QUANTUM LOCAL ASYMPTOTIC NORMALITY BASED ON A NEW QUANTUM LIKELIHOOD RATIO

BY KOICHI YAMAGATA, AKIO FUJIWARA AND RICHARD D. GILL

Osaka University, Osaka University and Leiden University

We develop a theory of local asymptotic normality in the quantum domain based on a novel quantum analogue of the log-likelihood ratio. This formulation is applicable to any quantum statistical model satisfying a mild smoothness condition. As an application, we prove the asymptotic achievability of the Holevo bound for the local shift parameter.

1. Introduction. Suppose that one has n copies of a quantum system each in the same state depending on an unknown parameter θ, and one wishes to estimate θ by making some measurement on the n systems together. This yields data whose distribution depends on θ and on the choice of the measurement. Given the measurement, we therefore have a classical parametric statistical model, though not necessarily an i.i.d. model, since we are allowed to bring the n systems together before measuring the resulting joint system as one quantum object. In that case the resulting data need not consist of (a function of) n i.i.d. observations, and a key quantum feature is that we can generally extract more information about θ using such "collective" or "joint" measurements than when we measure the systems separately. What is the best we can do as $n \to \infty$, when we are allowed to optimise both over the measurement and over the ensuing data processing? The objective of this paper is to study this question by extending the theory of local asymptotic normality (LAN), which is known to form an important part of the classical asymptotic theory, to quantum statistical models.

Let us recall the classical LAN theory first. Given a statistical model $\mathcal{S} = \{p_\theta; \theta \in \Theta\}$ on a probability space $(\Omega, \mathcal{F}, \mu)$ indexed by a parameter θ that ranges over an open subset Θ of $\mathbb{R}^d$, let us introduce a local parameter $h := \sqrt{n}(\theta - \theta_0)$ around a fixed $\theta_0 \in \Theta$. If the parametrisation $\theta \mapsto p_\theta$ is sufficiently smooth, it is known that the statistical properties of the model $\{p_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^d\}$ is similar to that of the Gaussian shift model $\{N(h, J_{\theta_0}^{-1}); h \in \mathbb{R}^d\}$ for large n, where $p_\theta^{\otimes n}$ is the nth i.i.d. extension of p_θ, and J_{θ_0} is the Fisher information matrix of the model p_θ at θ_0. This property is called the local asymptotic normality of the model $\mathcal{S}$ [21].

More generally, a sequence $\{p_\theta^{(n)}; \theta \in \Theta \subset \mathbb{R}^d\}$ of statistical models on $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mu^{(n)})$ is called *locally asymptotically normal* (LAN) at $\theta_0 \in \Theta$ if there

exist a $d \times d$ positive matrix J and random vectors $\Delta^{(n)} = (\Delta_1^{(n)}, \ldots, \Delta_d^{(n)})$ such that $\Delta^{(n)} \overset{0}{\rightsquigarrow} N(0, J)$ and

$$\log \frac{p_{\theta_0+h/\sqrt{n}}^{(n)}}{p_{\theta_0}^{(n)}} = h^i \Delta_i^{(n)} - \frac{1}{2} h^i h^j J_{ij} + o_{p_{\theta_0}}(1)$$

for all $h \in \mathbb{R}^d$. Here the arrow $\overset{h}{\rightsquigarrow}$ stands for the convergence in distribution under $p_{\theta_0+h/\sqrt{n}}^{(n)}$, the remainder term $o_{p_{\theta_0}}(1)$ converges in probability to zero under $p_{\theta_0}^{(n)}$, and Einstein's summation convention is used. The above expansion is similar in form to the log-likelihood ratio of the Gaussian shift model:

$$\log \frac{dN(h, J^{-1})}{dN(0, J^{-1})}(X^1, \ldots, X^d) = h^i (X^j J_{ij}) - \frac{1}{2} h^i h^j J_{ij}.$$

This is the underlying mechanism behind the statistical similarities between models $\{p_{\theta_0+h/\sqrt{n}}^{(n)}; h \in \mathbb{R}^d\}$ and $\{N(h, J^{-1}); h \in \mathbb{R}^d\}$.

In order to put the similarities to practical use, one needs some mathematical devices. In general, a statistical theory comprises two parts. One is to prove the existence of a statistic that possesses a certain desired property (direct part), and the other is to prove the nonexistence of a statistic that exceeds that property (converse part). In the problem of asymptotic efficiency, for example, the converse part, the impossibility to do asymptotically better than the best which can be done in the limit situation, is ensured by the following proposition, which is usually referred to as "Le Cam's third lemma" [21].

PROPOSITION 1.1. *Suppose* $\{p_\theta^{(n)}; \theta \in \Theta \subset \mathbb{R}^d\}$ *is LAN at* $\theta_0 \in \Theta$, *with* $\Delta^{(n)}$ *and* J *being as above, and let* $X^{(n)} = (X_1^{(n)}, \ldots, X_r^{(n)})$ *be a sequence of random vectors. If the joint distribution of* $X^{(n)}$ *and* $\Delta^{(n)}$ *converges to a Gaussian distribution, in that*

$$\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix} \overset{0}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ {}^t\tau & J \end{pmatrix} \right),$$

then $X^{(n)} \overset{h}{\rightsquigarrow} N(\tau h, \Sigma)$ *for all* $h \in \mathbb{R}^d$. *Here* ${}^t\tau$ *stands for the transpose of* τ.

Now, it appears from this lemma that it already tells us something about the direct problem. In fact, by putting $X^{(n)j} := \sum_{k=1}^d [J^{-1}]^{jk} \Delta_k^{(n)}$, we have

$$\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix} \overset{0}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J^{-1} & I \\ I & J \end{pmatrix} \right),$$

so that $X^{(n)} \overset{h}{\rightsquigarrow} N(h, J^{-1})$ follows from Proposition 1.1. This proves the existence of an asymptotically efficient estimator for h. In the real world, however, we do

not know θ_0 (obviously). Thus, the existence of an asymptotically optimal estimator for h does not translate into the existence of an asymptotically optimal estimator of θ. In fact, the usual way that Le Cam's third lemma is used in the subsequent analysis is in order to prove the so-called representation theorem, [21], Theorem 7.10. This theorem can be used to tell us in several precise mathematical senses that no estimator can asymptotically do better than what can be achieved in the limiting Gaussian model.

For instance, Van der Vaart's version of the representation theorem leads to the asymptotic minimax theorem, telling us that the worst behaviour of an estimator as θ varies in a shrinking (1 over root n) neighbourhood of θ_0 cannot improve on what we expect from the limiting problem. This theorem applies to *all* possible estimators, but only discusses their *worst* behaviour in a neighbourhood of θ. Another option is to use the representation theorem to derive the convolution theorem, which tells us that *regular* estimators (estimators whose asymptotic behaviour in a small neighbourhood of θ is more or less stable as the parameter varies) have a limiting distribution which in a very strong sense is more disperse than the optimal limiting distribution which we expect from the limiting statistical problem.

This paper addresses a quantum extension of LAN (abbreviated as QLAN). As in the classical statistics, one of the important subjects of QLAN is to show the existence of an estimator (direct part) that enjoys certain desired properties. Some earlier works of asymptotic quantum parameter estimation theory revealed the asymptotic achievability of the Holevo bound, a quantum extension of the Cramér–Rao type bound (see Section B.1 and B.2 in [23]). Using a group representation theoretical method, Hayashi and Matsumoto [11] showed that the Holevo bound for the quantum statistical model $\mathcal{S}(\mathbb{C}^2) = \{\rho_\theta; \theta \in \Theta \subset \mathbb{R}^3\}$ comprising the totality of density operators on the Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$ is asymptotically achievable at a given single point $\theta_0 \in \Theta$. Following their work, Guţă and Kahn [9, 14] developed a theory of strong QLAN, and proved that the Holevo bound is asymptotically uniformly achievable around a given $\theta_0 \in \Theta$ for the quantum statistical model $\mathcal{S}(\mathbb{C}^D) = \{\rho_\theta; \theta \in \Theta \subset \mathbb{R}^{D^2-1}\}$ comprising the totality of density operators on the finite dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^D$. They proved that an i.i.d. model $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^{D^2-1}\}$ and a certain quantum Gaussian shift model can be translated by quantum channels to each other asymptotically. Although their result is powerful, their QLAN has several drawbacks. First of all, their method works only for i.i.d. extension of the totality $\mathcal{S}(\mathcal{H})$ of the quantum states on the Hilbert space $\mathcal{H}$, and is not applicable to generic submodels of $\mathcal{S}(\mathcal{H})$. Moreover, it makes use of a special parametrisation θ of $\mathcal{S}(\mathcal{H})$, in which the change of eigenvalues and eigenvectors are treated as essential. Furthermore, it does not work if the reference state ρ_{θ_0} has a multiplicity of eigenvalues. Since these difficulties are inevitable in representation theoretical approach advocated by Hayashi and Matsumoto [11], Guţă and Jençová [8] also tried a different approach to QLAN via the Connes cocycle derivative, which was put forward in the literature as an appropriate quantum analogue of the likelihood ratio. However they did not formally establish an

expansion which would be directly analogous to the classical LAN. In addition, their approach is limited to faithful state models.

