\alpha}](L_2)\right]}{(\alpha-1)\,\mathrm{Tr}\left[\rho^\alpha\sigma^{\bar{\alpha}}\right]}, \tag{37}$$

as claimed in (12).

To show the above, we compute the Fréchet derivatives of the operator-valued function $f(A_1,A_2) = A_1^\alpha A_2^{1-\alpha}$. Using the integral representation [96, Lemma 2.8]

$$A^\alpha = c_\alpha \int_0^\infty \tau^\alpha \left(\frac{1}{\tau I} - \frac{1}{\tau I + A}\right), \;\alpha \in (0,1),$$

we have via chain and product rule for Fréchet derivatives that

$$\begin{aligned} D^{(1,0)}[f(A_1,A_2)](H) &= c_\alpha \int_0^\infty \tau^\alpha (\tau I + A_1)^{-1} H (\tau I + A_1)^{-1} d\tau A_2^{\bar{\alpha}}, \\ D^{(0,1)}[f(A_1,A_2)](H) &= c_{\bar{\alpha}} A_1^\alpha \int_0^\infty \tau^{\bar{\alpha}} (\tau I + A_2)^{-1} H (\tau I + A_2)^{-1} d\tau, \\ D^{(2,0)}[f(A_1,A_2)](H_1,H_2) &= -c_\alpha \int_0^\infty \tau^\alpha (\tau I + A_1)^{-1} H_1 (\tau I + A_1)^{-1} H_2 (\tau I + A_1)^{-1} d\tau A_2^{\bar{\alpha}} \\ &\quad - c_\alpha \int_0^\infty \tau^\alpha (\tau I + A_1)^{-1} H_2 (\tau I + A_1)^{-1} H_1 (\tau I + A_1)^{-1} d\tau A_2^{\bar{\alpha}}, \\ D^{(0,2)}[f(A_1,A_2)](H_1,H_2) &= -c_{\bar{\alpha}} A_1^\alpha \int_0^\infty \tau^{\bar{\alpha}} (\tau I + A_2)^{-1} H_1 (\tau I + A_2)^{-1} H_2 (\tau I + A_2)^{-1} d\tau \\ &\quad - c_{\bar{\alpha}} A_1^\alpha \int_0^\infty \tau^{\bar{\alpha}} (\tau I + A_2)^{-1} H_2 (\tau I + A_2)^{-1} H_1 (\tau I + A_2)^{-1} d\tau, \\ D^{(1,1+)}[f(A_1,A_2)](H_1,H_2) &= c_\alpha c_{\bar{\alpha}} \int_0^\infty \tau^\alpha (\tau I + A_1)^{-1} H_1 (\tau I + A_1)^{-1} d\tau \\ &\quad \int_0^\infty \tau^{\bar{\alpha}} (\tau I + A_2)^{-1} H_2 (\tau I + A_2)^{-1} d\tau, \\ D^{(1+,1)}[f(A_1,A_2)](H_1,H_2) &= c_\alpha c_{\bar{\alpha}} \int_0^\infty \tau^\alpha (\tau I + A_1)^{-1} H_2 (\tau I + A_1)^{-1} d\tau \\ &\quad \int_0^\infty \tau^{\bar{\alpha}} (\tau I + A_2)^{-1} H_1 (\tau I + A_2)^{-1} d\tau. \end{aligned}$$

Then, from (22) with $B_1 = \rho_n$, $B_2 = \sigma_n$, $A_1 = \rho$, and $A_2 = \sigma$ and $v(\rho_n, \rho, \tau, t) := \big(\tau I + (1-t)\rho + t\rho_n\big)^{-1}$, we obtain

$$\begin{aligned}
\rho_n^{\alpha}\sigma_n^{\bar{\alpha}} &= \rho^{\alpha}\sigma^{\bar{\alpha}} + c_{\alpha}\int_0^{\infty} \tau^{\alpha}(\tau I+\rho)^{-1}(\rho_n-\rho)(\tau I+\rho)^{-1}d\tau\,\sigma^{\bar{\alpha}} \\
&\quad + c_{\bar{\alpha}}\rho^{\alpha}\int_0^{\infty} \tau^{\bar{\alpha}}(\tau I+\sigma)^{-1}(\sigma_n-\sigma)(\tau I+\sigma)^{-1}d\tau \\
&\quad - 2c_{\alpha}\int_0^1 (1-t)\left[\int_0^{\infty} \tau^{\alpha} v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\right] \\
&\qquad\qquad \big((1-t)\sigma+t\sigma_n\big)^{\bar{\alpha}}dt \\
&\quad - 2c_{\bar{\alpha}}\int_0^1 (1-t)\big((1-t)\rho+t\rho_n\big)^{\alpha}\left[\int_0^{\infty} \tau^{\bar{\alpha}} v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma)v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma)\right. \\
&\qquad\qquad \left. v(\sigma_n,\sigma,\tau,t)d\tau\right]dt \\
&\quad + 2c_{\alpha}c_{\bar{\alpha}}\int_0^1 (1-t)\left[\int_0^{\infty} \tau^{\alpha} v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\right] \\
&\qquad\qquad \left[\int_0^{\infty} \tau^{\bar{\alpha}} v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma)v(\sigma_n,\sigma,\tau,t)d\tau\right]dt.
\end{aligned}$$

Multiplying by r_n, taking trace, and subsequent limits leads to (37) using similar arguments as in Theorem 1, provided

$$\begin{aligned}
\bar{p}_n(r_n) &:= (1-t)\left[\int_0^{\infty} \tau^{\alpha} v(\rho_n,\rho,\tau,t) r_n^{\frac{1}{2}}(\rho_n-\rho)v(\rho_n,\rho,\tau,t) r_n^{\frac{1}{2}}(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\right] \\
&\qquad\qquad \big((1-t)\sigma+t\sigma_n\big)^{\bar{\alpha}}, \\
\bar{q}_n(r_n) &:= (1-t)\big((1-t)\rho+t\rho_n\big)^{\alpha}\left[\int_0^{\infty} \tau^{\bar{\alpha}} v(\sigma_n,\sigma,\tau,t) r_n^{\frac{1}{2}}(\sigma_n-\sigma)v(\sigma_n,\sigma,\tau,t) r_n^{\frac{1}{2}}(\sigma_n-\sigma)\right. \\
&\qquad\qquad \left. v(\sigma_n,\sigma,\tau,t)d\tau\right], \\
\bar{s}_n(r_n) &:= (1-t)\left[\int_0^{\infty} \tau^{\alpha} v(\rho_n,\rho,\tau,t) r_n^{\frac{1}{2}}(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\right]\left[\int_0^{\infty} \tau^{\bar{\alpha}} v(\sigma_n,\sigma,\tau,t) r_n^{\frac{1}{2}}(\sigma_n-\sigma)\right. \\
&\qquad\qquad \left. v(\sigma_n,\sigma,\tau,t)d\tau\right],
\end{aligned}$$

are uniformly Bochner-integrable sequences along a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\sigma)\big) \rightarrow (L_1, L_2)$ in $\|\cdot\|_1$ a.s. To see the required uniform Bochner-integrability, observe that

$$\|\bar{p}_n(r_n)\|_1 \leq \left\|r_n^{\frac{1}{2}}(\rho_n-\rho)\right\|_1^2\left[\int_0^{\infty}\tau^{\alpha}\|v(\rho_n,\rho,\tau,t)\|_{\infty}^3 d\tau\right]\left\|\big((1-t)\sigma+t\sigma_n\big)^{\bar{\alpha}}\right\|_1, \tag{38a}$$

$$\|\bar{q}_n(r_n)\|_1 \leq \left\|r_n^{\frac{1}{2}}(\sigma_n-\sigma)\right\|_1^2\left[\int_0^{\infty}\tau^{\bar{\alpha}}\|v(\sigma_n,\sigma,\tau,t)\|_{\infty}^3 d\tau\right]\left\|\big((1-t)\rho+t\rho_n\big)^{\alpha}\right\|_1, \tag{38b}$$

$$\|\bar{s}_n(r_n)\|_1 \leq \left\|r_n^{\frac{1}{2}}(\rho_n-\rho)\right\|_1\left\|r_n^{\frac{1}{2}}(\sigma_n-\sigma)\right\|_1\left[\int_0^{\infty}\tau^{\alpha}\|v(\rho_n,\rho,\tau,t)\|_{\infty}^2 d\tau\right]\left[\int_0^{\infty}\|v(\sigma_n,\sigma,\tau,t)\|_{\infty}^2 d\tau\right]. \tag{38c}$$

For $\alpha \in (0,1)$, we have by concavity of the function $x \mapsto x^{\alpha}$ that

$$\begin{aligned}
\left\|\big((1-t)\sigma+t\sigma_n\big)^{\bar{\alpha}}\right\|_1 &\leq d^{\alpha}\|(1-t)\sigma+t\sigma_n\|_1^{\bar{\alpha}} \leq d^{\alpha}, \\
\left\|\big((1-t)\rho+t\rho_n\big)^{\alpha}\right\|_1 &\leq d^{\bar{\alpha}}\|(1-t)\rho+t\rho_n\|_1^{\alpha} \leq d^{\bar{\alpha}}.
\end{aligned}$$

Also, since $\rho_{n_{k_j}} \to \rho$ and $\sigma_{n_{k_j}} \to \sigma$ in $\|\cdot\|_1$ a.s., we have that $(1-t)\rho + t\rho_{n_{k_j}} \geq c\rho$ and $(1-t)\sigma + t\sigma_{n_{k_j}} \geq c\sigma$ a.s. for a constant $0 < c < 1$ and sufficiently large j. Consequently, we obtain that the integrals in (38) are finite for $\alpha \in (0,1)$. For instance, the integrand in (38a) is $O(\tau^{\alpha})$ for τ close to zero and $O(\tau^{\alpha-3})$ as $\tau \to \infty$ which implies its finiteness. Hence,

$$\left\|\bar{p}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_{d,\rho,\alpha} \left\|r^{\frac{1}{2}}_{n_{k_j}}(\rho_{n_{k_j}} - \rho)\right\|_1^2, \tag{39a}$$

$$\left\|\bar{q}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_{d,\sigma,\alpha} \left\|r^{\frac{1}{2}}_{n_{k_j}}(\sigma_{n_{k_j}} - \sigma)\right\|_1^2, \tag{39b}$$

$$\left\|\bar{s}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_{d,\rho,\sigma,\alpha} \left\|r^{\frac{1}{2}}_{n_{k_j}}(\rho_{n_{k_j}} - \rho)\right\|_1 \left\|r^{\frac{1}{2}}_{n_{k_j}}(\sigma_{n_{k_j}} - \sigma)\right\|_1. \tag{39c}$$

Then, the desired uniform Bochner-integrability follows from those of the sequences $\big((r^{1/2}_{n_{k_j}}(\rho_{n_{k_j}} - \rho)\big)_{j\in\mathbb{N}}$ and $\big(r^{1/2}_{n_{k_j}}(\sigma_{n_{k_j}} - \sigma)\big)_{j\in\mathbb{N}}$. This completes the proof of the claim for $\alpha \in (0,1)$.

Next, consider the case $\alpha \in (1,2)$. Using the integral representation [96, Lemma 2.8]

$$A^{\alpha} = c_{\alpha-1}\int_0^{\infty} \left(\tau^{\alpha-2}A + \tau^{\alpha}(\tau I + A)^{-1} - \tau^{\alpha-1}I\right) d\tau, \tag{40a}$$

$$A^{\bar{\alpha}} = c_{\bar{\alpha}+1}\int_0^{\infty} \tau^{\bar{\alpha}}(\tau I + A)^{-1} d\tau, \tag{40b}$$

we have

$$\begin{aligned}
D^{(1,0)}[f(A_1,A_2)](H) &= c_{\alpha-1}\int_0^{\infty}\left(\tau^{\alpha-2}H - \tau^{\alpha}(\tau I + A_1)^{-1}H(\tau I + A_1)^{-1}\right) d\tau A_2^{\bar{\alpha}},\\
D^{(0,1)}[f(A_1,A_2)](H) &= -c_{\bar{\alpha}+1}A_1^{\alpha}\int_0^{\infty}\tau^{\bar{\alpha}}(\tau I + A_2)^{-1}H(\tau I + A_2)^{-1}d\tau,\\
D^{(2,0)}[f(A_1,A_2)](H_1,H_2) &= c_{\alpha-1}\int_0^{\infty}\tau^{\alpha}(\tau I + A_1)^{-1}H_1(\tau I + A_1)^{-1}H_2(\tau I + A_1)^{-1}d\tau A_2^{\bar{\alpha}}\\
&\quad + c_{\alpha-1}\int_0^{\infty}\tau^{\alpha}(\tau I + A_1)^{-1}H_2(\tau I + A_1)^{-1}H_1(\tau I + A_1)^{-1}d\tau A_2^{\bar{\alpha}},\\
D^{(0,2)}[f(A_1,A_2)](H_1,H_2) &= c_{\bar{\alpha}+1}A_1^{\alpha}\int_0^{\infty}\tau^{\bar{\alpha}}(\tau I + A_2)^{-1}H_1(\tau I + A_2)^{-1}H_2(\tau I + A_2)^{-1}d\tau\\
&\quad + c_{\bar{\alpha}+1}A_1^{\alpha}\int_0^{\infty}\tau^{\bar{\alpha}}(\tau I + A_2)^{-1}H_2(\tau I + A_2)^{-1}H_1(\tau I + A_2)^{-1}d\tau,\\
D^{(1,1_+)}[f(A_1,A_2)](H_1,H_2) &= -c_{\alpha-1}c_{\bar{\alpha}+1}\left[\int_0^{\infty}\left(\tau^{\alpha-2}H_1 - \tau^{\alpha}(\tau I + A_1)^{-1}H_1(\tau I + A_1)^{-1}\right) d\tau\right]\\
&\qquad \left[\int_0^{\infty}\tau^{\bar{\alpha}}(\tau I + A_2)^{-1}H_2(\tau I + A_2)^{-1}d\tau\right],\\
D^{(1_+,1)}[f(A_1,A_2)](H_1,H_2) &= -c_{\bar{\alpha}+1}c_{\alpha-1}\left[\int_0^{\infty}\left(\tau^{\alpha-2}H_2 - \tau^{\alpha}(\tau I + A_1)^{-1}H_2(\tau I + A_1)^{-1}\right) d\tau\right]\\
&\qquad \left[\int_0^{\infty}\tau^{\bar{\alpha}}(\tau I + A_2)^{-1}H_1(\tau I + A_2)^{-1}d\tau\right].
\end{aligned}$$

