Strong consistency and asymptotic efficiency for adaptive quantum estimation problems

Akio Fujiwara
Department of Mathematics, Osaka University
Toyonaka, Osaka 560-0043, Japan

Abstract

It is shown that for an adaptive quantum estimation scheme based on locally unbiased measurements, the sequence of maximum likelihood estimators is strongly consistent and asymptotically efficient.

1 Introduction

Let $\mathcal{S} = \{\rho_\theta;\, \theta = (\theta^1, \dots, \theta^d) \in \Theta\}$ be a smooth parametric family of density operators, called a *quantum statistical model*, on a Hilbert space $\mathcal{H}$ with compact parameter space $\Theta \subset \mathbb{R}^d$. Our problem is to estimate the true value of θ by means of a certain quantum estimation scheme. An *estimator* is represented by a pair $(M, \check{\theta})$, where $M = \{M(x);\, x \in \mathcal{X}\}$ is a positive operator-valued measure (POVM) that takes values on a set $\mathcal{X}$, and $\check{\theta} : \mathcal{X} \to \Theta$ is a map that gives the estimated value $\check{\theta}(x)$ from each observed data $x \in \mathcal{X}$. A POVM is also called a measurement. The observed data $x \in \mathcal{X}$ has probability density

$$f(x; \theta, M) := \operatorname{Tr} \rho_\theta M(x), \tag{1}$$

which depends on both the parameter θ and the measurement M.

An estimator $(M, \check{\theta})$ is called *unbiased* if

$$E_\theta[M, \check{\theta}] = \theta \tag{2}$$

is satisfied for all $\theta \in \Theta$, where $E_\theta[\,\cdot\,]$ denotes the expectation with respect to the density (1). An estimator $(M, \check{\theta})$ is called *locally unbiased* [1] at a given point $\theta_0 \in \Theta$ if the condition (2) is satisfied around $\theta = \theta_0$ up to the first order of the Taylor expansion, that is, if

$$E_{\theta_0}[M, \check{\theta}] = \theta_0, \qquad \frac{\partial}{\partial \theta^j} E_\theta[M, \check{\theta}^i]\bigg|_{\theta=\theta_0} = \delta^i_j, \qquad (i, j = 1, \dots, d). \tag{3}$$

Clearly, an estimator is unbiased if and only if it is locally unbiased at all $\theta \in \Theta$.

It is well known that an estimator $(M, \check{\theta})$ that is locally unbiased at θ_0 satisfies the quantum Cramér-Rao inequality [1, 2]

$$V_{\theta_0}[M, \check{\theta}] \geq (J_{\theta_0})^{-1}, \tag{4}$$

where $V_\theta[\,\cdot\,]$ denotes the covariance matrix, and J_θ is the quantum Fisher information matrix of the model $\mathcal{S}$ whose (i,j)th entry is given by

$$(J_\theta)_{ij} = \frac{1}{2}\mathrm{Tr}\,\rho_\theta(L_{\theta,i}L_{\theta,j} + L_{\theta,j}L_{\theta,i}),$$

where $L_{\theta,i}$ is the ith symmetric logarithmic derivative (SLD) defined by the selfadjoint operator satisfying the equation

$$\frac{\partial\rho_\theta}{\partial\theta^i} = \frac{1}{2}\left(L_{\theta,i}\rho_\theta + \rho_\theta L_{\theta,i}\right).$$

As to the achievability of the lower bound in (4), there is a serious difficulty intrinsic to quantum statistics: due to the noncommutative nature of the SLDs, there is, in general, no locally unbiased estimator (LUE) $(M,\check{\theta})$ that satisfies the equality $V_{\theta_0}[M,\check{\theta}] = (J_{\theta_0})^{-1}$. To avoid this difficulty, we must introduce a different criterion for measuring the goodness of an estimator. A possible alternative is the weighted sum $\mathrm{tr}\,GV_\theta[M,\check{\theta}]$ of covariances [1, 2], where G is a positive definite weight matrix, and we seek, for each value of the parameter θ, the best LUE $(M,\check{\theta})$ that minimizes the scalar quantity $\mathrm{tr}\,GV_\theta[M,\check{\theta}]$.