The purpose of the present paper is to develop a theory of weak QLAN based on a new quantum extension of the log-likelihood ratio. This formulation is applicable to any quantum statistical model satisfying a mild smoothness condition, and is free from artificial setups such as the use of a special coordinate system and/or nondegeneracy of eigenvalues of the reference state at which QLAN works. We also prove asymptotic achievability of the Holevo bound for the local shift parameter h that belong to a dense subset of $\mathbb{R}^d$.

This paper is organised as follows. The main results are summarised in Section 2. We first introduce a novel type of quantum log-likelihood ratio, and define a quantum extension of local asymptotic normality in a quite analogous way to the classical LAN. We then explore some basic properties of QLAN, including a sufficient condition for an i.i.d. model to be QLAN, and a quantum extension of Le Cam's third lemma. Section 3 is devoted to application of QLAN, including the asymptotic achievability of the Holevo bound and asymptotic estimation theory for some typical qubit models. Proofs of main results are deferred to Section A of supplementary material [23]. Furthermore, since we assume some basic knowledge of quantum estimation theory throughout the paper, we provide, for the reader's convenience, a brief exposition of quantum estimation theory in Section B of supplementary material [23], including quantum logarithmic derivatives, the commutation operator and the Holevo bound (Section B.1), estimation theory for quantum Gaussian shift models (Section B.2), and estimation theory for pure state models (Section B.3).

It is also important to notice the limits of this work, which means that there are many open problems left to study in the future. In the classical case, the theory of LAN builds, of course, on the rich theory of convergence in distribution, as studied in probability theory. In the quantum case, there still does not exist a full parallel theory. Some of the most useful lemmas in the classical theory simply are not true when translated in the quantum domain. For instance, in the classical case, we know that if the sequence of random variables X_n converges in distribution to a random variable X, and at the same time the sequence Y_n converges in probability to a constant c, then this implies joint convergence in distribution of (X_n, Y_n) to the pair (X, c). The obvious analogue of this in the quantum domain is simply untrue. In fact, there is not even a general theory of convergence in distribution at all: there is only a theory of convergence in distribution toward quantum Gaussian limits. Unfortunately, even in this special case the natural analogue of the just mentioned result simply fails to be true.

Because of these obstructions we are not at present able to follow the standard route from Le Cam's third lemma to the representation theorem, and from there to asymptotic minimax or convolution theorems.

However we believe that the paper presents some notable steps in this direction. Moreover, just as with Le Cam's third lemma, one is able to use the lemma to

construct what can be conjectured to be asymptotically optimal measurement and estimation schemes. We make some more remarks on these possibilities later in the paper.

2. Main results.

2.1. *Quantum log-likelihood ratio.* In developing the theory of QLAN, it is crucial what quantity one should adopt as the quantum counterpart of the likelihood ratio. One may conceive of the Connes cocycle

$$[D\sigma, D\rho]_t := \sigma^{\sqrt{-1}t} \rho^{-\sqrt{-1}t}$$

as *the* proper counterpart since it plays an essential role in discussing the sufficiency of a subalgebra in quantum information theory [20]. Nevertheless, we shall take a different route to the theory of QLAN, paying attention to the fact that a "quantum exponential family"

$$\rho_\theta = e^{(1/2)(\theta L - \psi(\theta)I)} \rho_0 e^{(1/2)(\theta L - \psi(\theta)I)}$$

inherits nice properties of the classical exponential family [1, 2].

DEFINITION 2.1 (Quantum log-likelihood ratio). We say a pair of density operators ρ and σ on a finite dimensional Hilbert space $\mathcal{H}$ are *mutually absolutely continuous, $\rho \sim \sigma$* in symbols, if there exist a Hermitian operator $\mathcal{L}$ that satisfies

$$\sigma = e^{(1/2)\mathcal{L}} \rho e^{(1/2)\mathcal{L}}.$$

We shall call such a Hermitian operator $\mathcal{L}$ a *quantum log-likelihood ratio*. When the reference states ρ and σ need to be specified, $\mathcal{L}$ shall be denoted by $\mathcal{L}(\sigma|\rho)$, so that

$$\sigma = e^{(1/2)\mathcal{L}(\sigma|\rho)} \rho e^{(1/2)\mathcal{L}(\sigma|\rho)}.$$

We use the convention that $\mathcal{L}(\rho|\rho) = 0$.

EXAMPLE 2.2. We say a state on $\mathcal{H} \simeq \mathbb{C}^d$ is *faithful* if its density operator is positive definite. Any two faithful states are always mutually absolutely continuous, and the corresponding quantum log-likelihood ratio is unique. In fact, given $\rho > 0$ and $\sigma > 0$, they are related as $\sigma = e^{(1/2)\mathcal{L}(\sigma|\rho)} \rho e^{(1/2)\mathcal{L}(\sigma|\rho)}$, where

$$\mathcal{L}(\sigma|\rho) = 2\log\left(\sqrt{\rho^{-1}}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\sqrt{\rho^{-1}}\right).$$

Note that $\operatorname{Tr}\rho e^{(1/2)\mathcal{L}(\sigma|\rho)}$ is identical to the fidelity between ρ and σ, and $e^{(1/2)\mathcal{L}(\sigma|\rho)}$ is nothing but the operator geometric mean $\rho^{-1}\#\sigma$, where $A\#B := A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ for positive operators A, B [15]. Since $A\#B = B\#A$, the quantum log-likelihood ratio can also be written as

$$\mathcal{L}(\sigma|\rho) = 2\log\left(\sqrt{\sigma}(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}})^{-1}\sqrt{\sigma}\right).$$

EXAMPLE 2.3. Pure states $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\xi\rangle\langle\xi|$ are mutually absolutely continuous if and only if $\langle\xi|\psi\rangle \neq 0$. In fact, the "only if" part is obvious. For the "if" part, consider $\mathcal{L}(\sigma|\rho) := 2\log R$ where

$$R := I + \frac{1}{|\langle\xi|\psi\rangle|}|\xi\rangle\langle\xi| - |\psi\rangle\langle\psi|.$$

Now

$$e^{(1/2)\mathcal{L}(\sigma|\rho)}|\psi\rangle = R|\psi\rangle = \frac{\langle\xi|\psi\rangle}{|\langle\xi|\psi\rangle|}|\xi\rangle,$$

showing that $\rho \sim \sigma$.