Substituting $B_1 = \rho_n$, $B_2 = \sigma_n$, $A_1 = \rho$, and $A_2 = \sigma$, multiplying by r_n, taking trace and following similar arguments as above leads to the claim provided

$$\tilde{p}_n(r_n) := (1-t)\left[\int_0^{\infty}\tau^{\alpha}v(\rho_n,\rho,\tau,t)r_n^{\frac{1}{2}}(\rho_n - \rho)v(\rho_n,\rho,\tau,t)r_n^{\frac{1}{2}}(\rho_n - \rho)v(\rho_n,\rho,\tau,t)d\tau\right]$$
$$\big((1-t)\sigma + t\sigma_n\big)^{\bar{\alpha}},$$

$$\tilde{q}_n(r_n) := (1-t)\big((1-t)\rho + t\rho_n\big)^{\alpha}\bigg[\int_0^{\infty} \tau^{\bar{\alpha}} v(\sigma_n,\sigma,\tau,t) r_n^{\frac{1}{2}}(\sigma_n-\sigma) v(\sigma_n,\sigma,\tau,t) r_n^{\frac{1}{2}}(\sigma_n-\sigma) \\ v(\sigma_n,\sigma,\tau,t) d\tau\bigg],$$

$$\tilde{s}_n(r_n) := (1-t)\bigg[\int_0^{\infty} \Big(\tau^{\alpha-2} r_n^{\frac{1}{2}}(\rho_n-\rho) - \tau^{\alpha} v(\rho_n,\rho,\tau,t)\ r_n^{\frac{1}{2}}(\rho_n-\rho)\ v(\rho_n,\rho,\tau,t)\Big) d\tau\bigg] \\ \bigg[\int_0^{\infty} \tau^{\bar{\alpha}} v(\sigma_n,\sigma,\tau,t) r_n^{\frac{1}{2}}(\sigma_n-\sigma) v(\sigma_n,\sigma,\tau,t) d\tau\bigg],$$

are uniformly Bochner-integrable along a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ satisfying $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\sigma)\big) \to (L_1, L_2)$ in $\|\cdot\|_1$ a.s. Observe that

$$\|\tilde{p}_n(r_n)\|_1 \lesssim_{d,\alpha} \Big\|r_n^{\frac{1}{2}}(\rho_n-\rho)\Big\|_1^2 \bigg[\int_0^{\infty} \tau^{\alpha} \|v(\rho_n,\rho,\tau,t)\|_{\infty}^3 d\tau\bigg], \tag{41a}$$

$$\|\tilde{q}_n(r_n)\|_1 \lesssim_{d,\alpha} \Big\|r_n^{\frac{1}{2}}(\sigma_n-\sigma)\Big\|_1^2 \bigg[\int_0^{\infty} \tau^{\bar{\alpha}} \|v(\sigma_n,\sigma,\tau,t)\|_{\infty}^3 d\tau\bigg], \tag{41b}$$

$$\|\tilde{s}_n(r_n)\|_1 \lesssim_{d,\alpha} \Big\|r_n^{\frac{1}{2}}(\sigma_n-\sigma)\Big\|_1 \int_0^{\infty} \tau^{\bar{\alpha}} \|v(\sigma_n,\sigma,\tau,t)\|_{\infty}^2 d\tau \\ \bigg[\int_0^{\infty} \Big\|\tau^{\alpha-2} r_n^{\frac{1}{2}}(\rho_n-\rho) - \tau^{\alpha} v(\rho_n,\rho,\tau,t)\ r_n^{\frac{1}{2}}(\rho_n-\rho) v(\rho_n,\rho,\tau,t)\Big\|_1 d\tau\bigg]. \tag{41c}$$

It is straightforward to see the finiteness of the integrals in (41), except perhaps the last integral in (41c). For the latter, observe that along a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\rho_{n_{k_j}} \to \rho$, $\sigma_{n_{k_j}} \to \rho$ in $\|\cdot\|_1$ (a.s.), there exists constants $c, c' > 0$ (depending on ρ and the realizations $\rho_{n_{k_j}}, \sigma_{n_{k_j}}$) such that $c'I \geq (1-t)\rho + t\rho_{n_{k_j}}, (1-t)\rho + t\sigma_{n_{k_j}} \geq cI$. Then, we have

$$\big(\tau^{\alpha-2} - \tau^{\alpha}(\tau+c')^{-2}\big)\big(\rho_{n_{k_j}}-\rho\big) \geq \tau^{\alpha-2}\big(\rho_{n_{k_j}}-\rho\big) - \tau^{\alpha} v\big(\rho_{n_{k_j}},\rho,\tau,t\big)\big(\rho_{n_{k_j}}-\rho\big) v\big(\rho_{n_{k_j}},\rho,\tau,t\big) \\ \geq \big(\tau^{\alpha-2} - \tau^{\alpha}(\tau+c)^{-2}\big)\big(\rho_{n_{k_j}}-\rho\big).$$

From Lemma 1(iii), it then follows that

$$\int_0^{\infty} \Big\|\tau^{\alpha-2} r_{n_{k_j}}^{\frac{1}{2}}\big(\rho_{n_{k_j}}-\rho\big) - \tau^{\alpha} v\big(\rho_{n_{k_j}},\rho,\tau,t\big)\ r_{n_{k_j}}^{\frac{1}{2}}\big(\rho_{n_{k_j}}-\rho\big) v\big(\rho_{n_{k_j}},\rho,\tau,t\big)\Big\|_1 d\tau \\ \leq \Big\|r_{n_{k_j}}^{\frac{1}{2}}(\rho_{n_{k_j}}-\rho)\Big\|_1 \bigg(\int_0^{\infty} \big(\tau^{\alpha-2} - \tau^{\alpha}(\tau+c)^{-2}\big) d\tau + \int_0^{\infty} \big(\tau^{\alpha-2} - \tau^{\alpha}(\tau+c')^{-2}\big) d\tau\bigg).$$

The last two integrals are finite for $\alpha \in (1,2)$. Thus, the RHS of the equations in (41) are bounded similar to the RHS of (39), and the desired Bochner-integrability follows. This completes the proof of (12) for $\alpha \in (0,2)$. The case $\alpha = 2$ is simpler as $f(A_1,A_2) = A_1^2 A_2^{-1}$ and the relevant derivatives can be computed using the rules $D[A^2](H) = AH + HA$ and (2). Since rest of the proof is similar to above, we omit the details.

Two-sample null: Consider $\alpha \in (0,1)$. We will show that

$$r_n^2\big(Q_\alpha(\rho_n,\sigma_n) - Q_\alpha(\rho,\rho)\big) = r_n^2\big(Q_\alpha(\rho_n,\sigma_n) - 1\big) \xrightarrow{w} \frac{1}{2}\operatorname{Tr}\big[\rho^{\bar{\alpha}} D^2[\rho^{\alpha}](L_1,L_1) + \rho^{\alpha} D^2[\rho^{\bar{\alpha}}](L_2,L_2) \\ + 2D[\rho^{\alpha}](L_1) D[\rho^{\bar{\alpha}}](L_2)\big]. \tag{42}$$

Then, an application of the functional delta method yields the claim in (13) by noting that $\log 1 = 0$. From (22) with $B_1 = \rho_n$, $B_2 = \sigma_n$, $A_1 = A_2 = \rho$, and $v(\rho_n,\rho,\tau,t) := \big(\tau I + (1-t)\rho + t\rho_n\big)^{-1}$, we obtain

$$\rho_n^{\alpha}\sigma_n^{\bar{\alpha}}$$

$$
\begin{aligned}
&= \rho + c_\alpha \int_0^\infty \tau^\alpha (\tau I+\rho)^{-1}(\rho_n-\rho)(\tau I+\rho)^{-1} d\tau \rho^{\bar{\alpha}} + c_{\bar{\alpha}} \rho^\alpha \int_0^\infty \tau^{\bar{\alpha}} (\tau I+\rho)^{-1}(\sigma_n-\rho)(\tau I+\rho)^{-1} d\tau \\
&\quad - 2c_\alpha \int_0^1 (1-t) \left[\int_0^\infty \tau^\alpha v(\rho_n,\rho,\tau,t)(\rho_n-\rho) v(\rho_n,\rho,\tau,t)(\rho_n-\rho) v(\rho_n,\rho,\tau,t) d\tau \right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \big((1-t)\rho + t\sigma_n\big)^{\bar{\alpha}} dt \\
&\quad - 2c_{\bar{\alpha}} \int_0^1 (1-t) \big((1-t)\rho + t\rho_n\big)^\alpha \left[\int_0^\infty \tau^{\bar{\alpha}} v(\sigma_n,\rho,\tau,t)(\sigma_n-\rho) v(\sigma_n,\rho,\tau,t)(\sigma_n-\rho) \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. v(\sigma_n,\rho,\tau,t) d\tau \right] dt \\
&\quad + 2c_\alpha c_{\bar{\alpha}} \int_0^1 (1-t) \left[\int_0^\infty \tau^\alpha v(\rho_n,\rho,\tau,t)(\rho_n-\rho) v(\rho_n,\rho,\tau,t) d\tau \right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \left[\int_0^\infty \tau^{\bar{\alpha}} v(\sigma_n,\rho,\tau,t)(\sigma_n-\sigma) v(\sigma_n,\rho,\tau,t) d\tau \right] dt.
\end{aligned}
$$

Multiplying by r_n^2, taking trace and following similar arguments as above leads to (42), provided

$$
\mathrm{Tr}\left[\int_0^\infty \tau^\alpha (\tau I+\rho)^{-1}(\rho_n-\rho)(\tau I+\rho)^{-1} d\tau \rho^{\bar{\alpha}} \right] = 0, \tag{43a}
$$

$$
\mathrm{Tr}\left[\rho^\alpha \int_0^\infty \tau^{\bar{\alpha}} (\tau I+\rho)^{-1}(\sigma_n-\rho)(\tau I+\rho)^{-1} d\tau \right] = 0, \tag{43b}
$$

and the sequences

$$
\begin{aligned}
&\bar{p}_n(r_n) \\
&:= (1-t) \left[\int_0^\infty \tau^\alpha v(\rho_n,\rho,\tau,t) r_n(\rho_n-\rho) v(\rho_n,\rho,\tau,t) r_n(\rho_n-\rho) v(\rho_n,\rho,\tau,t) d\tau \right] \big((1-t)\rho + t\sigma_n\big)^{\bar{\alpha}}, \\
&\tilde{q}_n(r_n) \\
&:= (1-t)\big((1-t)\rho + t\rho_n\big)^\alpha \left[\int_0^\infty \tau^{\bar{\alpha}} v(\sigma_n,\rho,\tau,t) r_n(\sigma_n-\rho) v(\sigma_n,\rho,\tau,t) r_n(\sigma_n-\rho) v(\sigma_n,\rho,\tau,t) d\tau \right], \\
&\tilde{s}_n(r_n) := (1-t) \left[\int_0^\infty \left(\tau^{\alpha-2} r_n(\rho_n-\rho) - \tau^\alpha v(\rho_n,\rho,\tau,t)\ r_n(\rho_n-\rho)\ v(\rho_n,\rho,\tau,t) \right) d\tau \right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \left[\int_0^\infty \tau^{\bar{\alpha}} v(\sigma_n,\rho,\tau,t) r_n(\sigma_n-\rho) v(\sigma_n,\rho,\tau,t) d\tau \right],
\end{aligned}
$$

are uniformly Bochner-integrable along a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ satisfying $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\big) \to (L_1, L_2)$ in $\|\cdot\|_1$ a.s. To show (43a), note that we have

$$
\begin{aligned}
\mathrm{Tr}\left[\int_0^\infty \tau^\alpha (\tau I+\rho)^{-1}(\rho_n-\rho)(\tau I+\rho)^{-1} d\tau \rho^{\bar{\alpha}} \right] &\overset{(a)}{=} \mathrm{Tr}\left[(\rho_n-\rho)\rho^{\bar{\alpha}} \int_0^\infty \tau^\alpha (\tau I+\rho)^{-2} d\tau \right] \\
&\overset{(b)}{=} c_\alpha^{-1}\, \mathrm{Tr}\left[\rho_n - \rho \right] = 0,
\end{aligned}
$$

where (a) follows by similar arguments leading to (28) and (b) is due to

$$
\int_0^\infty \tau^\alpha (\tau I+\rho)^{-2} d\tau = -\tau^\alpha (\tau I+\rho)^{-1}\big|_0^\infty - \alpha \int_0^\infty \tau^{\alpha-1} (\tau I+\rho)^{-1} d\tau = c_\alpha^{-1} \rho^{-\bar{\alpha}}.
$$

In the above, the first equality uses integration by parts and the second equality uses (40b). Similarly, (43b) also holds. The desired Bochner-integrability can be shown similar to (39). This completes the proof of (13) for the case $\alpha \in (0,1)$. The proof when $\alpha \in (1,2]$ is similar and hence omitted.