Even if such a modified strategy is adopted, we still have a difficulty which often becomes the target of criticism: since the best LUE for estimating the parameter θ depends, in general, on the unknown parameter θ itself, the estimation strategy based on LUEs would be impracticable. In a different yet analogous context, Cochran [3] ingeniously described this kind of dilemma as follows: "You tell me the value of θ and I promise to design the best experiment for estimating θ." Apparently, such a criticism is strong enough to invalidate estimation strategies based on LUEs.

To surmount this difficulty, Nagaoka [4] advocated an adaptive quantum estimation scheme as follows. Suppose that, by prior investigation of the quantum statistical model $\mathcal{S}$, one has the list of optimal LUEs $\big(M(\,\cdot\,;\theta),\check{\theta}(\,\cdot\,;\theta)\big)$ for each $\theta\in\Theta$. One begins with an arbitrary initial guess $\hat{\theta}_0\in\Theta$, and applies the measurement $M(\,\cdot\,;\hat{\theta}_0)$ that is optimal at $\hat{\theta}_0$. Suppose the data x_1 is observed. One then applies the maximum likelihood method to the likelihood function $L_1(\theta) = f(x_1;\theta,M(\,\cdot\,;\hat{\theta}_0))$, to obtain the next guess $\hat{\theta}_1$. At stage n (≥ 2), one applies the measurement $M(\,\cdot\,;\hat{\theta}_{n-1})$, where $\hat{\theta}_{n-1}$ is the maximum likelihood estimator (MLE) obtained at the previous stage. The likelihood function is then given by

$$L_n(\theta) := \prod_{i=1}^{n} f(x_i;\theta,M(\,\cdot\,;\hat{\theta}_{i-1})),$$

where x_i is the observed data at stage i, and one obtains the nth MLE $\hat{\theta}_n$ that maximizes $L_n(\theta)$. It is quite natural to expect that, under certain regularity conditions, the sequence $\hat{\theta}_n$ of MLEs would converge to the true value of the parameter θ [4]. However, such an asymptotic property has not been investigated so far in the context of adaptive quantum estimation.

The purpose of this paper is to establish the strong consistency and the asymptotic efficiency for the above sequence $\hat{\theta}_n$ of MLEs. For simplicity, we confine ourselves to finite quantum systems. The paper is organized as follows. In Section 2, we show that the set $\mathcal{X}$ of observed data can be taken to be a finite set $\mathcal{X} = \{1,2,\ldots,n_0\}$, where $n_0 = (\dim\mathcal{H})^2 + d(d+1)$. In Section 3, we point out that Nagaoka's adaptive quantum estimation scheme is a variant of the so-called sequential design problem [5, 6, 7, 8], and spell out the strong consistency and the asymptotic efficiency in a form relevant to our problem. The proofs of those results are deferred to Sections 4 and 5 respectively. In Section 6, we apply those results to Nagaoka's adaptive quantum estimation scheme, and clarify an

operational meaning of the Cramér-Rao bound. In Section 7, we remark that, with some additional regularity conditions, those asymptotic properties can be naturally extended to infinite quantum systems.

2 Size of optimal LUE

We first show that the size of an optimal measurement for estimating θ can be taken to be at most $(\dim \mathcal{H})^2 + d(d+1)$. Let $\mathcal{M}_n$ be the set of POVMs that take values on $\mathcal{X}_n = \{1, \ldots, n\}$, that is

$$\mathcal{M}_n = \left\{ M = (M(1), \ldots, M(n)) \; ; \; \forall M(x) \geq 0, \; \sum_{x=1}^{n} M(x) = I \right\}.$$