REMARK 2.4. In general, density operators ρ and σ are mutually absolutely continuous if and only if

$$(2.1) \qquad \sigma\rfloor_{\mathrm{supp}\,\rho} > 0 \quad \text{and} \quad \mathrm{rank}\,\rho = \mathrm{rank}\,\sigma,$$

where $\sigma\rfloor_{\mathrm{supp}\,\rho}$ denotes the "excision" of σ, the operator on the subspace $\mathrm{supp}\,\rho := (\ker\rho)^\perp$ of $\mathcal{H}$ defined by

$$\sigma\rfloor_{\mathrm{supp}\,\rho} := \iota_\rho^* \sigma \iota_\rho,$$

where $\iota_\rho : \mathrm{supp}\,\rho \hookrightarrow \mathcal{H}$ is the inclusion map. In fact, the "only if" part is immediate. To prove the "if" part, let ρ and σ be represented in the form of block matrices

$$\rho = \begin{pmatrix} \rho_0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \sigma = \begin{pmatrix} \sigma_0 & \alpha \\ \alpha^* & \beta \end{pmatrix}$$

with $\rho_0 > 0$. Since the first condition in (2.1) is equivalent to $\sigma_0 > 0$, the matrix σ is further decomposed as

$$\sigma = E^* \begin{pmatrix} \sigma_0 & 0 \\ 0 & \beta - \alpha^*\sigma_0^{-1}\alpha \end{pmatrix} E, \qquad E := \begin{pmatrix} I & \sigma_0^{-1}\alpha \\ 0 & I \end{pmatrix},$$

and the second condition in (2.1) turns out to be equivalent to $\beta - \alpha^*\sigma_0^{-1}\alpha = 0$. Now let $\mathcal{L}(\sigma|\rho) := 2\log R$, where

$$R := E^* \begin{pmatrix} \rho_0^{-1}\#\sigma_0 & 0 \\ 0 & \gamma \end{pmatrix} E$$

with γ being an arbitrary positive matrix. Then a simple calculation shows that $\sigma = R\rho R$.

The above argument demonstrates that a quantum log-likelihood ratio, if it exists, is not unique when the reference states are not faithful. To be precise, the operator $e^{(1/2)\mathcal{L}(\sigma|\rho)}$ is determined up to an additive constant Hermitian operator K satisfying $\rho K = 0$. This fact also proves that the quantity $\mathrm{Tr}\,\rho e^{(1/2)\mathcal{L}(\sigma|\rho)}$ is well defined regardless of the uncertainty of $\mathcal{L}(\sigma|\rho)$, and is identical to the fidelity.

2.2. *Quantum central limit theorem.* In quantum mechanics, canonical observables are represented by the following canonical commutation relations (CCR):

$$[Q_i, P_j] = \sqrt{-1}\hbar\delta_{ij}I, \qquad [Q_i, Q_j] = 0, \qquad [P_i, P_j] = 0,$$

where $\hbar$ is the Planck constant. In what follows, we shall treat a slightly generalised form of the CCR:

$$\frac{\sqrt{-1}}{2}[X_i, X_j] = S_{ij}I \qquad (1 \leq i, j \leq d),$$

where $S = [S_{ij}]$ is a $d \times d$ real skew-symmetric matrix. The algebra generated by the observables $(X_1, \ldots, X_d)$ is denoted by $\text{CCR}(S)$, and $X := (X_1, \ldots, X_d)$ is called the basic canonical observables of the algebra $\text{CCR}(S)$. (See [12, 13, 16, 19] for a rigorous definition of the CCR algebra.)

A state ϕ on the algebra $\text{CCR}(S)$ is characterised by the *characteristic function*

$$\mathcal{F}_\xi\{\phi\} := \phi\big(e^{\sqrt{-1}\xi^i X_i}\big),$$

where $\xi = (\xi^i)_{i=1}^d \in \mathbb{R}^d$ and Einstein's summation convention is used. A state ϕ on $\text{CCR}(S)$ is called a *quantum Gaussian state*, denoted by $\phi \sim N(h, J)$, if the characteristic function takes the form

$$\mathcal{F}_\xi\{\phi\} = e^{\sqrt{-1}\xi^i h_i - (1/2)\xi^i \xi^j V_{ij}},$$

where $h = (h_i)_{i=1}^d \in \mathbb{R}^d$ and $V = (V_{ij})$ is a real symmetric matrix such that the Hermitian matrix $J := V + \sqrt{-1}S$ is positive semidefinite. When the canonical observables X need to be specified, we also use the notation $(X, \phi) \sim N(h, J)$. (See [4, 7, 12, 14] for more information about quantum Gaussian states.)

We will discuss relationships between a quantum Gaussian state ϕ on a CCR and a state on another algebra. In such a case, we need to use the *quasi-characteristic function*

$$(2.2) \quad \begin{aligned} &\phi\left(\prod_{t=1}^r e^{\sqrt{-1}\xi_t^i X_i}\right) \\ &= \exp\left(\sum_{t=1}^r \left(\sqrt{-1}\xi_t^i h_i - \frac{1}{2}\xi_t^i \xi_t^j J_{ji}\right) - \sum_{t=1}^r \sum_{s=t+1}^r \xi_t^i \xi_s^j J_{ji}\right), \end{aligned}$$

of a quantum Gaussian state, where $(X, \phi) \sim N(h, J)$ and $\{\xi_t\}_{t=1}^r$ is a finite subset of $\mathbb{C}^d$ [13].

Given a sequence $\mathcal{H}^{(n)}$, $n \in \mathbb{N}$, of finite dimensional Hilbert spaces, let $X^{(n)} = (X_1^{(n)}, \ldots, X_d^{(n)})$ and $\rho^{(n)}$ be a list of observables and a density operator on

each $\mathcal{H}^{(n)}$. We say the sequence $(X^{(n)}, \rho^{(n)})$ *converges in law to a quantum Gaussian state* $N(h, J)$, denoted as $(X^{(n)}, \rho^{(n)}) \rightsquigarrow_q N(h, J)$, if

$$\lim_{n \to \infty} \operatorname{Tr} \rho^{(n)} \left(\prod_{t=1}^r e^{\sqrt{-1} \xi_t^i X_i^{(n)}} \right) = \phi \left(\prod_{t=1}^r e^{\sqrt{-1} \xi_t^i X_i} \right)$$

for any finite subset $\{\xi_t\}_{t=1}^r$ of $\mathbb{C}^d$, where $(X, \phi) \sim N(h, J)$. Here we do not intend to introduce the notion of "quantum convergence in law" in general. We use this notion only for quantum Gaussian states in the sense of convergence of quasi-characteristic function.

The following is a version of the quantum central limit theorem (see [13], e.g.).

PROPOSITION 2.5 (Quantum central limit theorem). *Let* A_i $(1 \le i \le d)$ *and* ρ *be observables and a state on a finite dimensional Hilbert space* $\mathcal{H}$ *such that* $\operatorname{Tr} \rho A_i = 0$, *and let*

$$X_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^n I^{\otimes(k-1)} \otimes A_i \otimes I^{\otimes(n-k)}.$$

Then $(X^{(n)}, \rho^{\otimes n}) \rightsquigarrow_q N(0, J)$, *where* J *is the Hermitian matrix whose* (i, j)*th entry is given by* $J_{ij} = \operatorname{Tr} \rho A_j A_i$.

For later convenience, we introduce the notion of an "infinitesimal" object relative to the convergence $(X^{(n)}, \rho^{(n)}) \rightsquigarrow_q N(0, J)$ as follows. Given a list $X^{(n)} = (X_1^{(n)}, \ldots, X_d^{(n)})$ of observables and a state $\rho^{(n)}$ on each $\mathcal{H}^{(n)}$ that satisfy $(X^{(n)}, \rho^{(n)}) \rightsquigarrow_q N(0, J) \sim (X, \phi)$, we say a sequence $R^{(n)}$ of observables, each being defined on $\mathcal{H}^{(n)}$, is *infinitesimal relative to the convergence* $(X^{(n)}, \rho^{(n)}) \rightsquigarrow_q N(0, J)$ if it satisfies

$$(2.3) \qquad \lim_{n \to \infty} \operatorname{Tr} \rho^{(n)} \left(\prod_{t=1}^r e^{\sqrt{-1}(\xi_t^i X_i^{(n)} + \eta_t R^{(n)})} \right) = \phi \left(\prod_{t=1}^r e^{\sqrt{-1} \xi_t^i X_i} \right)$$

for any finite subset of $\{\xi_t\}_{t=1}^r$ of $\mathbb{C}^d$ and any finite subset $\{\eta_t\}_{t=1}^r$ of $\mathbb{C}$. This is equivalent to saying that

$$\left(\begin{pmatrix} X^{(n)} \\ R^{(n)} \end{pmatrix}, \rho^{(n)} \right) \rightsquigarrow_q N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J & 0 \\ 0 & 0 \end{pmatrix} \right),$$

and is much stronger a requirement than

$$\left(R^{(n)}, \rho^{(n)} \right) \rightsquigarrow_q N(0, 0).$$

An infinitesimal object $R^{(n)}$ relative to $(X^{(n)}, \rho^{(n)}) \rightsquigarrow_q N(0, J)$ will be denoted as $o(X^{(n)}, \rho^{(n)})$.