If $[\rho_n,\rho]=[\sigma_n,\rho]=[\sigma_n,\rho_n]=0$, then $[L_1,L_2]=[L_1,\rho]=[L_2,\rho]=0$. Hence,

$$\rho^{\bar{\alpha}}D^2[\rho^{\alpha}](L_1,L_1)=-2\rho^{\bar{\alpha}}L_1^2c_\alpha\int_0^\infty \tau^\alpha(\tau I+\rho)^{-3}d\tau=\rho^{\bar{\alpha}}L_1^2\rho^{\alpha-2}=L_1^2\rho^{-1}.$$

Similarly,

$$\rho^{\alpha}D^2[\rho^{\bar{\alpha}}](L_2,L_2)=L_2^2\rho^{-1},$$

$$2D[\rho^{\alpha}](L_1)D[\rho^{\bar{\alpha}}](L_2)=2L_1L_2\rho^{-1}.$$

Substituting this in (13) leads to (15).

D. Proof of Theorem 3

We will prove the claim for $\rho,\sigma>0$. The general case $\rho\ll\sigma$ follows by an application of Lemma 1 via similar arguments as in the proof of Theorem 1. We begin by noting that the following variational form holds for the sandwiched Rényi divergence as given in [13] (see [97] for a further application of this form in generalizing sandwiched Rényi divergence to infinite dimensional quantum settings):

$$\tilde{\mathsf{D}}_\alpha(\rho\|\sigma):=\frac{\alpha}{\alpha-1}\log\left\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha=\max_{\substack{\eta\geq 0,\\ \|\eta\|_{\frac{\alpha}{\alpha-1}}\leq 1}}\tilde{\mathsf{D}}_\alpha(\rho\|\sigma;\eta), \tag{44}$$

where

$$\tilde{\mathsf{D}}_\alpha(\rho\|\sigma;\eta):=\big(\alpha/(\alpha-1)\big)\log\mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\eta\right].$$

The maximum above is achieved by

$$\eta=\eta^\star(\rho,\sigma,\alpha):=\big(\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\big)^{\alpha-1}/\big\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\big\|_\alpha^{\alpha-1}. \tag{45}$$

Defining $\eta_n^\star:=\eta^\star(\rho_n,\sigma_n,\alpha)$, we have

$$r_n\big(\tilde{\mathsf{D}}_\alpha(\rho_n\|\sigma_n)-\tilde{\mathsf{D}}_\alpha(\rho\|\sigma)\big)\leq r_n\big(\tilde{\mathsf{D}}_\alpha(\rho_n\|\sigma_n;\eta_n^\star)-\tilde{\mathsf{D}}_\alpha(\rho\|\sigma;\eta_n^\star)\big), \tag{46a}$$

$$r_n\big(\tilde{\mathsf{D}}_\alpha(\rho_n\|\sigma_n)-\tilde{\mathsf{D}}_\alpha(\rho\|\sigma)\big)\geq r_n\big(\tilde{\mathsf{D}}_\alpha(\rho_n\|\sigma_n;\eta^\star)-\tilde{\mathsf{D}}_\alpha(\rho\|\sigma;\eta^\star)\big). \tag{46b}$$

We will show that the limit on the RHS of (46a) and (46b) coincides with the RHS of (16), thus proving the desired claim.

To establish that the former limit equals the RHS of (16), it is sufficient to show that

$$\begin{aligned}&r_n\left(\mathrm{Tr}\left[\rho_n^{\frac{1}{2}}\sigma_n^{\frac{\bar{\alpha}}{\alpha}}\rho_n^{\frac{1}{2}}\eta_n^\star\right]-\mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\eta_n^\star\right]\right)\\ &\xrightarrow{w}\frac{\mathrm{Tr}\left[\left(D[\rho^{\frac{1}{2}}](L_1)\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}D[\sigma^{\frac{\bar{\alpha}}{\alpha}}](L_2)\rho^{\frac{1}{2}}\right)\left(\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right]}{\left\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha^{\alpha-1}}.\end{aligned} \tag{47}$$

Then, the functional delta-method applied with the function $x\mapsto\alpha\log x/(\alpha-1)$ yields

$$\begin{aligned}&r_n\big(\tilde{\mathsf{D}}_\alpha(\rho_n\|\sigma_n;\eta_n^\star)-\tilde{\mathsf{D}}_\alpha(\rho\|\sigma;\eta_n^\star)\big)\\ &\xrightarrow{w}\frac{\alpha}{\alpha-1}\frac{\mathrm{Tr}\left[\left(D[\rho^{\frac{1}{2}}](L_1)\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}D[\sigma^{\frac{\bar{\alpha}}{\alpha}}](L_2)\rho^{\frac{1}{2}}\right)\left(\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\right)^{\alpha-1}\right]}{\left\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha^{\alpha}},\end{aligned}$$

as desired. Further, invoking the subsequence argument, it suffices to prove (47) along a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\sigma)\big) \to (L_1, L_2)$ in trace-norm a.s.

To show (47), we compute the Fréchet derivatives of

$$f(A_1,A_2)=A_1^{\frac{1}{2}}A_2^{\frac{\bar{\alpha}}{\alpha}}A_1^{\frac{1}{2}}.$$

Consider $\alpha \in (0.5,1)\cup(1,\infty)$. Since $\bar{\alpha}/\alpha \in (0,1)\cup(-1,0)$, the following integral representations given in [96, Lemma 2.8] are relevant for our purpose:

$$A^{\alpha}=c_{\alpha}\int_0^{\infty}\tau^{\alpha}\left(\frac{1}{\tau I}-\frac{1}{\tau I+A}\right)d\tau, \quad \alpha\in(0,1),$$
$$A^{\alpha}=c_{\alpha+1}\int_0^{\infty}\frac{\tau^{\alpha}}{\tau I+A}d\tau, \quad \alpha\in(-1,0).$$

We have via the chain and product rule for Fréchet derivatives that

$$\begin{aligned}
D^{(1,0)}[f(A_1,A_2)](H) &= c_{\frac{1}{2}}\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H(\tau I+A_1)^{-1}d\tau A_2^{\frac{\bar{\alpha}}{\alpha}}A_1^{\frac{1}{2}}\\
&\qquad + c_{\frac{1}{2}}A_1^{\frac{1}{2}}A_2^{\frac{\bar{\alpha}}{\alpha}}\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H(\tau I+A_1)^{-1}d\tau,\\
D^{(0,1)}[f(A_1,A_2)](H) &= -c_{\frac{\bar{\alpha}}{\alpha}+1}A_1^{\frac{1}{2}}\left[\int_0^{\infty}\tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H(\tau I+A_2)^{-1}d\tau\right]A_1^{\frac{1}{2}},\\
D^{(2,0)}[f(A_1,A_2)](H_1,H_2) &= -c_{\frac{1}{2}}\left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}d\tau\right.\\
&\qquad \left.+\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}d\tau\right]A_2^{\frac{\bar{\alpha}}{\alpha}}A_1^{\frac{1}{2}}\\
&\qquad - c_{\frac{1}{2}}A_1^{\frac{1}{2}}A_2^{\frac{\bar{\alpha}}{\alpha}}\left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}d\tau\right.\\
&\qquad \left.+\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}d\tau\right]\\
&\qquad + c_{\frac{1}{2}}^2\left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}d\tau\right]A_2^{\frac{\bar{\alpha}}{\alpha}}\\
&\qquad\quad \left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}d\tau\right]\\
&\qquad + c_{\frac{1}{2}}^2\left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}d\tau\right]A_2^{\frac{\bar{\alpha}}{\alpha}}\\
&\qquad\quad \left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}d\tau\right],\\
D^{(0,2)}[f(A_1,A_2)](H_1,H_2) &= c_{\frac{\bar{\alpha}}{\alpha}+1}A_1^{\frac{1}{2}}\left[\int_0^{\infty}\tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H_1(\tau I+A_2)^{-1}H_2(\tau I+A_2)^{-1}d\tau\right.\\
&\qquad \left.+\int_0^{\infty}\tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H_2(\tau I+A_2)^{-1}H_1(\tau I+A_2)^{-1}d\tau\right]A_1^{\frac{1}{2}},\\
D^{(1,1+)}[f(A_1,A_2)](H_1,H_2) &= -c_{\frac{1}{2}}c_{\frac{\bar{\alpha}}{\alpha}+1}\left[\int_0^{\infty}\tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}d\tau\right]\\
&\qquad\quad \left[\int_0^{\infty}\tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H_2(\tau I+A_2)^{-1}d\tau\right]A_1^{\frac{1}{2}}\\
&\qquad - c_{\frac{1}{2}}c_{\frac{\bar{\alpha}}{\alpha}+1}A_1^{\frac{1}{2}}\left[\int_0^{\infty}\tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H_2(\tau I+A_2)^{-1}d\tau\right]
\end{aligned}$$

$$\Big[\int_0^\infty \tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_1(\tau I+A_1)^{-1}d\tau\Big],$$

$$D^{(1_+,1)}[f(A_1,A_2)](H_1,H_2) = -c_{\frac{1}{2}}c_{\frac{\bar{\alpha}}{\alpha}+1}\Big[\int_0^\infty \tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}d\tau\Big] \Big[\int_0^\infty \tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H_1(\tau I+A_2)^{-1}d\tau\Big]A_1^{\frac{1}{2}} - c_{\frac{1}{2}}c_{\frac{\bar{\alpha}}{\alpha}+1}A_1^{\frac{1}{2}}\Big[\int_0^\infty \tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+A_2)^{-1}H_1(\tau I+A_2)^{-1}d\tau\Big] \Big[\int_0^\infty \tau^{\frac{1}{2}}(\tau I+A_1)^{-1}H_2(\tau I+A_1)^{-1}d\tau\Big].$$

Then, substituting the expressions for Fréchet derivatives derived above in (22) with $B_1=\rho_n$, $B_2=\sigma_n$, $A_1=\rho$, and $A_2=\sigma$ leads to

$$\rho_n^{\frac{1}{2}}\sigma_n^{\frac{\bar{\alpha}}{\alpha}}\rho_n^{\frac{1}{2}}\eta_n^\star = \rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\eta_n^\star + c_{\frac{1}{2}}\Big[\int_0^\infty \tau^{\frac{1}{2}}(\tau I+\rho)^{-1}(\rho_n-\rho)(\tau I+\rho)^{-1}d\tau\Big]\sigma^{\frac{\bar{\alpha}}{\alpha}}\rho^{\frac{1}{2}}\eta_n^\star + c_{\frac{1}{2}}\rho^{\frac{1}{2}}\sigma^{\frac{\bar{\alpha}}{\alpha}}\Big[\int_0^\infty \tau^{\frac{1}{2}}(\tau I+\rho)^{-1}(\rho_n-\rho)(\tau I+\rho)^{-1}d\tau\Big]\eta_n^\star - c_{\frac{1}{\alpha}}\rho^{\frac{1}{2}}\Big[\int_0^\infty \tau^{\frac{\bar{\alpha}}{\alpha}}(\tau I+\sigma)^{-1}(\sigma_n-\sigma)(\tau I+\sigma)^{-1}d\tau\Big]\rho^{\frac{1}{2}}\eta_n^\star + R_n\eta_n^\star, \tag{48}$$

where $R_n := R_{1,n}+R_{2,n}+R_{3,n}$, and with $v(\rho_n,\rho,\tau,t) := \big(\tau I+(1-t)\rho+t\rho_n\big)^{-1}$,