There is a natural convex structure on $\mathcal{M}_n$: for $M, N \in \mathcal{M}_n$ and $0 \leq \lambda \leq 1$,

$$(1-\lambda)M + \lambda N = ((1-\lambda)M(x) + \lambda N(x))_{x=1}^{n}.$$

Moreover, $\mathcal{M}_n$ is naturally regarded as a compact set [9]. For an $M \in \mathcal{M}_n$, we call the set $\{x; M(x) \neq 0\}$ the *support* of M and denote it as $\mathrm{supp}\,(M)$. Further, given $\check{\theta} : \; \mathcal{X} \to \Theta$, let

$$\mathcal{M}_n(\check{\theta}) = \left\{ M \in \mathcal{M}_n \, ; \; (M, \check{\theta}) \text{ is locally unbiased at a given } \theta \in \Theta \right\}.$$

For a subset A of a convex set, a point $x \in A$ is called *extreme* if x cannot be represented as a nontrivial convex combination of points in A. The totality of extreme points in A is called the *extreme boundary* and is denoted by $\partial_e A$.

Lemma 1 *$\mathcal{M}_n(\check{\theta})$ is a convex compact subset of $\mathcal{M}_n$. Thus*

$$\mathcal{M}_n(\check{\theta}) = \mathrm{co}\, \partial_e \mathcal{M}_n(\check{\theta}), \tag{5}$$

where co *denotes the convex hull.*

Proof Convexity follows from the fact that convex combination of POVMs in $\mathcal{M}_n(\check{\theta})$ preserves the local unbiasedness conditions:

$$\sum_{x=1}^{n} \check{\theta}^i(x) \, \mathrm{Tr}\, \rho_\theta M(x) = \theta^i, \qquad (\forall i = 1, \ldots, d) \tag{6}$$

$$\sum_{x=1}^{n} \check{\theta}^i(x) \, \mathrm{Tr}\, \frac{\partial \rho_\theta}{\partial \theta^j} M(x) = \delta^i_j, \qquad (\forall i, \forall j = 1, \ldots, d) \tag{7}$$

and compactness follows from the fact that, due to (6) (7), $\mathcal{M}_n(\check{\theta})$ is closed. The relation (5) is an immediate consequence of Krein-Milman's extreme point theorem [10]. □

Let us call $M = (M(1), \ldots, M(n)) \in \mathcal{M}_n(\check{\theta})$ *linearly quasi-independent with respect to* $\check{\theta}$ if the three conditions

(i) $\sum_x \alpha(x) \, M(x) = 0,$

(ii) $\sum_x \alpha(x)\check{\theta}^i(x)\,\mathrm{Tr}\,\rho_\theta M(x) = 0$, for all $i = 1, \dots d$,

(iii) $\sum_x \alpha(x)\check{\theta}^i(x)\,\mathrm{Tr}\,\dfrac{\partial \rho_\theta}{\partial \theta^j} M(x) = 0$, for all $i, j = 1, \dots d$,

for a sequence $(\alpha(1), \dots, \alpha(n)) \in \mathbb{R}^n$ imply $\alpha(x) = 0$ for all $x \in \mathrm{supp}\,(M)$.

Lemma 2

$$\partial_e \mathcal{M}_n(\check{\theta}) \subset \left\{ M \in \mathcal{M}_n(\check{\theta})\ ;\ M \textit{ is linearly quasi-independent with respect to } \check{\theta} \right\}$$

Proof Assume that $M \in \mathcal{M}_n(\check{\theta})$ is *not* linearly quasi-independent with respect to $\check{\theta}$. Then there is a sequence $(\alpha(x))_{x=1}^n$ that fulfills conditions (i)-(iii) and that $(\alpha(x))_{x \in \mathrm{supp}\,(M)}$ are not all zero. Let

$$N^{(+)}(x) = (1 + \varepsilon\,\alpha(x))M(x), \qquad N^{(-)}(x) = (1 - \varepsilon\,\alpha(x))M(x)$$

for sufficiently small $\varepsilon > 0$. Then $N^{(\pm)} = (N^{(\pm)}(x))_{x=1}^n$ are different elements of $\mathcal{M}_n(\check{\theta})$. In fact,

$$\sum_x N^{(\pm)}(x) = \sum_x M(x) \pm \varepsilon \sum_x \alpha(x) M(x) = \sum_x M(x) = I,$$