The following is in essence a simple extension of Proposition 2.5, but will turn out to be useful in applications.

LEMMA 2.6. *In addition to assumptions of Proposition* 2.5, *let* $P(n)$, $n \in \mathbb{N}$, *be a sequence of observables on* $\mathcal{H}$, *and let*

$$R^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes P(n) \otimes I^{\otimes(n-k)}.$$

If $\lim_{n\to\infty} P(n) = 0$ *and* $\lim_{n\to\infty} \sqrt{n} \operatorname{Tr} \rho \, P(n) = 0$, *then* $R^{(n)} = o(X^{(n)}, \rho^{\otimes n})$.

This lemma gives a precise criterion for the convergence of quasi-characteristic function for quantum Gaussian states.

2.3. *Quantum local asymptotic normality.* We are now ready to extend the notion of local asymptotic normality to the quantum domain.

DEFINITION 2.7 (QLAN). Given a sequence $\mathcal{H}^{(n)}$ of finite dimensional Hilbert spaces, let $\mathcal{S}^{(n)} = \{\rho_\theta^{(n)}; \theta \in \Theta \subset \mathbb{R}^d\}$ be a quantum statistical model on $\mathcal{H}^{(n)}$, where $\rho_\theta^{(n)}$ is a parametric family of density operators and Θ is an open set. We say $\mathcal{S}^{(n)}$ is *quantum locally asymptotically normal* (QLAN) at $\theta_0 \in \Theta$ if the following conditions are satisfied:

(i) for any $\theta \in \Theta$ and $n \in \mathbb{N}$, $\rho_\theta^{(n)}$ is mutually absolutely continuous to $\rho_{\theta_0}^{(n)}$,

(ii) there exist a list $\Delta^{(n)} = (\Delta_1^{(n)}, \ldots, \Delta_d^{(n)})$ of observables on each $\mathcal{H}^{(n)}$ that satisfies

$$\left(\Delta^{(n)}, \rho_{\theta_0}^{(n)}\right) \underset{q}{\rightsquigarrow} N(0, J),$$

where J is a $d \times d$ Hermitian positive semidefinite matrix with $\operatorname{Re} J > 0$,

(iii) quantum log-likelihood ratio $\mathcal{L}_h^{(n)} := \mathcal{L}(\rho_{\theta_0+h/\sqrt{n}}^{(n)} | \rho_{\theta_0}^{(n)})$ is expanded in $h \in \mathbb{R}^d$ as

(2.4) $$\mathcal{L}_h^{(n)} = h^i \Delta_i^{(n)} - \tfrac{1}{2}(J_{ij} h^i h^j) I^{(n)} + o(\Delta^{(n)}, \rho_{\theta_0}^{(n)}),$$

where $I^{(n)}$ is the identity operator on $\mathcal{H}^{(n)}$.

It is also possible to extend Le Cam's third lemma (Proposition 1.1) to the quantum domain. To this end, however, we need a device to handle the infinitesimal residual term in (2.4) in a more elaborate way.

DEFINITION 2.8. Let $\mathcal{S}^{(n)} = \{\rho_\theta^{(n)}; \theta \in \Theta \subset \mathbb{R}^d\}$ be as in Definition 2.7, and let $X^{(n)} = (X_1^{(n)}, \ldots, X_r^{(n)})$ be a list of observables on $\mathcal{H}^{(n)}$. We say the pair $(\mathcal{S}^{(n)}, X^{(n)})$ is *jointly QLAN* at $\theta_0 \in \Theta$ if the following conditions are satisfied:

(i) for any $\theta \in \Theta$ and $n \in \mathbb{N}$, $\rho_\theta^{(n)}$ is mutually absolutely continuous to $\rho_{\theta_0}^{(n)}$,

(ii) there exist a list $\Delta^{(n)} = (\Delta_1^{(n)}, \ldots, \Delta_d^{(n)})$ of observables on each $\mathcal{H}^{(n)}$ that satisfies

$$(2.5) \qquad \left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{(n)} \right) \underset{q}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^* & J \end{pmatrix} \right),$$

where Σ and J are Hermitian positive semidefinite matrices of size $r \times r$ and $d \times d$, respectively, with $\mathrm{Re}\, J > 0$, and τ is a complex matrix of size $r \times d$.

(iii) quantum log-likelihood ratio $\mathcal{L}_h^{(n)} := \mathcal{L}(\rho_{\theta_0+h/\sqrt{n}}^{(n)} | \rho_{\theta_0}^{(n)})$ is expanded in $h \in \mathbb{R}^d$ as

$$(2.6) \qquad \mathcal{L}_h^{(n)} = h^i \Delta_i^{(n)} - \frac{1}{2}(J_{ij} h^i h^j) I^{(n)} + o\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{(n)} \right).$$

With Definition 2.8, we can state a quantum extension of Le Cam's third lemma as follows.

THEOREM 2.9.   *Let $\mathcal{S}^{(n)}$ and $X^{(n)}$ be as in Definition 2.8. If $(\rho_\theta^{(n)}, X^{(n)})$ is jointly QLAN at $\theta_0 \in \Theta$, then*

$$\left(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{(n)} \right) \underset{q}{\rightsquigarrow} N\big((\mathrm{Re}\, \tau) h, \Sigma \big)$$

for any $h \in \mathbb{R}^d$.

It should be emphasised that assumption (2.6), which was superfluous in classical theory, is in fact crucial in proving Theorem 2.9.

In applications, we often handle i.i.d. extensions. In classical statistics, a sequence of i.i.d. extensions of a model is LAN if the log-likelihood ratio is twice differentiable [21]. Quite analogously, we can prove, with the help of Lemma 2.6, that a sequence of i.i.d. extensions of a quantum statistical model is QLAN if the quantum log-likelihood ratio is twice differentiable.

THEOREM 2.10.   *Let $\{\rho_\theta; \theta \in \Theta \subset \mathbb{R}^d\}$ be a quantum statistical model on a finite dimensional Hilbert space $\mathcal{H}$ satisfying $\rho_\theta \sim \rho_{\theta_0}$ for all $\theta \in \Theta$, where $\theta_0 \in \Theta$ is an arbitrarily fixed point. If $\mathcal{L}_h := \mathcal{L}(\rho_{\theta_0+h} | \rho_{\theta_0})$ is differentiable around $h = 0$ and twice differentiable at $h = 0$, then $\{\rho_\theta^{\otimes n}; \theta \in \Theta \subset \mathbb{R}^d\}$ is QLAN at θ_0: that is, $\rho_\theta^{\otimes n} \sim \rho_{\theta_0}^{\otimes n}$, and*

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes L_i \otimes I^{\otimes(n-k)}$$

and $J_{ij} := \mathrm{Tr}\, \rho_{\theta_0} L_j L_i$, with L_i being the ith symmetric logarithmic derivative at $\theta_0 \in \Theta$, satisfy conditions (ii) *and* (iii) *in Definition 2.7.*

By combining Theorem 2.10 with Theorem 2.9 and Lemma 2.6, we obtain the following.

COROLLARY 2.11. *Let $\{\rho_\theta; \theta \in \Theta \subset \mathbb{R}^d\}$ be a quantum statistical model on $\mathcal{H}$ satisfying $\rho_\theta \sim \rho_{\theta_0}$ for all $\theta \in \Theta$, where $\theta_0 \in \Theta$ is an arbitrarily fixed point. Further, let $\{B_i\}_{1 \le i \le r}$ be observables on $\mathcal{H}$ satisfying $\operatorname{Tr} \rho_{\theta_0} B_i = 0$ for $i = 1, \dots, r$. If $\mathcal{L}_h := \mathcal{L}(\rho_{\theta_0+h} | \rho_{\theta_0})$ is differentiable around $h = 0$ and twice differentiable at $h = 0$, then the pair $(\{\rho_\theta^{\otimes n}\}, X^{(n)})$ of i.i.d. extension model $\{\rho_\theta^{\otimes n}\}$ and the list $X^{(n)} = \{X_i^{(n)}\}_{1 \le i \le r}$ of observables defined by*

$$X_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes B_i \otimes I^{\otimes(n-k)}$$

is jointly QLAN at θ_0, and

$$(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{\otimes n}) \underset{q}{\rightsquigarrow} N((\operatorname{Re} \tau)h, \Sigma)$$

for any $h \in \mathbb{R}^d$, where Σ is the $r \times r$ positive semidefinite matrix defined by $\Sigma_{ij} = \operatorname{Tr} \rho_{\theta_0} B_j B_i$ and τ is the $r \times d$ matrix defined by $\tau_{ij} = \operatorname{Tr} \rho_{\theta_0} L_j B_i$ with L_i being the ith symmetric logarithmic derivative at θ_0.