$$R_{1,n} := -2c_{\frac{1}{2}}\int_0^1(1-t)\int_0^\infty \tau^{\frac{1}{2}}v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau \big((1-t)\sigma+t\sigma_n\big)^{\frac{\bar{\alpha}}{\alpha}}\big((1-t)\rho+t\rho_n\big)^{\frac{1}{2}}dt - 2c_{\frac{1}{2}}\int_0^1(1-t)\big((1-t)\rho+t\rho_n\big)^{\frac{1}{2}}\big((1-t)\sigma+t\sigma_n\big)^{\frac{\bar{\alpha}}{\alpha}} \int_0^\infty \tau^{\frac{1}{2}}v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau dt + 4c_{\frac{1}{2}}^2\int_0^1(1-t)\Big[\int_0^\infty \tau^{\frac{1}{2}}v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\Big]\big((1-t)\sigma+t\sigma_n\big)^{\frac{\bar{\alpha}}{\alpha}} \Big[\int_0^\infty \tau^{\frac{1}{2}}v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\Big]dt,$$

$$R_{2,n} := 2c_{\frac{1}{\alpha}}\int_0^1(1-t)\big((1-t)\rho+t\rho_n\big)^{\frac{1}{2}}\Big[\int_0^\infty \tau^{\frac{\bar{\alpha}}{\alpha}}v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma)v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma) v(\sigma_n,\sigma,\tau,t)d\tau\Big]\big((1-t)\rho+t\rho_n\big)^{\frac{1}{2}}dt,$$

$$R_{3,n} := -2c_{\frac{1}{\alpha}}c_{\frac{1}{2}}\int_0^1(1-t)\Big[\int_0^\infty \tau^{\frac{1}{2}}v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\Big] \Big[\int_0^\infty \tau^{\frac{\bar{\alpha}}{\alpha}}v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma)v(\sigma_n,\sigma,\tau,t)d\tau\Big]\big((1-t)\rho+t\rho_n\big)^{\frac{1}{2}}dt - 2c_{\frac{1}{\alpha}}c_{\frac{1}{2}}\int_0^1(1-t)(1-t)\rho+t\rho_n\big)^{\frac{1}{2}}\Big[\int_0^\infty \tau^{\frac{1}{2}}v(\rho_n,\rho,\tau,t)(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau\Big] \Big[\int_0^\infty \tau^{\frac{\bar{\alpha}}{\alpha}}v(\sigma_n,\sigma,\tau,t)(\sigma_n-\sigma)v(\sigma_n,\sigma,\tau,t)d\tau\Big]dt.$$

Multiplying by r_n and taking limits, the desired claim will follow provided $\mathrm{Tr}\left[r_n R_n \eta_n^\star\right] \to 0$ along the subsequence $(n_{k_j})_{j\in\mathbb{N}}$ mentioned above. To show this, we require an interchange of limits and integral which can be justified via the uniform Bochner-integrability conditions stated below:

$$\begin{aligned}
&\|r_n R_{1,n}\eta_n^\star\|_1 \\
&\lesssim_{d,\alpha} \left\|r_n^{\frac{1}{2}}(\rho_n-\rho)\right\|_1^2 \left[\int_0^\infty \tau^{\frac{1}{2}} \|v(\rho_n,\rho,\tau,t)\|_\infty^3 \, d\tau\right] \|(1-t)\sigma+t\sigma_n\|_1^{\frac{\bar\alpha}{\alpha}} \|(1-t)\rho+t\rho_n\|_1^{\frac{1}{2}} \|\eta_n^\star\|_1 \\
&\qquad\qquad + \left\|r_n^{\frac{1}{2}}(\rho_n-\rho)\right\|_1^2 \|(1-t)\sigma+t\sigma_n\|^{\frac{\bar\alpha}{\alpha}} \left[\int_0^\infty \tau^{\frac{1}{2}} \|v(\rho_n,\rho,\tau,t)\|_\infty^2 \, d\tau\right]^2 \|\eta_n^\star\|_1 , \\
&\|r_n R_{2,n}\eta_n^\star\|_1 \lesssim_{d,\alpha} \|(1-t)\rho+t\rho_n\|_1 \left\|r_n^{\frac{1}{2}}(\sigma_n-\sigma)\right\|_1^2 \left[\int_0^\infty \tau^{\frac{\bar\alpha}{\alpha}} \|v(\sigma_n,\sigma,\tau,t)\|_\infty^3 \, d\tau\right] \|\eta_n^\star\|_1 , \\
&\|r_n R_{3,n}\eta_n^\star\|_1 \lesssim_{d,\alpha} \left\|r_n^{\frac{1}{2}}(\rho_n-\rho)\right\|_1 \left\|r_n^{\frac{1}{2}}(\sigma_n-\sigma)\right\|_1 \|(1-t)\rho+t\rho_n\|_1^{\frac{1}{2}} \left[\int_0^\infty \tau^{\frac{1}{2}} \|v(\rho_n,\rho,\tau,t)\|_\infty^2 \, d\tau\right] \\
&\qquad\qquad \left[\int_0^\infty \tau^{\frac{\bar\alpha}{\alpha}} \|v(\sigma_n,\sigma,\tau,t)\|_\infty^2 \, d\tau\right] \|\eta_n^\star\|_1 .
\end{aligned}$$

Hence, similar to (39), the integrals above are finite, and we have a.s. along the subsequence $(n_{k_j})_{j\in\mathbb{N}}$ that

$$\begin{aligned}
\left\|r_{n_{k_j}} R_{1,n_{k_j}} \eta_{n_{k_j}}^\star\right\|_1 &\lesssim_{d,\rho,\sigma,\alpha} \left\|r_{n_{k_j}}^{\frac{1}{2}}(\rho_{n_{k_j}}-\rho)\right\|_1^2, \\
\left\|r_{n_{k_j}} R_{2,n_{k_j}} \eta_{n_{k_j}}^\star\right\|_1 &\lesssim_{d,\rho,\sigma,\alpha} \left\|r_{n_{k_j}}^{\frac{1}{2}}(\rho_{n_{k_j}}-\rho)\right\|_1^2, \\
\left\|r_{n_{k_j}} R_{3,n_{k_j}} \eta_{n_{k_j}}^\star\right\|_1 &\lesssim_{d,\rho,\sigma,\alpha} \left\|r_{n_{k_j}}^{\frac{1}{2}}(\rho_{n_{k_j}}-\rho)\right\|_1 \left\|r_{n_{k_j}}^{\frac{1}{2}}(\sigma_{n_{k_j}}-\sigma)\right\|_1 .
\end{aligned}$$

From this, the desired integrability follows from those of the terms in the RHS. Also, note that since $\rho_{n_{k_j}} \to \rho$ and $\sigma_{n_{k_j}} \to \sigma$ a.s. in trace norm, $\eta_{k_j}^\star$ converges a.s. in trace norm to $\eta^\star$ due to $\rho,\sigma>0$. This is because for all sufficiently large j (depending on the realizations), the eigenvalues (and eigenvectors) of both $\rho_{n_{k_j}}$ and $\sigma_{n_{k_j}}$ are arbitrarily close to that of ρ and σ, respectively, which are all bounded away from zero. Hence, we obtain $\mathrm{Tr}\left[r_{n_{k_j}} R_{n_{k_j}} \eta_{n_{k_j}}^\star\right] \to 0$. Consequently, $r_n\big(\mathrm{Tr}\left[\rho_n^{\frac{1}{2}}\sigma_n^{\frac{\bar\alpha}{\alpha}}\rho_n^{\frac{1}{2}}\eta_n^\star\right] - \mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{\frac{\bar\alpha}{\alpha}}\rho^{\frac{1}{2}}\eta_n^\star\right]\big)$ converges to the RHS of (47) along $(n_{k_j})_{j\in\mathbb{N}}$. In a similar vein, $r_n\big(\mathrm{Tr}\left[\rho_n^{\frac{1}{2}}\sigma_n^{\frac{\bar\alpha}{\alpha}}\rho_n^{\frac{1}{2}}\eta^\star\right] - \mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{\frac{\bar\alpha}{\alpha}}\rho^{\frac{1}{2}}\eta^\star\right]\big)$ converges a.s. to RHS of (47) along the same subsequence, thus proving (47) and (16). The proof for $\alpha=0.5$ can be shown similar to above. In this case, the Fréchet derivatives are easier to compute since $f(A_1,A_2)=A_1^{\frac{1}{2}}A_2A_1^{\frac{1}{2}}$.

To show that (16) simplifies to (14) when all relevant operators commute, observe that $D[\rho^{\frac{1}{2}}](L_1)=L_1\rho^{\frac{1}{2}}/2$ and $D[\sigma^{\frac{\bar\alpha}{\alpha}}](L_2)=(\bar\alpha/\alpha)L_2\sigma^{\frac{\bar\alpha}{\alpha}-1}$, and hence

$$\left(D[\rho^{\frac{1}{2}}](L_1)\sigma^{\frac{\bar\alpha}{\alpha}}\rho^{\frac{1}{2}} + \rho^{\frac{1}{2}}\sigma^{\frac{\bar\alpha}{\alpha}}D[\rho^{\frac{1}{2}}](L_1) + \rho^{\frac{1}{2}}D[\sigma^{\frac{\bar\alpha}{\alpha}}](L_2)\rho^{\frac{1}{2}}\right)\left(\rho^{\frac{1}{2}}\sigma^{\frac{\bar\alpha}{\alpha}}\rho^{\frac{1}{2}}\right)^{\alpha-1} = L_1\sigma^{\bar\alpha}\rho^{-\bar\alpha} + L_2\frac{\bar\alpha}{\alpha}\sigma^{-\alpha}\rho^{\alpha}.$$

Substituting this in (16) and noting that $\left\|\rho^{\frac{1}{2}}\sigma^{\frac{\bar\alpha}{\alpha}}\rho^{\frac{1}{2}}\right\|_\alpha^\alpha = \mathrm{Tr}\left[\rho^\alpha\sigma^{\bar\alpha}\right]$ proves the claim.

E. *Proof of Corollary 4*

By setting $\alpha=0.5$ in (16), we obtain

$$r_n\left(\tilde{\mathsf{D}}_{\frac{1}{2}}(\rho_n\|\sigma_n) - \tilde{\mathsf{D}}_{\frac{1}{2}}(\rho\|\sigma)\right) \xrightarrow{w} \frac{-\mathrm{Tr}\left[\left(D[\rho^{\frac{1}{2}}](L_1)\sigma\rho^{\frac{1}{2}} + \rho^{\frac{1}{2}}\sigma D[\rho^{\frac{1}{2}}](L_1) + \rho^{\frac{1}{2}}L_2\rho^{\frac{1}{2}}\right)\left(\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}\right)^{-\frac{1}{2}}\right]}{\sqrt{F(\rho,\sigma)}}.$$

The claim for fidelity then follows by noting that $F(\rho,\sigma)=e^{-\tilde{\mathsf{D}}_{1/2}(\rho\|\sigma)}$, and applying the functional delta method to the above equation for the map $x\mapsto e^{-x}$.

Next, consider max-divergence given in (6) which corresponds to infinite-order sandwiched Rényi divergence. The variational form in (44) with $\alpha=\infty$ becomes

$$\mathsf{D}_{\max}(\rho\|\sigma)=\max_{\eta\geq 0:\|\eta\|_1\leq 1}\mathsf{D}_{\max}(\rho\|\sigma;\eta),$$

where

$$\mathsf{D}_{\max}(\rho\|\sigma;\eta):=\log\mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\eta\right].$$

The maximum above is achieved by $\eta^\star=\Pi_{\max}$, where $\Pi_{\max}$ is the eigenprojection corresponding to the maximal eigenvalue of $\rho^{1/2}\sigma^{-1}\rho^{1/2}$. The rest of the proof is similar to that of Theorem 3 and proceeds by arguing that

$$\eta_n^\star=\Pi_{n,\max}\to\Pi_{\max},$$

a.s. in trace norm,

$$\begin{aligned}
&r_n\left(\mathrm{Tr}\left[\rho_n^{\frac{1}{2}}\sigma_n^{-1}\rho_n^{\frac{1}{2}}\Pi_{n,\max}\right]-\mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\Pi_{n,\max}\right]\right)\\
&\qquad\xrightarrow{w}\mathrm{Tr}\left[\left(D[\rho^{\frac{1}{2}}](L_1)\sigma^{-1}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{-1}D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}D[\sigma^{-1}](L_2)\rho^{\frac{1}{2}}\right)\Pi_{\max}\right],\\
&r_n\left(\mathrm{Tr}\left[\rho_n^{\frac{1}{2}}\sigma_n^{-1}\rho_n^{\frac{1}{2}}\Pi_{\max}\right]-\mathrm{Tr}\left[\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\Pi_{\max}\right]\right)\\
&\qquad\xrightarrow{w}\mathrm{Tr}\left[\left(D[\rho^{\frac{1}{2}}](L_1)\sigma^{-1}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{-1}D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}D[\sigma^{-1}](L_2)\rho^{\frac{1}{2}}\right)\Pi_{\max}\right],
\end{aligned}$$

where $\Pi_{n,\max}$ is the eigenprojection corresponding to the maximal eigenvalue of $\rho_n^{1/2}\sigma_n^{-1}\rho_n^{1/2}$. These together imply that

$$\begin{aligned}
&r_n\left(e^{\mathsf{D}_{\max}(\rho_n\|\sigma_n)}-e^{\mathsf{D}_{\max}(\rho\|\sigma)}\right)\\
&\qquad\xrightarrow{w}\mathrm{Tr}\left[\left(D[\rho^{\frac{1}{2}}](L_1)\sigma^{-1}\rho^{\frac{1}{2}}+\rho^{\frac{1}{2}}\sigma^{-1}D[\rho^{\frac{1}{2}}](L_1)+\rho^{\frac{1}{2}}D[\sigma^{-1}](L_2)\rho^{\frac{1}{2}}\right)\Pi_{\max}\right].
\end{aligned}$$

Then, applying the functional delta-method to $x\mapsto\log x$ at $x=e^{\mathsf{D}_{\max}(\rho\|\sigma)}$ yields the desired claim.