Corollary 2.11 is an i.i.d. version of the quantum Le Cam third lemma, and will play a key role in demonstrating the asymptotic achievability of the Holevo bound.

3. Applications to quantum statistics.

3.1. *Achievability of the Holevo bound.* Corollary 2.11 prompts us to expect that, for sufficiently large n, the estimation problem for the parameter h of $\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}$ could be reduced to that for the shift parameter h of the quantum Gaussian shift model $N((\operatorname{Re} \tau)h, \Sigma)$. The latter problem has been well-established to date (see Section B.2 in [23]). In particular, the best strategy for estimating the shift parameter h is the one that achieves the Holevo bound $C_h(N((\operatorname{Re} \tau)h, \Sigma), G)$ (see Theorem B.7 in [23]). Moreover, it is shown (see Corollary B.6 in [23]) that the Holevo bound $C_h(N((\operatorname{Re} \tau)h, \Sigma), G)$ is identical to the Holevo bound $C_{\theta_0}(\rho_\theta, G)$ for the model ρ_θ at θ_0. These facts suggest the existence of a sequence $M^{(n)}$ of estimators for the parameter h of $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\}_n$ that asymptotically achieves the Holevo bound $C_{\theta_0}(\rho_\theta, G)$. The following theorem materialises this program.

THEOREM 3.1. *Let $\{\rho_\theta; \theta \in \Theta \subset \mathbb{R}^d\}$ be a quantum statistical model on a finite dimensional Hilbert space $\mathcal{H}$, and fix a point $\theta_0 \in \Theta$. Suppose that $\rho_\theta \sim \rho_{\theta_0}$ for all $\theta \in \Theta$, and that the quantum log-likelihood ratio $\mathcal{L}_h := \mathcal{L}(\rho_{\theta_0+h} | \rho_{\theta_0})$ is differentiable in h around $h = 0$ and twice differentiable at $h = 0$. For any countable dense subset D of $\mathbb{R}^d$ and any weight matrix G, there exist a sequence $M^{(n)}$ of estimators on the model $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^d\}$ that enjoys*

$$\lim_{n \to \infty} E_h^{(n)}[M^{(n)}] = h$$

and

$$\lim_{n \to \infty} \mathrm{Tr}\, G V_h^{(n)}[M^{(n)}] = C_{\theta_0}(\rho_\theta, G)$$

for every $h \in D$. Here $C_{\theta_0}(\rho_\theta, G)$ is the Holevo bound at θ_0. Here $E_h^{(n)}[\cdot]$ and $V_h^{(n)}[\cdot]$ stand for the expectation and the covariance matrix under the state $\rho_{\theta_0 + h/\sqrt{n}}^{\otimes n}$.

Theorem 3.1 asserts that there is a sequence $M^{(n)}$ of estimators on $\{\rho_{\theta_0 + h/\sqrt{n}}^{\otimes n}\}_n$ that is asymptotically unbiased and achieves the Holevo bound $C_{\theta_0}(\rho_\theta, G)$ for all h that belong to a dense subset of $\mathbb{R}^d$. Since this result requires only twice differentiability of the quantum log-likelihood ratio of the base model ρ_θ, it will be useful in a wide range of statistical estimation problems.

3.2. *Application to qubit state estimation.* In order to demonstrate the applicability of our theory, we explore qubit state estimation problems.

EXAMPLE 3.2 (3-dimensional faithful state model). The first example is an ordinary one, comprising the totality of faithful qubit states:

$$\mathcal{S}(\mathbb{C}^2) = \{\rho_\theta = \tfrac{1}{2}(I + \theta^1 \sigma_1 + \theta^2 \sigma_2 + \theta^3 \sigma_3); \theta = (\theta^i)_{1 \le i \le 3} \in \Theta\},$$

where σ_i $(i = 1, 2, 3)$ are the standard Pauli matrices and Θ is the open unit ball in $\mathbb{R}^3$. Due to the rotational symmetry, we take the reference point to be $\theta_0 = (0, 0, r)$, with $0 \le r < 1$. By a direct calculation, we see that the symmetric logarithmic derivatives (SLDs) of the model ρ_θ at $\theta = \theta_0$ are $(L_1, L_2, L_3) = (\sigma_1, \sigma_2, (rI + \sigma_3)^{-1})$, and the SLD Fisher information matrix $J^{(S)}$ at θ_0 is given by the real part of the matrix

$$J := [\mathrm{Tr}\, \rho_{\theta_0} L_j L_i]_{ij} = \begin{pmatrix} 1 & -r\sqrt{-1} & 0 \\ r\sqrt{-1} & 1 & 0 \\ 0 & 0 & 1/(1 - r^2) \end{pmatrix}.$$

Given a 3×3 real positive definite matrix G, the minimal value of the weighted covariances at $\theta = \theta_0$ is given by

$$\min_{\hat{M}} \mathrm{Tr}\, G V_{\theta_0}[\hat{M}] = C_{\theta_0}^{(1)}(\rho_\theta, G),$$

where the minimum is taken over all estimators $\hat{M}$ that are locally unbiased at θ_0, and

$$C_{\theta_0}^{(1)}(\rho_\theta, G) = (\mathrm{Tr}\, \sqrt{\sqrt{G} J^{(S)-1} \sqrt{G}})^2$$

is the Hayashi–Gill–Massar bound [6, 10] (see also [22]). On the other hand, the SLD tangent space (i.e., the linear span of the SLDs) is obviously invariant under the action of the commutation operator $\mathcal{D}$, and the Holevo bound is given by

$$C_{\theta_0}(\rho_\theta, G) := \operatorname{Tr} G J^{(R)^{-1}} + \operatorname{Tr}|\sqrt{G}\operatorname{Im}J^{(R)^{-1}}\sqrt{G}|,$$

where

$$J^{(R)^{-1}} := (\operatorname{Re}J)^{-1} J (\operatorname{Re}J)^{-1} = \begin{pmatrix} 1 & -r\sqrt{-1} & 0 \\ r\sqrt{-1} & 1 & 0 \\ 0 & 0 & 1-r^2 \end{pmatrix}$$

is the inverse of the right logarithmic derivative (RLD) Fisher information matrix (see Corollary B.2 in [23]).

It can be shown that the Hayashi–Gill–Massar bound is greater than the Holevo bound:

$$C_{\theta_0}^{(1)}(\rho_\theta, G) > C_{\theta_0}(\rho_\theta, G).$$

Let us check this fact for the special case when $G = J^{(S)}$. A direct computation shows that

$$C_{\theta_0}^{(1)}(\rho_\theta, J^{(S)}) = 9$$

and

$$C_{\theta_0}(\rho_\theta, J^{(S)}) = 3 + 2r.$$

The left panel of Figure 1 shows the behaviour of $C_{\theta_0}(\rho_\theta, J^{(S)})$ (solid) and $C_{\theta_0}^{(1)}(\rho_\theta, J^{(S)})$ (dashed) as functions of r. We see that the Holevo bound $C_{\theta_0}(\rho_\theta, J^{(S)})$ is much smaller than $C_{\theta_0}^{(1)}(\rho_\theta, J^{(S)})$.

Does this fact imply that the Holevo bound is of no use? The answer is contrary, as Theorem 3.1 asserts. We will demonstrate the asymptotic achievability of the Holevo bound. Let

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes k-1} \otimes L_i \otimes I^{\otimes n-k}$$

and let $X_i^{(n)} := \Delta_i^{(n)}$ for $i = 1, 2, 3$. It follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{\otimes n} \right) \underset{q}{\rightsquigarrow} N\left(0, \begin{pmatrix} J & J \\ J & J \end{pmatrix} \right).$$

Since

$$\mathcal{L}(\theta) := \mathcal{L}(\rho_\theta | \rho_{\theta_0}) = 2\log\left(\sqrt{\rho_{\theta_0}^{-1}} \sqrt{\sqrt{\rho_{\theta_0}} \rho_\theta \sqrt{\rho_{\theta_0}}} \sqrt{\rho_{\theta_0}^{-1}} \right)$$
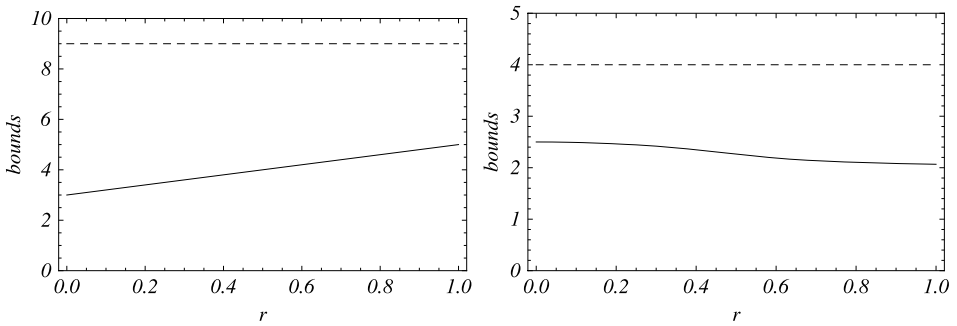
FIG. 1. *The left panel displays the Holevo bound $C_{(0,0,r)}(\rho_\theta, J^{(S)})$ (solid) and the Hayashi–Gill–Massar bound $C_{(0,0,r)}^{(1)}(\rho_\theta, J^{(S)})$ (dashed) for the 3-D model $\rho_\theta = \frac{1}{2}(I + \theta^1\sigma_1 + \theta^2\sigma_2 + \theta^3\sigma_3)$ as functions of $r = \|\theta\|$. The right panel displays the Holevo bound $C_{(0,r)}(\rho_\theta, J^{(S)})$ (solid) and the Nagaoka bound $C_{(0,r)}^{(1)}(\rho_\theta, J^{(S)})$ (dashed) for the 2-D model $\rho_\theta = \frac{1}{2}(I + \theta^1\sigma_1 + \theta^2\sigma_2 + \frac{1}{4}\sqrt{1 - \|\theta\|^2}\sigma_3)$.*

is obviously of class C^∞ in θ, Corollary 2.11 shows that $(\{\rho_\theta^{\otimes n}\}, X^{(n)})$ is jointly QLAN at θ_0, and that

$$\left(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right) \overset{\rightsquigarrow}{q} N\left((\operatorname{Re} J)h, J\right)$$

for all $h \in \mathbb{R}^3$. This implies that a sequence of models $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^d\}$ converges to a quantum Gaussian shift model $\{N((\operatorname{Re} J)h, J); h \in \mathbb{R}^3\}$. Note that the imaginary part

$$S = \begin{pmatrix} 0 & -r\sqrt{-1} & 0 \\ r\sqrt{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

of the matrix J determines the CCR(S), as well as the corresponding basic canonical observables $X = (X^1, X^2, X^3)$. When $r \neq 0$, the above S has the following physical interpretation: X^1 and X^2 form a canonical pair of quantum Gaussian observables, while X^3 is a classical Gaussian random variable. In this way, the matrix J automatically tells us the structure of the limiting quantum Gaussian shift model.

Now, the best strategy for estimating the shift parameter h of the quantum Gaussian shift model $\{N((\operatorname{Re} J)h, J); h \in \mathbb{R}^d\}$ is the one that achieves the Holevo bound $C_h(N((\operatorname{Re} J)h, J), G)$ (see Theorem B.7 in [23]). Moreover, this Holevo bound $C_h(N((\operatorname{Re} J)h, J), G)$ is identical to the Holevo bound $C_{\theta_0}(\rho_\theta, G)$ for the model ρ_θ at θ_0 (see Corollary B.6 in [23]. Recall that the matrix J is evaluated at θ_0 of the model ρ_θ). Theorem 3.1 combines these facts, and concludes that there exist a sequence $M^{(n)}$ of estimators on the model $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^3\}$ that is asymptotically unbiased and achieves the common values of the Holevo bound:

$$\lim_{n \to \infty} \operatorname{Tr} G V_h^{(n)}[M^{(n)}] = C_h\left(N((\operatorname{Re} J)h, J), G\right) = C_{\theta_0}(\rho_\theta, G)$$

for all h that belong to a countable dense subset of $\mathbb{R}^3$.

It should be emphasised that the matrix J becomes the identity at the origin $\theta_0 = (0, 0, 0)$. This means that the limiting Gaussian shift model $\{N(h, J); h \in \mathbb{R}^3\}$ is "classical." Since such a degenerate case cannot be treated in [9, 11, 14], our method has a clear advantage in applications.

EXAMPLE 3.3 (Pure state model). The second example is to demonstrate that our formulation allows us to treat pure state models. Let us consider the model $\mathcal{S} = \{|\psi(\theta)\rangle\langle\psi(\theta)|; \theta = (\theta^i)_{1 \le i \le 2} \in \Theta\}$ defined by

$$\psi(\theta) := \frac{1}{\sqrt{\cosh \|\theta\|}} e^{(1/2)(\theta^1 \sigma_1 + \theta^2 \sigma_2)} \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

where Θ is an open subset of $\mathbb{R}^2$ containing the origin, and $\|\cdot\|$ denotes the Euclid norm. By a direct computation, the SLDs at $\theta_0 = (0, 0)$ are $(L_1, L_2) = (\sigma_1, \sigma_2)$, and the SLD Fisher information matrix $J^{(S)}$ is the real part of the matrix

$$J = [\operatorname{Tr} \rho_{\theta_0} L_j L_i]_{ij} = \begin{pmatrix} 1 & -\sqrt{-1} \\ \sqrt{-1} & 1 \end{pmatrix},$$

that is, $J^{(S)} = I$. Since the SLD tangent space is $\mathcal{D}$ invariant [3], the Holevo bound for a weight $G > 0$ is represented as

$$C_{\theta_0}(\rho_\theta, G) := \operatorname{Tr} G J^{(R)-1} + \operatorname{Tr} |\sqrt{G} \operatorname{Im} J^{(R)-1} \sqrt{G}|,$$

where

$$J^{(R)-1} := (\operatorname{Re} J)^{-1} J (\operatorname{Re} J)^{-1} = \begin{pmatrix} 1 & -\sqrt{-1} \\ \sqrt{-1} & 1 \end{pmatrix}$$

is the inverse RLD Fisher information matrix (see Corollary B.2 in [23]).

Let us demonstrate that our QLAN is applicable also to pure state models. Let

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^n I^{\otimes k-1} \otimes L_i \otimes I^{\otimes n-k}$$

and let $X_i^{(n)} := \Delta_i^{(n)}$ for $i = 1, 2$. It follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{\otimes n} \right) \underset{q}{\rightsquigarrow} N\left(0, \begin{pmatrix} J & J \\ J & J \end{pmatrix} \right).$$

Since

$$\mathcal{L}(\theta) := \mathcal{L}(\rho_\theta | \rho_{\theta_0}) = \theta^1 \sigma_1 + \theta^2 \sigma_2 - \log \cosh \|\theta\|$$

is of class C^∞ with respect to θ, it follows from Corollary 2.11 that $(\{\rho_\theta^{\otimes n}\}, X^{(n)})$ is jointly QLAN at θ_0, and that

$$\left(X^{(n)}, \rho_{\theta_0 + h/\sqrt{n}}^{\otimes n} \right) \rightsquigarrow N\left((\operatorname{Re} J)h, J \right) = N\left(h, J^{(R)-1} \right)$$

for all $h \in \mathbb{R}^2$. Theorem 3.1 further asserts that there exist a sequence $M^{(n)}$ of estimators on the model $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^2\}$ that is asymptotically unbiased and achieves the Holevo bound:

$$\lim_{n\to\infty} \operatorname{Tr} G V_h^{(n)}[M^{(n)}] = C_h\big(N(h, J^{(R)^{-1}}), G\big) = C_{(0,0)}(\rho_\theta, G)$$

for all h that belong to a dense subset of $\mathbb{R}^3$. In fact, the sequence $M^{(n)}$ can be taken to be a separable one, making no use of quantum correlations [17]. (See also Section B.3 in [23] for a simple proof.) Note that the matrix $J^{(R)^{-1}}$ is degenerate, and the derived quantum Gaussian shift model $\{N(h, J^{(R)^{-1}})\}_h$ is a canonical coherent model [3].