F. Proof of Theorem 4

As in the proof of Theorem 1, we will prove the claim assuming $\rho,\sigma>0$. The general case $\rho\ll\sigma$ follows using similar arguments as in the proof of Theorem 1 by using Lemma 1. Following the proof of Theorem 1, it suffices to show that the terms in a Taylor's expansion of the quantum relative entropy are well-defined and the uniform Bochner-integrability of the remainder terms are ensured. Note that $\mathsf{D}\left(\rho_n\|\sigma_n\right)<\infty$ for all $n\in\mathbb{N}$ and $\mathsf{D}\left(\rho\|\sigma\right)<\infty$ by assumption.

Two-sample null: Consider a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho),r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\big)\to(L_1,L_2)$ in $\|\cdot\|_1$ a.s. This is again possible by Skorokhods representation theorem by separability of the space of trace-class operators (i.e., set of operators with finite trace norm). However, the key difference from the finite dimensional case is that the argument that there exists a constant $0<c<1$ such that $(1-t)\rho+t\sigma_{n_{k_j}}\geq c\rho$ for sufficiently large

j does not hold. Hence, we need a different argument to ensure uniform Bochner-integrability of the terms which we provide next under the additional assumption that $\mathbb{P}\big(\left\|\rho_n\sigma_n^{-1}\right\|_\infty > c\big) \to 0$.

Since the high-level proof is similar to that of Theorem 1, we will only highlight the differences. To begin, we note that the steps from (22) to (29) hold. Next, we show uniform Bochner-integrability of the terms defined in (30):

$$p_n(r_n) := (1-t)r_n(\rho_n-\rho)\int_0^\infty v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau, \tag{49a}$$

$$q_n(r_n) := (1-t)\big((1-t)\rho+t\rho_n\big)\int_0^\infty v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)r_n(\rho_n-\rho)v(\rho_n,\rho,\tau,t)d\tau, \tag{49b}$$

$$\tilde{p}_n(r_n) := (1-t)\big((1-t)\rho+t\rho_n\big)\int_0^\infty v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)d\tau, \tag{49c}$$

$$\tilde{q}_n(r_n) := (1-t)r_n(\rho_n-\rho)\int_0^\infty v(\sigma_n,\rho,\tau,t)r_n(\sigma_n-\rho)v(\sigma_n,\rho,\tau,t)d\tau, \tag{49d}$$

where $v(\rho_n,\rho,\tau,t) := \big(\tau I+(1-t)\rho+t\rho_n\big)^{-1}$. The first and last term can be bounded as

$$\begin{aligned}
\left\|p_n(r_n)\right\|_1 &\leq (1-t)\left\|r_n(\rho_n-\rho)\right\|_1^2\int_0^\infty \left\|v(\rho_n,\rho,\tau,t)\right\|_\infty^2 d\tau \\
&\lesssim_\rho (1-t)\left\|r_n(\rho_n-\rho)\right\|_1^2(1-t)^{-1} \\
&\leq \left\|r_n(\rho_n-\rho)\right\|_1^2, \qquad (50a)\\
\left\|\tilde{q}_n(r_n)\right\|_1 &\leq (1-t)\left\|r_n(\rho_n-\rho)\right\|_1\left\|r_n(\sigma_n-\rho)\right\|_1\int_0^\infty \left\|v(\sigma_n,\rho,\tau,t)\right\|_\infty^2 d\tau \\
&\lesssim_\rho \left\|r_n(\rho_n-\rho)\right\|_1\left\|r_n(\sigma_n-\rho)\right\|_1. \qquad (50b)
\end{aligned}$$

For the second and third terms, we have by using Hölder's inequality for Schatten-norms and $\left\|A\right\|_p \geq \left\|A\right\|_q$ for any linear operator A and $1\leq p\leq q\leq\infty$, that

$$\begin{aligned}
\left\|q_n(r_n)\right\|_1 &\leq (1-t)\left\|r_n(\rho_n-\rho)\right\|_1^2\int_0^\infty \left\|\big((1-t)\rho+t\rho_n\big)v(\rho_n,\rho,\tau,t)\right\|_\infty\left\|v(\rho_n,\rho,\tau,t)\right\|_\infty^2 d\tau \\
&\overset{(a)}{\leq} (1-t)\left\|r_n(\rho_n-\rho)\right\|_1^2\int_0^\infty \left\|v(\rho_n,\rho,\tau,t)\right\|_\infty^2 d\tau \\
&\lesssim_\rho (1-t)\left\|r_n(\rho_n-\rho)\right\|_1^2(1-t)^{-1} \\
&\leq \left\|r_n(\rho_n-\rho)\right\|_1^2, \\
\left\|\tilde{p}_n(r_n)\right\|_1 &\leq (1-t)\left\|r_n(\sigma_n-\rho)\right\|_1^2\int_0^\infty \left\|\big((1-t)\rho+t\rho_n\big)v(\sigma_n,\rho,\tau,t)\right\|_\infty\left\|v(\sigma_n,\rho,\tau,t)\right\|_\infty^2 d\tau \\
&\overset{(b)}{\leq} (1-t)\left\|r_n(\sigma_n-\rho)\right\|_1^2\left\|I+\rho_n\sigma_n^{-1}\right\|_\infty\int_0^\infty \left\|v(\sigma_n,\rho,\tau,t)\right\|_\infty^2 d\tau.
\end{aligned}$$

Here, we used

$$\begin{aligned}
&\big((1-t)\rho+t\rho_n\big)v(\rho_n,\rho,\tau,t)\leq I, \\
&\big((1-t)\rho+t\rho_n\big)v(\sigma_n,\rho,\tau,t)\leq I+\rho_n\sigma_n^{-1},
\end{aligned}$$

in (a) and (b), respectively. Since $\mathbb{P}\big(\|\rho_n\sigma_n^{-1}\|_\infty > c\big) \to 0$, there exists a constant $c > 0$ and a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\big) \to (L_1, L_2)$ in $\|\cdot\|_1$ and $\|\rho_{n_{k_j}}\sigma_{n_{k_j}}^{-1}\|_\infty \leq c$ a.s. Hence, along this subsequence, we have

$$\left\|p_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right\|_1^2, \tag{51a}$$

$$\left\|\tilde{q}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right\|_1 \left\|r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\right\|_1, \tag{51b}$$

$$\left\|q_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_\rho \left\|r_{n_{k_j}}(\rho_{n_{k_j}}-\rho)\right\|_1^2, \tag{51c}$$

$$\left\|\tilde{p}_{n_{k_j}}(r_{n_{k_j}})\right\|_1 \lesssim_{\rho,c} \left\|r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\right\|_1^2. \tag{51d}$$

This verifies the required uniform Bochner-integrability conditions. The rest of the proof is same as that of Theorem 1, and hence omitted.

Two-sample alternative: Consider the expansions in (34). By taking trace, we obtain (35) using (29) as well as $\mathsf{D}(\rho_n\|\sigma_n) < \infty$ and $\mathsf{D}(\rho\|\sigma) < \infty$. We need to verify that the remaining terms in the expansion are well-defined and that the second-order terms satisfy a uniform Bochner-integrability condition along a subsequence $(n_{k_j})_{j\in\mathbb{N}}$ such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}}-\rho), r_{n_{k_j}}(\sigma_{n_{k_j}}-\rho)\big) \to (L_1, L_2)$ in $\|\cdot\|_1$ and $\|\rho_{n_{k_j}}\sigma_{n_{k_j}}^{-1}\|_\infty \leq c$ a.s. Note that since $\rho,\sigma > 0$, we have

$$\|r_n(\rho_n-\rho)(\log\rho - \log\sigma)\|_1 \leq \|r_n(\rho_n-\rho)\|_1 \|\log\rho-\log\sigma\|_\infty,$$
$$\left\|\rho\int_0^\infty (\tau I+\sigma)^{-1} r_n(\sigma_n-\sigma)(\tau I+\sigma)^{-1}d\tau\right\|_1 \leq \|r_n(\sigma_n-\sigma)\|_1 \int_0^\infty \|(\tau I+\sigma)^{-1}\|_\infty^2 \lesssim_\sigma \|r_n(\sigma_n-\sigma)\|_1,$$

which further implies by a version of Slutsky's theorem (see e.g. [22, Lemma 4]) that

$$\begin{aligned}&\mathrm{Tr}\left[r_n(\rho_n-\rho)(\log\rho-\log\sigma) - \rho\int_0^\infty (\tau I+\sigma)^{-1} r_n(\sigma_n-\sigma)(\tau I+\sigma)^{-1}d\tau\right]\\ &\xrightarrow{w} \mathrm{Tr}\left[L_1(\log\rho-\log\sigma) - \rho\int_0^\infty (\tau I+\sigma)^{-1} L_2 (\tau I+\sigma)^{-1}d\tau\right].\end{aligned}$$

It remains to show that $p_n(r_n), q_n(r_n), \bar{p}_n(r_n), \bar{q}_n(r_n)$ as defined in (30) and (36) are uniformly Bochner-integrable along the subsequence $(n_{k_j})_{j\in\mathbb{N}}$. But, this follows similar to (51), thus completing the proof.

G. Proof of Proposition 1

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

$$\hat{s}_j^{(n)}(\rho) := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{O_k(j,\rho)\in\Lambda_j^+} - \mathbb{1}_{O_k(j,\rho)\in\Lambda_j^-},\ 1\leq j\leq d^2-1.$$

Since $\gamma_0 = I$, by setting $\hat{s}_0^{(n)}(\rho) = 1$, we have

$$\sqrt{n}(\hat{\rho}_n-\rho) = \sqrt{n}\left(\hat{\rho}_n - \frac{1}{d}\sum_{j=0}^{d^2-1}\hat{s}_j^{(n)}(\rho)\gamma_j\right) + \sqrt{n}\left(\frac{1}{d}\sum_{j=0}^{d^2-1}\hat{s}_j^{(n)}(\rho)\gamma_j - \rho\right). \tag{52}$$

Note that the first term above can be written as

$$\sqrt{n}\left(\hat{\rho}_n - \frac{1}{d}\sum_{j=0}^{d^2-1}\hat{s}_j^{(n)}(\rho)\gamma_j\right) = \mathbb{1}_{\|\hat{s}^{(n)}(\rho)\|_2>1}\frac{\sqrt{n}}{d}\sum_{j=1}^{d^2-1}\left(\frac{\hat{s}_j^{(n)}(\rho)}{\|\hat{s}^{(n)}(\rho)\|_2} - \hat{s}_j^{(n)}(\rho)\right)\gamma_j.$$

From this, it follows that

$$\sqrt{n}\left\|\hat{\rho}_n-\frac{1}{d}\sum_{j=0}^{d^2-1}\hat{s}_j^{(n)}(\rho)\gamma_j\right\|\lesssim_d\sqrt{n}.$$

Consequently

$$\mathbb{E}\left[\sqrt{n}\left\|\hat{\rho}_n-\frac{1}{d}\sum_{j=0}^{d^2-1}\hat{s}_j^{(n)}(\rho)\gamma_j\right\|_1\right]\lesssim_d\mathbb{P}\left(\left\|\hat{s}^{(n)}(\rho)\right\|_2>1\right)\sqrt{n}. \tag{53}$$

Since $\mathbb{E}\big[\hat{s}^{(n)}(\rho)\big]=s(\rho)$ and $\|s(\rho)\|_2<1$ (as ρ is a mixed state), we have by an application of Hoeffding's inequality that

$$\mathbb{P}\left(\left\|\hat{s}^{(n)}(\rho)\right\|_2>1\right)\leq\mathbb{P}\left(\left\|\hat{s}^{(n)}(\rho)-s(\rho)\right\|_2>1-\left\|s(\rho)\right\|_2\right)\leq e^{-nc_{\rho,d}},$$

where $c_{\rho,d}>0$ is some constant. Consequently, the LHS of (53) converges to zero which implies that the first term in the RHS of (52) converges weakly to zero.