EXAMPLE 3.4 (2-dimensional faithful state model). The third example treats the case when the SLD tangent space is not $\mathcal{D}$ invariant. Let us consider the model

$$\mathcal{S} = \{\rho_\theta = \tfrac{1}{2}(I + \theta^1 \sigma_1 + \theta^2 \sigma_2 + z_0\sqrt{1 - \|\theta\|^2}\sigma_3); \theta = (\theta^i)_{1\le i\le 2} \in \Theta\},$$

where $0 \le z_0 < 1$, and Θ is the open unit disk. Due to the rotational symmetry around z-axis, we take the reference point to be $\theta_0 = (0, r)$, with $0 \le r < 1$. By a direct calculation, we see that the SLDs at θ_0 are $(L_1, L_2) = (\sigma_1, \frac{1}{1-r^2}(\sigma_2 - rI))$. It is important to notice that the SLD tangent space $\operatorname{span}\{L_i\}_{i=1}^2$ is not $\mathcal{D}$ invariant unless $r = 0$. In fact

$$\mathcal{D}\sigma_1 = z(r)\sigma_2 - r\sigma_3, \qquad \mathcal{D}\sigma_2 = -z(r)\sigma_1,$$

where $z(r) := E[\sigma_3] = z_0\sqrt{1 - r^2}$. The minimal $\mathcal{D}$ invariant extension $\mathcal{T}$ of the SLD tangent space has a basis $(D_1, D_2, D_3) := (L_1, L_2, \sigma_3 - z(r)I)$. The matrices Σ, J, and τ appeared in Definition 2.8 and Corollary 2.11 are calculated as

$$\Sigma := [\operatorname{Tr}\rho_{\theta_0} D_j D_i]_{ij}$$

$$= \begin{pmatrix} 1 & -\sqrt{-1}\dfrac{z_0^2}{z(r)} & r\sqrt{-1} - z(r) \\[3mm] \sqrt{-1}\dfrac{z_0^2}{z(r)} & \dfrac{z_0^2}{z(r)^2} & -\left(\dfrac{r}{z(r)} + \sqrt{-1}\right)z_0^2 \\[3mm] -r\sqrt{-1} - z(r) & -\left(\dfrac{r}{z(r)} - \sqrt{-1}\right)z_0^2 & 1 \end{pmatrix},$$

$$J := [\operatorname{Tr}\rho_{\theta_0} L_j L_i]_{ij} = \begin{pmatrix} 1 & -\sqrt{-1}\dfrac{z_0^2}{z(r)} \\[3mm] \sqrt{-1}\dfrac{z_0^2}{z(r)} & \dfrac{z_0^2}{z(r)^2} \end{pmatrix},$$

$$\tau := [\operatorname{Tr} \rho_{\theta_0} L_j \sigma_i]_{ij} = \begin{pmatrix} 1 & -\sqrt{-1}\dfrac{z_0^2}{z(r)} \\[6pt] \sqrt{-1}\dfrac{z_0^2}{z(r)} & \dfrac{z_0^2}{z(r)^2} \\[6pt] -r\sqrt{-1}-z(r) & -\left(\dfrac{r}{z(r)}-\sqrt{-1}\right)z_0^2 \end{pmatrix}.$$

Given a 2×2 real positive definite matrix G, the minimal value of the weighted covariances at $\theta = \theta_0$ is given by

$$\min_{\hat{M}} \operatorname{Tr} G V_{\theta_0}[\hat{M}] = C_{\theta_0}^{(1)}(\rho_\theta, G),$$

where the minimum is taken over all estimators $\hat{M}$ that are locally unbiased at θ_0, and

$$C_{\theta_0}^{(1)}(\rho_\theta, G) = (\operatorname{Tr}\sqrt{\sqrt{G}J^{(S)-1}\sqrt{G}})^2$$

is the Nagaoka bound [18] (see also [22]).

It can be shown that the Nagaoka bound is greater than the Holevo bound:

$$C_{\theta_0}^{(1)}(\rho_\theta, G) > C_{\theta_0}(\rho_\theta, G).$$

Let us check this fact for the special case when $G = J^{(S)}$. A direct computation shows that

$$C_{\theta_0}^{(1)}(\rho_\theta, J^{(S)}) = 4$$

and

$$C_{\theta_0}(\rho_\theta, J^{(S)}) = \begin{cases} 2(1+z_0) - r^2(1-z_0^2), & \text{if } 0 \le r \le \sqrt{\dfrac{z_0}{1-z_0^2}}, \\[10pt] 2 + \dfrac{z_0^2}{r^2(1-z_0^2)}, & \text{if } \sqrt{\dfrac{z_0}{1-z_0^2}} < r. \end{cases}$$

The right panel of Figure 1 shows the behaviour of $C_{\theta_0}(\rho_\theta, J^{(S)})$ (solid) and $C_{\theta_0}^{(1)}(\rho_\theta, J^{(S)})$ with $z_0 = \frac{1}{4}$ (dashed) as functions of r. We see that Holevo bound $C_{\theta_0}(\rho_\theta, J^{(S)})$ is much smaller than $C_{(0,r)}^{(1)}(\rho_\theta, J^{(S)})$.

As in Example 3.2, we demonstrate that the Holevo bound is asymptotically achievable. Let

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes k-1} \otimes L_i \otimes I^{\otimes n-k} \qquad (i = 1, 2),$$

and let

$$X_j^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes k-1} \otimes D_j \otimes I^{\otimes n-k} \qquad (j = 1, 2, 3).$$

It then follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N\left(0, \begin{pmatrix} \Sigma & \tau \\ \tau^* & J \end{pmatrix}\right).$$

Therefore, Corollary 2.11 shows that $(\{\rho_\theta^{\otimes n}\}, X^{(n)})$ is jointly QLAN at θ_0, and that

$$\left(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N\left((\mathrm{Re}\,\tau)h, \Sigma\right)$$

for all $h \in \mathbb{R}^2$.

It should be noted that the off-diagonal block τ of the "quantum covariance" matrix is not a square matrix. This means that the derived quantum Gaussian shift model $\{N((\mathrm{Re}\,\tau)h, \Sigma); h \in \mathbb{R}^2\}$ forms a submanifold of the total quantum Gaussian shift model derived in Example 3.2, corresponding to a 2-dimensional linear subspace in the shift parameter space. Nevertheless, Theorem 3.1 asserts that there exists a sequence $M^{(n)}$ of estimators on the model $\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}; h \in \mathbb{R}^3\}$ that is asymptotically unbiased and achieves the Holevo bound:

$$\lim_{n\to\infty} \mathrm{Tr}\, G V_h^{(n)}[M^{(n)}] = C_h\left(N((\mathrm{Re}\,\tau)h, \Sigma), G\right) = C_{\theta_0}(\rho_\theta, G)$$

for all h that belong to a dense subset of $\mathbb{R}^3$.

3.3. *Translating estimation of h to estimation of θ.* As we have seen in the previous subsections, our theory enables us to construct asymptotically optimal estimators of h in the local models indexed by the parameter $\theta_0 + h/\sqrt{n}$. In practice of course, θ_0 is unknown and hence estimation of h, with θ_0 known, is irrelevant. The actual sequence of measurements which we have constructed depends in all interesting cases on θ_0.

However, the results immediately inspire two-step (or adaptive) procedures, in which we first measure a small proportion of the quantum systems, in number n_1 say, using some standard measurement scheme, for instance, separate particle quantum tomography. From these measurement outcomes we construct an initial estimate of θ, let us call it $\widetilde{\theta}$. We can now use our theory to compute the asymptotically optimal measurement scheme which corresponds to the situation $\theta_0 = \widetilde{\theta}$. We proceed to implement this measurement on the remaining quantum systems collectively, estimating h in the model $\theta = \widetilde{\theta} + h/\sqrt{n_2}$ where n_2 is the number of systems still available for the second stage.