Next, consider the second term in the RHS of (52). Since the measurements for different Pauli operators are done on independent copies of ρ, $\hat{s}_j^{(n)}(\rho)$ are independent across different j. Setting $X_k(j,\rho):=\mathbb{1}_{O_k(j,\rho)\in\Lambda_j^+}-\mathbb{1}_{O_k(j,\rho)\in\Lambda_j^-}$ and noting that $\mathbb{E}[X_k(j,\rho)]=s_j(\rho)$, we have by the classical central limit theorem (CLT) that

$$\frac{n^{\frac{1}{2}}}{d}\sum_{j=0}^{d^2-1}\left(\hat{s}_j^{(n)}(\rho)\gamma_j-\rho\right)=\sum_{j=1}^{d^2-1}\frac{\gamma_j n^{-\frac{1}{2}}}{d}\sum_{k=1}^{n}\left(X_k(j,\rho)-s_j(\rho)\right)\xrightarrow{w}\sum_{j=1}^{d^2-1}\gamma_j Z_j(\rho)=\underbrace{\sum_{j=1}^{d^2-1}\gamma_j Z_j(\rho)}_{L_\rho},$$

where $Z_j(\rho)\sim N\big(0,4s_j^+(\rho)s_j^-(\rho)/d^2\big)$ are independent for different j. Combining the above with Slutsky's theorem yields $\sqrt{n}(\hat{\rho}_n-\rho)\xrightarrow{w}L_\rho$. Repeating the same arguments with ρ replaced by σ and noting that $\|I\|_1/n\to 0$ leads to $\sqrt{n}(\hat{\sigma}_n-\sigma)\xrightarrow{w}L_\sigma$. Consequently, $\big(\sqrt{n}(\hat{\rho}_n-\rho),\sqrt{n}(\hat{\sigma}_n-\sigma)\big)\xrightarrow{w}(L_\rho,L_\sigma)$, where L_ρ and L_σ are independent due to the independence of the measurements on ρ and σ. Moreover, $\hat{\rho}_n\ll\hat{\sigma}_n\ll\sigma$ and $\hat{\rho}_n\ll\rho\ll\sigma$ since $\hat{\rho}_n\ll\hat{\sigma}_n$ and $0<\rho\ll\sigma$. Hence, Theorem 1 applies and (7) yields

$$\begin{aligned}\sqrt{n}\big(\mathsf{D}\left(\hat{\rho}_n\|\hat{\sigma}_n\right)-\mathsf{D}\left(\rho\|\sigma\right)\big)&\xrightarrow{w}\operatorname{Tr}\left[L_\rho(\log\rho-\log\sigma)-\rho D[\log\sigma](L_\sigma)\right]\\&=\sum_{j=1}^{d^2-1}Z_j(\rho)\operatorname{Tr}\left[\gamma_j(\log\rho-\log\sigma)\right]-Z_j(\sigma)\operatorname{Tr}\left[\rho D[\log\sigma](\gamma_j)\right]\\&\sim N(0,v_2^2(\rho,\sigma)),\end{aligned}$$

where $v_2^2(\rho,\sigma)$ is defined in Proposition 1. This completes the proof.

H. Proof of Proposition 2

We will use Proposition 1 to prove the claim. To that end, we first bound the relevant variances $v_1^2(\rho_k,\sigma)$ for $k\in\mathcal{I}$ given by

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

Also, $s_j^+(\rho_k)s_j^-(\rho_k) \leq 1/4$ since $s_j^+(\rho_k) = 1 - s_j^-(\rho_k)$. Hence, we have

$$\begin{aligned}
v_1^2(\rho_k,\sigma) &\leq \frac{1}{d^2}\sum_{j=1}^{d^2-1} \mathrm{Tr}\left[\gamma_j(\log\rho_k - \log\sigma)\right]^2 \\
&\leq \frac{2}{d^2}\sum_{j=1}^{d^2-1} \mathrm{Tr}\left[\gamma_j \log\rho_k\right]^2 + \mathrm{Tr}\left[\gamma_j\log\sigma)\right]^2 \\
&\leq \frac{2}{d^2}\sum_{j=1}^{d^2-1} \|\gamma_j\|_2^2\,\|\log\rho_k\|_2^2 + \|\gamma_j\|_2^2\,\|\log\sigma)\|_2^2 \\
&\leq 4d^2(\log b)^2,
\end{aligned}$$

where, the penultimate inequality, follows by Cauchy-Schawarz inequality, and the final inequality uses $\|\gamma_j\|_2^2 \leq d$ and $\|\log\sigma\|_2^2 \vee \|\log\rho_k\|_2^2 \leq d(\log b)^2$ for all $k \in \mathcal{I}$. Then,

$$\begin{aligned}
\alpha_{i,n}\big(\mathcal{T}_n^{\mathrm{tom}},\rho_i^{\otimes n}\big) &= \mathbb{P}\big(\hat{D}_n \notin (\epsilon_i + cn^{-\frac{1}{2}}, \epsilon_{i+1} + cn^{-\frac{1}{2}})|H=i\big) \\
&= \mathbb{P}\big(\hat{D}_n - \mathsf{D}\left(\rho_i\|\sigma\right) \notin (\epsilon_i - \mathsf{D}\left(\rho_i\|\sigma\right) + cn^{-\frac{1}{2}}, \epsilon_{i+1} - \mathsf{D}\left(\rho_i\|\sigma\right) + cn^{-\frac{1}{2}})|H=i\big) \\
&\leq \mathbb{P}\big(n^{\frac{1}{2}}\big(\hat{D}_n - \mathsf{D}\left(\rho_i\|\sigma\right)\big) \notin (n^{\frac{1}{2}}(\epsilon_i - \mathsf{D}\left(\rho_i\|\sigma\right)) + c, c)|H=i\big).
\end{aligned}$$

Note that $n^{\frac{1}{2}}\big(\hat{D}_n - \mathsf{D}\left(\rho_i\|\sigma\right)\big) \xrightarrow{w} N(0, v_1^2(\rho_i,\sigma))$ given H_i is the true hypothesis by (20a) in Proposition 1. Then, taking limits in the equation above and applying Portmanteaus theorem [90, Theorem 2.1] yields

$$\begin{aligned}
\limsup_{n\to\infty} \alpha_{i,n}\big(\mathcal{T}_n^{\mathrm{tom}},\rho_i^{\otimes n}\big) &\leq \limsup_{n\to\infty} \mathbb{P}\big(n^{\frac{1}{2}}\big(\hat{D}_n - \mathsf{D}\left(\rho_i\|\sigma\right)\big) \notin (n^{\frac{1}{2}}(\epsilon_i - \mathsf{D}\left(\rho_i\|\sigma\right)) + c, c)|H=i\big) \\
&= \mathbb{P}\big(W \notin (-\infty, c)\big) \\
&= Q\left(c/v_1(\rho_i,\sigma)\right),
\end{aligned}$$

where $W \sim N(0, v_1^2(\rho_i,\sigma))$, $Q(x) := (2\pi)^{-1/2}\int_x^\infty e^{-x^2}dx$ is the complementary error function. In the inequality above, we used $\epsilon_i < \mathsf{D}\left(\rho_i\|\sigma\right) \leq \epsilon_{i+1}$. Setting $c = 2dQ^{-1}(\tau)|\log b|$, the RHS above is bounded by τ for all $i \in \mathcal{I}$.

Appendix A
Proof of Lemma 1

Let $\ker A$ denote the kernel of an operator A. To prove (i), we show that if $u \in (\ker AB)^\perp$, then $u \in \ker A^\perp$. Suppose on the contrary, there exists u such that $ABu \neq 0$ and $Au = 0$. Then $Bu \neq 0$. Let $u = Bu + (u - Bu)$, where $u - Bu \in \ker B$. Since $\ker B \subseteq \ker A$ due to $A \ll B$, $A(u - Bu) = 0$. But, then $0 = Au = ABu + A(u - Bu) = ABu \neq 0$, which is the desired contradiction.

Next, we prove (ii). Consider any eigenvector u of AB corresponding to an eigenvalue λ. Let $0 \neq u = Pu + v$ for $v = u - Pu \in \ker P$. Then, we have $\lambda u = ABu = ABPu + ABv = ABPu$ since $v \in \ker P$ implies $v \in \ker(AB)$ due to $AB \ll A \ll P$ by Part (i). From the above, it follows that λ is an eigenvalue of AB if and only if it is an eigenvalue of ABP. Next, suppose u is any eigenvector of ABP corresponding to an eigenvalue λ. We have $BPu = PBPu + w$, where $w = BPu - PBPu \in \ker P$. Then, $\lambda u = ABPu = APBPu + Aw = APBPu$ since $w \in \ker P$ implies $w \in \ker A$. This implies that ABP and $APBP$ have the same eigenvalues.

Finally, suppose u is an eigenvector of APB corresponding to an eigenvalue λ. We have $Bu = PBu + z$, where $z = Bu - PBu \in \ker P \subseteq \ker A$. Hence, we have $\lambda u = APBu = ABu - Az = ABu$. It follows that AB, APB, ABP and $APBP$ have the same eigenvalues. Since the trace of a square linear operator is the sum of its eigenvalues, the claim follows.

Finally, we prove (iii). By the spectral theorem, we have $A = \sum_{i=1}^{d} \lambda_{i,A} |u_{i,A}\rangle\langle u_{i,A}|$, where the $\{\lambda_{i,A}\}_{i=1}^{d}$ and $\{u_{i,A}\}_{i=1}^{d}$ are the set of eigenvalues and corresponding set of orthonormal eigenvectors of A, respectively. Then, $B \leq A \leq C$ implies that for all $1 \leq j \leq d$,

$$\sum_{i=1}^{d} \lambda_{i,B} \left|\langle u_{j,A}, u_{i,B}\rangle\right|^2 \leq \langle u_{j,A}|A|u_{j,A}\rangle = \lambda_{j,A} \leq \sum_{i=1}^{d} \lambda_{i,C} \left|\langle u_{j,A}, u_{i,C}\rangle\right|^2.$$

Hence, we have

$$\begin{aligned}\|A\|_1 = \sum_{j=1}^{d} |\lambda_{j,A}| &\leq \sum_{j=1}^{d} \max\left\{\sum_{i=1}^{d} |\lambda_{i,B}| \left|\langle u_{j,A}, u_{i,B}\rangle\right|^2, \sum_{i=1}^{d} |\lambda_{i,C}| \left|\langle u_{j,A}, u_{i,C}\rangle\right|^2\right\} \\ &\leq \sum_{j=1}^{d}\sum_{i=1}^{d} |\lambda_{i,B}| \left|\langle u_{j,A}, u_{i,B}\rangle\right|^2 + |\lambda_{i,C}| \left|\langle u_{j,A}, u_{i,C}\rangle\right|^2 \\ &= \sum_{i=1}^{d} |\lambda_{i,B}| + |\lambda_{i,C}| = \|B\|_1 + \|C\|_1,\end{aligned}$$

where in the penultimate equality, we interchanged the summations and used that the eigenvectors form an orthonormal basis implying that

$$\sum_{j=1}^{d} \left|\langle u_{j,A}, u_{i,B}\rangle\right|^2 = \sum_{j=1}^{d} \left|\langle u_{j,A}, u_{i,C}\rangle\right|^2 = 1, \ \forall\ 1 \leq i \leq d.$$

This completes the proof of the lemma.

Appendix B
Limit distributions for Measured Relative Entropy

Here, we derive limit distributions for estimators of measured relative entropy with respect to a general class of measurements (see [18], [87]–[89] for certain important classes). Before stating our result, we need to introduce some terminology. Let $\mathcal{M}$ denote a set of POVMs (see the books [15]–[17]), where a POVM here refers to a map M from a discrete set $\mathcal{I}$ to $\mathcal{P}_d$ satisfying $\sum_{i\in\mathcal{I}} M(i) = I$. The measured relative entropy between density operators ρ and σ with respect to $\mathcal{M}$ is

$$\mathsf{D}_{\mathcal{M}}\left(\rho\|\sigma\right) := \sup_{M\in\mathcal{M}} \mathsf{D}\left(P_{\rho,M}\|P_{\sigma,M}\right), \tag{54}$$

where $P_{\rho,M}$ denotes the probability measure defined via $P_{\rho,M}(i) := \mathrm{Tr}\left[M(i)\rho\right]$. In what follows, we will also use the notation $P_{A,M}(\cdot)$ for a general operator A and POVM M, in which case it is no longer necessarily a probability distribution.

Each POVM M defines a measurement (quantum to classical) channel given by $M(\rho) := \sum_{i\in\mathcal{I}} \mathrm{Tr}\left[M(i)\rho\right]|i\rangle\langle i|$, where $|i\rangle$ denotes the i^{th} computational basis element. With this notation, $\mathsf{D}_{\mathcal{M}}\left(\rho\|\sigma\right) = \sup_{M\in\mathcal{M}} \mathsf{D}\left(M(\rho)\|M(\sigma)\right)$. We identify two POVMs M and $\tilde{M}$ in $\mathcal{M}$ if for every $\rho,\sigma \in \mathcal{S}_d$, $P_{\rho,M}(\cdot)$ and $P_{\sigma,M}(\cdot)$ are equivalent (up to the same permutation) to $P_{\rho,\tilde{M}}(\cdot)$ and $P_{\sigma,\tilde{M}}(\cdot)$, respectively, since measured relative entropy remains the same with this identification. Also, we will restrict to $\mathcal{M}$ such that $|\mathcal{I}| \leq m$ for all $M \in \mathcal{M}$ and some $m \in \mathbb{N}$. This is not a restriction when $\mathcal{M}$ contains the set of all projective measurements (see [98]). By adding zero matrices as necessary, we may assume without loss of generality that $\mathcal{I} = \{0, \ldots, m-1\}$.

Let $M^\star(\rho, \sigma, \mathcal{M})$ denote an optimizer achieving the supremum (if it exists) in (54). The next result characterizes limit distribution for measured relative entropy estimation in terms of quantities induced by optimal measurement.

Theorem 5 (Limit distribution for measured relative entropy) *Let $\rho_n \ll \sigma_n \ll \sigma$, $\rho_n \ll \rho \ll\gg \sigma$, and $\mathcal{M}$ be compact such that $M^\star := M^\star(\rho, \sigma, \mathcal{M})$ is unique with the identification of POVMs as mentioned above. If $\big(r_n(\rho_n - \rho), r_n(\sigma_n - \sigma)\big) \xrightarrow{w} (L_1, L_2)$, then*

$$r_n\big(\mathsf{D}_{\mathcal{M}}\left(\rho_n \| \sigma_n\right) - \mathsf{D}_{\mathcal{M}}\left(\rho \| \sigma\right)\big) \xrightarrow{w} \sum_{i \in \mathcal{I}} P_{L_1,M^\star}(i) \log \frac{P_{\rho,M^\star}(i)}{P_{\sigma,M^\star}(i)} - \frac{P_{L_2,M^\star}(i) P_{\rho,M^\star}(i)}{P_{\sigma,M^\star}(i)}. \tag{55}$$

Before, we proceed with the proof of Proposition 5, a few remarks are in order. Since the definition of $\mathsf{D}_{\mathcal{M}}$ itself involves a supremum, a method of proof similar to that of Theorem 3 applies. However, a key technical difference arises due to the fact that without additional assumptions, the maximizer in (54) is not unique. Moreover, the maximizer does not have a closed form expression in general. Hence, to establish the above claim, our proof involves showing that the optimal measurement POVM of the empirical measured relative entropy, $\mathsf{D}_{\mathcal{M}}\left(\rho_n \| \sigma_n\right)$, converges to $M^\star$ in operator norm. This convergence necessitates the requirement of a unique $M^\star$ as stated above.

Proof. It suffices to show that for every subsequence of $\mathbb{N}$, there exists a further subsequence along which the convergence in (55) holds. Let n_{k_j} be a subsequence such that $\big((r_{n_{k_j}}(\rho_{n_{k_j}} - \rho), r_{n_{k_j}}(\sigma_{n_{k_j}} - \sigma)\big) \to (L_1, L_2)$ in trace-norm a.s., which exists by the same reasons stated in the proof of Theorem 1. Let $M_n^\star$ be such that

$$\mathsf{D}_{\mathcal{M}}\left(\rho_n \| \sigma_n\right) = \mathsf{D}\left(P_{\rho_n,M_n^\star} \middle\| P_{\sigma_n,M_n^\star}\right).$$

Such an $M_n^\star$ exists since $\mathcal{M}$ is compact and $\mathsf{D}\left(P_{\rho_n,M} \| P_{\sigma_n,M}\right)$ is a continuous functional of M for $\rho_n \ll \sigma_n$.

We will show that $M_n^\star \to M^\star$ a.s. in operator norm. Note that since $\rho \ll\gg \sigma$, there exists constants $c_1, c_2 > 0$ such that $c_1\sigma \leq \rho \leq c_2\sigma$. From the Choi-Kraus representation of a quantum channel, we also have $c_1 M(\sigma) \leq M(\rho) \leq c_2 M(\sigma)$. Moreover, since $\rho_{n_{k_j}} \to \rho$ and $\sigma_{n_{k_j}} \to \sigma$ in trace norm a.s., there exists constants $\tilde{c}_1, \tilde{c}_2$ (depending on the realization $\rho_{n_{k_j}}$ and $\sigma_{n_{k_j}}$) such that for all j sufficiently large, $0 < \tilde{c}_1 \leq c_1$, $0 < c_2 \leq \tilde{c}_2$, $\tilde{c}_1\sigma_{n_{k_j}} \leq \rho_{n_{k_j}} \leq \tilde{c}_2\sigma_{n_{k_j}}$ and $\tilde{c}_1 M\big(\sigma_{n_{k_j}}\big) \leq M\big(\rho_{n_{k_j}}\big) \leq \tilde{c}_2 M\big(\sigma_{n_{k_j}}\big)$. Then, denoting by A_-, the negative part of $A = A_+ - A_-$, we have

$$\begin{aligned}
\mathsf{D}\left(M(\rho) \| M(\sigma)\right) &\overset{(a)}{=} \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{s} \operatorname{Tr}\left[\big(M(\rho) - sM(\sigma)\big)_-\right] + \log \tilde{c}_2 + 1 - \tilde{c}_2 \\
&\overset{(b)}{=} \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{2s} \left(s - 1 + \left\|\big(M(\rho) - sM(\sigma)\big)\right\|_1\right) + \log \tilde{c}_2 + 1 - \tilde{c}_2,
\end{aligned}$$

where (a) follows from the integral representation of quantum relative entropy (see [99, Theorem 6] and [100, Corollary 1]), while (b) uses $\|A\|_1 = A_+ + A_-$ and $\operatorname{Tr}\left[M(\rho) - sM(\sigma)\right] = 1 - s$. Using similar representation for $\mathsf{D}\left(M(\rho_{n_{k_j}}) \middle\| M(\sigma_{n_{k_j}})\right)$ and subtracting from previous equation, we have

$$\left|\mathsf{D}\left(M(\rho) \| M(\sigma)\right) - \mathsf{D}\left(M(\rho_{n_{k_j}}) \middle\| M(\sigma_{n_{k_j}})\right)\right|$$

$$
\begin{aligned}
&= \left| \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{2s} \left\| M(\rho) - sM(\sigma) \right\|_1 - \left\| M(\rho_{n_{k_j}}) - sM(\sigma_{n_{k_j}}) \right\|_1 \right| \\
&\leq \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{2s} \left| \left\| M(\rho) - sM(\sigma) \right\|_1 - \left\| M(\rho_{n_{k_j}}) - sM(\sigma_{n_{k_j}}) \right\|_1 \right| \\
&\overset{(a)}{\leq} \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{2s} \left\| M(\rho) - M(\rho_{n_{k_j}}) - sM(\sigma) + sM(\sigma_{n_{k_j}}) \right\|_1 \\
&\overset{(b)}{\leq} \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{2s} \left(\left\| M(\rho) - M(\rho_{n_{k_j}}) \right\|_1 + s \left\| M(\sigma) - M(\sigma_{n_{k_j}}) \right\|_1 \right) \\
&\overset{(c)}{\leq} \int_{\tilde{c}_1}^{\tilde{c}_2} \frac{ds}{2s} \left(\left\| \rho - \rho_{n_{k_j}} \right\|_1 + s \left\| \sigma - \sigma_{n_{k_j}} \right\|_1 \right),
\end{aligned}
$$

where (a) and (b) follows from triangle inequality, while (c) is due to data processing inequality for trace norm. By setting $h_n(M) = \mathsf{D}\left(M(\rho_n) \| M(\sigma_n)\right)$, it follows from the above inequality that $h_{n_{k_j}}(M)$ converges uniformly to $\mathsf{D}\left(M(\rho) \| M(\sigma)\right)$ (as a function of M) given $\rho_{n_{k_j}} \to \rho$ and $\sigma_{n_{k_j}} \to \sigma$. Also, note that for any $M, \tilde{M} \in \mathcal{M}$, we have via similar steps as above that

$$
\begin{aligned}
\left| \mathsf{D}\left(M(\rho) \| M(\sigma)\right) - \mathsf{D}\big(\tilde{M}(\rho) \| \tilde{M}(\sigma)\big) \right| &\leq \int_{c_1}^{c_2} \frac{ds}{2s} \left(\left\| M(\rho) - \tilde{M}(\rho) \right\|_1 + s \left\| M(\sigma) - \tilde{M}(\sigma) \right\|_1 \right) \\
&\leq \int_{c_1}^{c_2} \frac{ds}{2s} \left\| M - \tilde{M} \right\| (s+1) \\
&\leq \frac{\left\| M - \tilde{M} \right\|}{2} \left(\ln \frac{c_2}{c_1} + c_2 - c_1 \right),
\end{aligned}
$$

where $\|\cdot\|$ in the last two inequalities denotes the operator norm. Hence, $\mathsf{D}\left(M(\rho) \| M(\sigma)\right)$ is a uniformly continuous functional of M in operator norm. Since $\mathcal{M}$ is compact and $M^\star = \arg\max \mathsf{D}_{\mathcal{M}}\left(\rho \| \sigma\right)$ is unique, the aforementioned uniform convergence implies that $M^\star_{n_{k_j}} \to M^\star$ a.s. in operator norm.

Equipped with the above, we next prove (55). Note that by definition of $M_n^\star$ and $M^\star$, we have

$$
r_n \left(\mathsf{D}_{\mathcal{M}}\left(\rho_n \| \sigma_n\right) - \mathsf{D}_{\mathcal{M}}\left(\rho \| \sigma\right) \right) \leq r_n \left(\mathsf{D}\left(P_{\rho_n, M_n^\star} \middle\| P_{\sigma_n, M_n^\star}\right) - \mathsf{D}\left(P_{\rho, M_n^\star} \middle\| P_{\sigma, M_n^\star}\right) \right), \tag{56a}
$$

$$
r_n \left(\mathsf{D}_{\mathcal{M}}\left(\rho_n \| \sigma_n\right) - \mathsf{D}_{\mathcal{M}}\left(\rho \| \sigma\right) \right) \geq r_n \left(\mathsf{D}\left(P_{\rho_n, M^\star} \| P_{\sigma_n, M^\star}\right) - \mathsf{D}\left(P_{\rho, M^\star} \| P_{\sigma, M^\star}\right) \right). \tag{56b}
$$

Denoting the LHS and RHS of (55) by $g_n(\rho_n, \sigma_n)$ and $g(L_1, L_2)$, respectively, we will show that along the subsequence $(n_{k_j})_{j \in \mathbb{N}}$, the RHS of (56a) and (56b) converge a.s. to $g(L_1, L_2)$. This then implies that $g_{n_{k_j}} \to g(L_1, L_2)$ a.s. and also weakly.

From Taylor expansion applied to the function $f(x,y) = x \log \frac{x}{y}$, we have[3] for $u, v, \bar{u}, \bar{v}$ such that $u \geq 0, v, \bar{u}, \bar{v} > 0$ or $u = 0, v \geq 0, \bar{u}, \bar{v} > 0$ that

$$
\begin{aligned}
u \log \frac{u}{v} = \bar{u} \log \frac{\bar{u}}{\bar{v}} + \left(1 + \log \frac{\bar{u}}{\bar{v}}\right)(u - \bar{u}) - \frac{\bar{u}}{\bar{v}}(v - \bar{v}) + \int_0^1 \frac{(1-\tau)(u-\bar{u})^2}{(1-\tau)\bar{u} + \tau u} d\tau \\
+ \int_0^1 \frac{(1-\tau)((1-\tau)\bar{u} + \tau u)(v - \bar{v})^2}{\left((1-\tau)\bar{v} + \tau v\right)^2} d\tau - 2 \int_0^1 \frac{(1-\tau)(u-\bar{u})(v-\bar{v})}{\left((1-\tau)\bar{v} + \tau v\right)^2} d\tau.
\end{aligned} \tag{57}
$$

[3]The conditions required for validity of Taylor's expansion requires $u, v, \bar{u}, \bar{v} > 0$, but as noted in the proof of [22, Proposition 1], this expansion is also valid for all $u, v, \bar{u}, \bar{v}$ as mentioned above.