What can we say about such a procedure? If $n_1/n \to \alpha > 0$ as $n \to \infty$, then we can expect that the initial estimate $\widetilde{\theta}$ is root n consistent. In smooth models, one would expect that in this case the final estimate $\widehat{\theta} = \widetilde{\theta} + \widehat{h}/\sqrt{n_2}$ would be asymptotically optimal *up to a factor* $1 - \alpha$: its limiting variance will be a factor $(1-\alpha)^{-1}$ too large.

If however $n_1 \to \infty$ but $n_1/n \to \alpha = 0$, then one would expect this procedure to break down, unless the rate of growth of n_1 is very carefully chosen (and fast

enough). On the other hand, instead of a direct two-step procedure, with the final estimate computed as $\widetilde{\theta} + \widehat{h}/\sqrt{n_2}$, one could be more careful in how the data obtained from the second stage measurement is used. Given the second step measurement, which results in an observed value $\widehat{h}$, one could write down the likelihood for h based on the given measurement and the initially specified model, and compute instead of the just mentioned one-step iterate, the actual maximum likelihood estimator of θ based on the second stage data. Such procedures have earlier been studied by Gill and Massar [6] and others, and shown in special cases to perform very well.

However, in general, the computational problem of even calculating the likelihood given data, measurement, and model, is challenging, due to the huge size of the Hilbert space of n copies of a finite dimensional quantum system.

4. Concluding remarks. We have developed a new theory of local asymptotic normality in the quantum domain based on a quantum extension of the log-likelihood ratio. This formulation is applicable to any model satisfying a mild smoothness condition, and is free from artificial setups such as the use of a special coordinate system and/or nondegeneracy of eigenvalues of the reference state. We also have proved asymptotic achievability of the Holevo bound for the local shift parameter on a dense subset of the parameter space.

There are of course many open questions left. Among others, it is not clear whether every sequence of statistics on a QLAN model can be realised on the limiting quantum Gaussian shift model. In classical statistics, such a problem has been solved affirmatively as the representation theorem, which asserts that, given a weakly convergent sequence $T^{(n)}$ of statistics on $\{p_{\theta_0+h/\sqrt{n}}^{(n)}; h \in \mathbb{R}^d\}$, there exist a limiting statistics T on $\{N(h, J^{-1}); h \in \mathbb{R}^d\}$ such that $T^{(n)} \overset{h}{\rightsquigarrow} T$. Representation theorem is useful in proving, for example, the nonexistence of an asymptotically superefficient estimator (the converse part, as stated in Introduction). Moreover, the so-called convolution theorem and local asymptotic minimax theorem, which are the standard tools in discussing asymptotic lower bounds for estimation in LAN models, immediately follows [21]. Extending the representation theorem, convolution theorem, and local asymptotic minimax theorem to the quantum domain is an intriguing open problem. However it surely is possible to make some progress in this direction, as, for instance, the results of Gill and Guţă [4]. In that paper, the van Trees inequality was used to derive some results in a "poor man's" version of QLAN theory; see also [5].

It also remains to be seen whether our asymptotically optimal statistical procedures for the local model with local parameter h can be translated into useful statistical procedures for the real world case in which θ_0 is unknown.

SUPPLEMENTARY MATERIAL

Supplementary material to "Quantum local asymptotic normality based on a new quantum likelihood ratio" (DOI: 10.1214/13-AOS1147SUPP; .pdf). Sec-

tion A is devoted to proofs of Lemma 2.6, Theorems 2.9 and 2.10, Corollary 2.11, and Theorem 3.1. Section B is devoted to a brief account of quantum estimation theory, including quantum logarithmic derivatives, the commutation operator, the Holevo bound, estimation theory for quantum Gaussian shift models and for pure state models.

REFERENCES

[1] AMARI, S.-I. and NAGAOKA, H. (2000). *Methods of Information Geometry. Translations of Mathematical Monographs* **191**. Amer. Math. Soc., Providence, RI. MR1800071

[2] FUJIWARA, A. and NAGAOKA, H. (1995). Quantum Fisher metric and estimation for pure state models. *Phys. Lett. A* **201** 119–124. MR1329961

[3] FUJIWARA, A. and NAGAOKA, H. (1999). An estimation theoretical characterization of coherent states. *J. Math. Phys.* **40** 4227–4239. MR1708381

[4] GILL, R. D. and GUȚĂ, M. (2012). On asymptotic quantum statistical inference. *IMS Collections From Probability to Statistics and Back*: *High-Dimensional Models and Processes* **9** 105–127.

[5] GILL, R. D. and LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* **1** 59–79. MR1354456

[6] GILL, R. D. and MASSAR, S. (2000). State estimation for large ensembles. *Phys. Rev. A* (3) **61** 042312.

[7] GUȚĂ, M. and BUTUCEA, C. (2010). Quantum U-statistics. *J. Math. Phys.* **51** 102202, 24. MR2761295

[8] GUȚĂ, M. and JENČOVÁ, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys.* **276** 341–379. MR2346393

[9] GUȚĂ, M. and KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* (3) **73** 052108, 15. MR2229156

[10] HAYASHI, M. (1997). A linear programming approach to attainable Cramér–Rao type bounds. In *Quantum Communication, Computing, and Measurement* 99–108. Plenum, New York.

[11] HAYASHI, M. and MATSUMOTO, K. (2008). Asymptotic performance of optimal state estimation in qubit system. *J. Math. Phys.* **49** 102101, 33. MR2464597

[12] HOLEVO, A. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed. *Quaderni. Monographs* **1**. Edizioni della Normale, Pisa. MR2797301

[13] JAKŠIĆ, V., PAUTRAT, Y. and PILLET, C. A. (2010). A quantum central limit theorem for sums of independent identically distributed random variables. *J. Math. Phys.* **51** 015208, 8. MR2605841

[14] KAHN, J. and GUȚĂ, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. MR2506764

[15] KUBO, F. and ANDO, T. (1979/80). Means of positive linear operators. *Math. Ann.* **246** 205–224. MR0563399

[16] MANUCEAU, J., SIRUGUE, M., TESTARD, D. and VERBEURE, A. (1973). The smallest C^*-algebra for canonical commutations relations. *Comm. Math. Phys.* **32** 231–243. MR0339715

[17] MATSUMOTO, K. (2002). A new approach to the Cramér–Rao-type bound of the pure-state model. *J. Phys. A* **35** 3111–3123. MR1913859

[18] NAGAOKA, H. (1991). A generalization of the simultaneous diagonalization of Hermitian matrices and its relation to quantum estimation theory (in Japanese). *Transactions of the Japan Society for Industrial and Applied Mathematics* **1** 305–318.

[19] PETZ, D. (1990). *An Invitation to the Algebra of Canonical Commutation Relations*. *Leuven Notes in Mathematical and Theoretical Physics*. *Series A*: *Mathematical Physics* **2**. Leuven Univ. Press, Leuven. MR1057180

[20] PETZ, D. (2008). *Quantum Information Theory and Quantum Statistics*. Springer, Berlin. MR2363070

[21] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[22] YAMAGATA, K. (2011). Efficiency of quantum state tomography for qubits. *Int*. *J*. *Quantum Inf*. **9** 1167–1183. MR2931457

[23] YAMAGATA, K., FUJIWARA, A. and GILL, R. D. (2013). Supplement to "Quantum local asymptotic normality based on a new quantum likelihood ratio." DOI:10.1214/13-AOS1147SUPP.

K. YAMAGATA
A. FUJIWARA
DEPARTMENT OF MATHEMATICS
OSAKA UNIVERSITY
1-1 MACHIKANEYAMA
TOYONAKA, OSAKA 560-0043
JAPAN
E-MAIL: k-yamagata@cr.math.sci.osaka-u.ac.jp
       fujiwara@math.sci.osaka-u.ac.jp

R. D. GILL
MATHEMATICAL INSTITUTE
LEIDEN UNIVERSITY
P.O. BOX 9512
2300 RA LEIDEN
THE NETHERLANDS
E-MAIL: gill@math.leidenuniv.nl