Let $u = P_{\rho_n,M_n^\star}(i)$, $v = P_{\sigma_n,M_n^\star}(i)$, $\bar{u} = P_{\rho,M_n^\star}(i)$, $\bar{v} = P_{\sigma,M_n^\star}(i)$ for $i \in \mathcal{I}$ and note that the aforementioned constraints on $u, v, \bar{u}, \bar{v}$ are satisfied since $P_{\rho_n,M_n^\star} \ll P_{\rho,M_n^\star} \ll P_{\sigma,M_n^\star}$ and $P_{\rho_n,M_n^\star} \ll P_{\sigma_n,M_n^\star} \ll P_{\sigma,M_n^\star}$ due to $\rho_n \ll \sigma_n \ll \sigma$ and $\rho_n \ll \rho \ll \sigma$. Substituting the above and summing the resulting expression over i in the above equation leads to

$$\begin{aligned}
f_n &:= r_n\big(\mathsf{D}\left(M_n^\star(\rho_n)\|M_n^\star(\sigma_n)\right) - \mathsf{D}\left(M_n^\star(\rho)\|M_n^\star(\sigma)\right)\big) \\
&= \sum_{i\in\mathcal{I}} r_n(P_{\rho_n,M_n^\star}(i) - P_{\rho,M_n^\star}(i)) \log \frac{P_{\rho,M_n^\star}(i)}{P_{\sigma,M_n^\star}(i)} - \sum_{i\in\mathcal{I}} r_n(P_{\sigma_n,M_n^\star}(i) - P_{\sigma,M_n^\star}(i)) \frac{P_{\rho,M_n^\star}(i)}{P_{\sigma,M_n^\star}(i)} + R_n,
\end{aligned}$$

where $R_n = R_{n,1} + R_{n,2} - 2R_{n,3}$, and

$$\begin{aligned}
R_{n,1} &:= \sum_{i\in\mathcal{I}} \int_0^1 \frac{(1-\tau)\Big(r_n^{\frac{1}{2}}\left(P_{\rho_n,M_n^\star}(i) - P_{\rho,M_n^\star}(i)\right)\Big)^2}{(1-\tau)P_{\rho,M_n^\star}(i) + \tau P_{\rho_n,M_n^\star}(i)} d\tau, \\
R_{n,2} &:= \sum_{i\in\mathcal{I}} \int_0^1 \frac{(1-\tau)((1-\tau)P_{\rho,M_n^\star}(i) + \tau P_{\rho_n,M_n^\star}(i))\Big(r_n^{\frac{1}{2}}\left((P_{\sigma_n,M_n^\star}(i) - P_{\sigma,M_n^\star}(i)\right)\Big)^2}{\left((1-\tau)P_{\sigma_n,M_n^\star}(i) + \tau P_{\sigma,M_n^\star}(i)\right)^2} d\tau, \\
R_{n,3} &:= \sum_{i\in\mathcal{I}} \int_0^1 \frac{(1-\tau) r_n^{\frac{1}{2}}\left(P_{\rho_n,M_n^\star}(i) - P_{\rho,M_n^\star}(i)\right) r_n^{\frac{1}{2}}\left(P_{\sigma_n,M_n^\star}(i) - P_{\sigma,M_n^\star}(i)\right)}{\left((1-\tau)P_{\sigma,M_n^\star}(i) + \tau P_{\sigma_n,M_n^\star}(i)\right)^2} d\tau.
\end{aligned}$$

Note that since $M^\star_{n_{k_j}} \to M^\star$ in operator norm a.s. and $\big((r_{n_{k_j}}(\rho_{n_{k_j}} - \rho), r_{n_{k_j}}(\sigma_{n_{k_j}} - \rho)\big) \to (L_1, L_2)$ in trace-norm a.s., we have $P_{\rho_{n_{k_j}},M^*_{n_{k_j}}}(i) \to P_{\rho,M^\star}(i)$, $P_{\sigma_{n_{k_j}},M^*_{n_{k_j}}}(i) \to P_{\sigma,M^\star}(i)$, and

$$\begin{aligned}
&r_{n_{k_j}}\left(P_{\rho_{n_{k_j}},M^*_{n_{k_j}}}(i) - P_{\rho,M^*_{n_{k_j}}}(i)\right) = \mathrm{Tr}\left[M^*_{n_{k_j}}(i) r_{n_{k_j}}\left(\rho_{n_{k_j}} - \rho\right)\right] \to \mathrm{Tr}\left[M^\star(i)L_1\right], \\
&r_{n_{k_j}}\left(P_{\sigma_{n_{k_j}},M^*_{n_{k_j}}}(i) - P_{\sigma,M^*_{n_{k_j}}}(i)\right) = \mathrm{Tr}\left[M^*_{n_{k_j}}(i) r_{n_{k_j}}\left(\sigma_{n_{k_j}} - \sigma\right)\right] \to \mathrm{Tr}\left[M^\star(i)L_2\right], \\
&r^{\frac{1}{2}}_{n_{k_j}}\left(P_{\rho_{n_{k_j}},M^*_{n_{k_j}}}(i) - P_{\rho,M^*_{n_{k_j}}}(i)\right) \to 0, \quad \text{and} \quad r^{\frac{1}{2}}_{n_{k_j}}\left(P_{\sigma_{n_{k_j}},M^*_{n_{k_j}}}(i) - P_{\sigma,M^*_{n_{k_j}}}(i)\right) \to 0.
\end{aligned}$$

Hence, we have $R_{n_{k_j}} \to 0$ and $f_{n_{k_j}} \to g(L_1, L_2)$. This implies that the RHS of (56a) converges to $g(L_1, L_2)$. Likewise, the RHS of (56b) also converges a.s. to $g(L_1, L_2)$ along the sequence $(n_{k_j})_{j\in\mathbb{N}}$, which implies that $g_{n_{k_j}} \to g(L_1, L_2)$ a.s, and hence also weakly as claimed. This completes the proof of the proposition. □

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)*, 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] Q. Wang, S. R. Kulkarni, and S. Verdu, "Divergence estimation of continuous distributions based on data-dependent partitions," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3064–3074, Sep. 2005.

[28] 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.

[29] F. Perez-Cruz, "Kullback-Leibler divergence estimation of continuous distributions," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1666–1670.

[30] 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.

[31] 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.

[32] M. Noshad, K. R. Moon, S. Y. Sekeh, and A. O. Hero, "Direct estimation of information divergence using nearest neighbor ratios," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 903–907.

[33] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, "Mutual information neural estimation," in *Proc. Int. Conf. Mach. Learn.*, vol. 80, Stockholm Sweden, Jul. 2018, pp. 531–540.

[34] K. R. Moon, K. Sricharan, K. Greenewald, and A. O. Hero, "Ensemble estimation of information divergence," *Entropy*, vol. 20, no. 8, Aug. 2018.

[35] 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.

[36] 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.

[37] Y. Han, J. Jiao, T. Weissman, and Y. Wu, "Optimal rates of entropy estimation over Lipschitz balls," *Ann. Stat.*, vol. 48, no. 6, pp. 3228–3250, Dec. 2020.

[38] S. Sreekumar, Z. Zhang, and Z. Goldfeld, "Non-asymptotic performance guarantees for neural estimation of f-divergences," in *AISTATS*, 2021.

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

[40] A. Antos and I. Kontoyiannis, "Convergence properties of functional estimates for discrete distributions," *Random Struct. Algorithms*, vol. 19, no. 3-4, pp. 163–193, 2001.

[41] Z. Zhang and X. Zhang, "A normal law for the plug-in estimator of entropy," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2745–2747, 2012.

[42] I. Kontoyiannis and M. Skoularidou, "Estimating the directed information and testing for causality," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6053–6067, 2016.

[43] 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.

[44] 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.

[45] E. del Barrio, E. Giné, and C. Matrán, "Central limit theorems for the Wasserstein distance between the empirical and the true distributions," *Ann. Probab.*, vol. 27, no. 2, pp. 1009–1071, 1999.

[46] E. del Barrio, E. Giné, and F. Utzet, "Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances," *Bernoulli*, vol. 11, no. 1, pp. 131–189, Jan. 2005.

[47] M. Sommerfeld and A. Munk, "Inference for empirical Wasserstein distances on finite spaces," *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)*, vol. 80, 2016.

[48] E. del Barrio and J. Loubes, "Central limit theorems for empirical transportation cost in general dimension," *Ann. Probab.*, vol. 47, no. 2, pp. 926–951, 2019.

[49] C. Tameling, M. Sommerfeld, and A. Munk, "Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications," *Ann. Appl. Probab.*, vol. 29, no. 5, pp. 2744 – 2781, 2019.

[50] E. del Barrio, A. G. Sanz, J.-M. Loubes, and J. Niles-Weed, "An improved central limit theorem and fast convergence rates for entropic transportation costs," *SIAM J. Math. of Data Sci.*, vol. 5, no. 3, pp. 639–669, 2023.

[51] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman, "Plugin estimation of smooth optimal transport maps," *arXiv:2107.12364*, 2021.

[52] S. Hundrieser, M. Klatt, T. Staudt, and A. Munk, "A unifying approach to distributional limits for empirical optimal transport," *arXiv:2202.12790*, 2022.

[53] J. Bigot, E. Cazelles, and N. Papadakis, "Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications," *Electron. J. Stat.*, vol. 13, no. 2, pp. 5120 – 5150, 2019.

[54] G. Mena and J. Niles-Weed, "Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32. Curran Associates, Inc., 2019.

[55] M. Klatt, C. Tameling, and A. Munk, "Empirical regularized optimal transport: Statistical theory and applications," *SIAM J. Math. Data Sci.*, vol. 2, no. 2, pp. 419–443, 2020.

[56] Z. Harchaoui, L. Liu, and S. Pal, "Asymptotics of entropy-regularized optimal transport via chaos decomposition," *arXiv: 2011.08963*, 2020.

[57] Z. Goldfeld, K. Greenewald, and K. Kato, "Asymptotic guarantees for generative modeling based on the smooth Wasserstein distance," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, Dec. 2020.

[58] F. Gunsilius and Y. Xu, "Matching for causal effects via multimarginal unbalanced optimal transport," *arXiv:2112.04398*, 2021.

[59] A. Gonzalez-Sanz, J. Loubes, and J. Niles-Weed, "Weak limits of entropy regularized optimal transport; potentials, plans and divergences," *arXiv:2207.07427*, 2022.

[60] R. Sadhu, Z. Goldfeld, and K. Kato, "Limit distribution theory for the smooth 1-Wasserstein distance with applications," *arXiv:2107.13494*, 2021.

[61] Z. Goldfeld, K. Kato, S. Nietert, and G. Rioux, "Limit distribution theory for smooth p-Wasserstein distances," *arXiv:2203.00159*, 2022.

[62] Z. Goldfeld, K. Kato, G. Rioux, and R. Sadhu, "Limit theorems for entropic optimal transport maps and the Sinkhorn divergence," *arXiv:2207.08683*, 2022.

[63] ——, “Statistical inference with regularized optimal transport,” *arXiv:2205.04283*, 2022.

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

[65] Y. Subasi, L. Cincio, and P. J. Coles, “Entanglement spectroscopy with a depth-two quantum circuit,” *J. Phys. A: Math. Theor.*, vol. 52, no. 4, p. 44001, Mar. 2019.

[66] 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.

[67] 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.

[68] A. Gilyén and T. Li, “Distributional property testing in a quantum world,” in *11th Innov. in Theor. Comput. Sci. Conf. (ITCS)*, vol. 151. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020, pp. 25:1–25:19.

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

[70] R. Chen, Z. Song, X. Zhao, and X. Wang, “Variational quantum algorithms for trace distance and fidelity estimation,” *Quantum Science and Technology*, vol. 7, no. 1, pp. 015–019, 2021.

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

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

[73] Q. Wang, J. Guan, J. Liu, Z. Zhang, and M. Ying, “New quantum algorithms for computing quantum entropies and distances,” Mar. 2022, arXiv:2203.13522.

[74] A. Gilyén and A. Poremba, “Improved quantum algorithms for fidelity estimation,” Mar. 2022, arXiv:2203.15993.

[75] Y. Wang, L. Zhang, Z. Yu, and X. Wang, “Quantum phase processing and its applications in estimating phase and entropies,” *Phys. Rev. A*, vol. 108, p. 062413, Dec 2023.

[76] 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.

[77] Z. Goldfeld, D. Patel, S. Sreekumar, and M. M. Wilde, “Quantum neural estimation of entropies,” Jul. 2023, arXiv:2307.01171.

[78] Y. Quek, E. Kaur, and M. M. Wilde, “Multivariate trace estimation in constant quantum depth,” Jun. 2022, arXiv:2206.15405.

[79] 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.

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

[81] 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.

[82] 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.

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

[84] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.

[85] A. Uhlmann, “The ‘transition probability’ in the state space of a *-algebra,” *Rep. Math. Phys.*, vol. 9, pp. 273–279, 1976.

[86] R. Jozsa, “Fidelity for mixed quantum states,” *J. Mod. Opt.*, vol. 41, no. 12, pp. 2315–2323, 1994.

[87] M. J. Donald, “On the relative entropy,” *Commun. Math. Phys.*, vol. 105, no. 1, pp. 13–34, 1986.

[88] 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.

[89] E. Rains, “A semidefinite program for distillable entanglement,” *IEEE Trans. Inf. Theory*, vol. 47, pp. 2921 – 2933, 12 2001.

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

[91] 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.

[92] 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.

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

[94] S. Bochner, “Integration von funktionen, deren werte die elemente eines vektorraumes sind,” *Fundamenta Mathematicae*, vol. 20, p. 262–276, 1933.

[95] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*. Springer, New York, 1996.

[96] E. Carlen, “Trace inequalities and quantum entropy: An introductory course,” in *Entropy and the quantum*, ser. Contemp. Math. Amer. Math. Soc., Providence, RI, 2010, vol. 529, pp. 73–140.

[97] M. Berta, V. Scholz, and M. Tomamichel, “Rényi divergences as weighted non-commutative vector valued L_p-spaces,” *Annales Henri Poincaré*, vol. 19, p. 1843–1867, 03 2018.

[98] M. Berta, O. Fawzi, and M. Tomamichel, “On variational expressions for quantum relative entropies,” *Lett. Math. Phys.*, vol. 107, no. 12, pp. 2239–2265, Dec. 2015.

[99] P. E. Frenkel, “Integral formula for quantum relative entropy implies data processing inequality,” *Quantum*, vol. 7, p. 1102, Sep. 2023.

[100] A. Jenčová, “Recoverability of quantum channels via hypothesis testing,” *arxiv: 2303.11707*, 2023.