where (i) is used in the second equality. Similarly, the local unbiasedness conditions are verified by using (ii) and (iii) as follows:

$$\sum_x \check{\theta}^i(x)\ \mathrm{Tr}\,\rho_\theta N^{(\pm)}(x) = \sum_x \check{\theta}^i(x)\ \mathrm{Tr}\,\rho_\theta M(x) \pm \varepsilon \sum_x\ \alpha(x)\check{\theta}^i(x)\ \mathrm{Tr}\,\rho_\theta M(x) = \theta^i,$$

$$\sum_x \check{\theta}^i(x)\ \mathrm{Tr}\,\frac{\partial \rho_\theta}{\partial \theta^j} N^{(\pm)}(x) = \sum_x \check{\theta}^i(x)\ \mathrm{Tr}\,\frac{\partial \rho_\theta}{\partial \theta^j} M(x) \pm \varepsilon \sum_x\ \alpha(x)\check{\theta}^i(x)\ \mathrm{Tr}\,\frac{\partial \rho_\theta}{\partial \theta^j} M(x) = \delta^i_j.$$

Since $M = \frac{1}{2}N^{(+)} + \frac{1}{2}N^{(-)}$, M is not an extreme point. □

Corollary 3 *For $M \in \partial_e \mathcal{M}_n(\check{\theta})$*

$$|\mathrm{supp}\,(M)| \le (\dim \mathcal{H})^2 + d + d^2.$$

Proof Conditions (i)-(iii) of the quasi-independence impose at most $(\dim \mathcal{H})^2 + d + d^2$ linear constraints on $(\alpha(x))_{x \in \mathrm{supp}\,(M)}$. As a consequence, $M \in \mathcal{M}_n(\check{\theta})$ cannot be linearly quasi-independent with respect to $\check{\theta}$ if $|\mathrm{supp}\,(M)| > (\dim \mathcal{H})^2 + d + d^2$. □

Lemma 4 *For all $M \in \mathcal{M}_n(\check{\theta})$, there is an $N \in \partial_e \mathcal{M}_n(\check{\theta})$ such that*

$$\mathrm{tr}\, G V_\theta[M, \check{\theta}] \ge \mathrm{tr}\, G V_\theta[N, \check{\theta}].$$

Proof As the covariance

$$v_\theta^{ij}[M,\check{\theta}] = \sum_{x=1}^{n} \left(\check{\theta}^i(x) - \theta^i\right) \left(\check{\theta}^j(x) - \theta^j\right) \operatorname{Tr} \rho_\theta M(x) \tag{8}$$

is linear in M, so is $\operatorname{tr} GV_\theta[M,\check{\theta}]$. Assume that $M = (1-\lambda)N^{(1)} + \lambda N^{(2)}$ for $N^{(1)}, N^{(2)} \in \mathcal{M}_n(\check{\theta})$ and $0 \leq \lambda \leq 1$. Then

$$\begin{aligned} \operatorname{tr} GV_\theta[M,\check{\theta}] &= (1-\lambda) \operatorname{tr} GV_\theta[N^{(1)},\check{\theta}] + \lambda \operatorname{tr} GV_\theta[N^{(2)},\check{\theta}] \\ &\geq \min\left\{\operatorname{tr} GV_\theta[N^{(1)},\check{\theta}],\ \operatorname{tr} GV_\theta[N^{(2)},\check{\theta}]\right\}. \end{aligned}$$

This inequality, as well as Lemma 1, proves the lemma. □

Now Corollary 3 and Lemma 4 immediately leads to the following:

Theorem 5 *The size of the support of an optimal measurement that attains*

$$\min_{M \in \mathcal{M}_n(\check{\theta})} \operatorname{tr} GV_\theta[M,\check{\theta}]$$

can be taken to be at most $(\dim \mathcal{H})^2 + d(d+1)$.

Since

$$\inf_{(M,\check{\theta}):\,\text{LUE at } \theta} \operatorname{tr} GV_\theta[M,\check{\theta}] = \inf_{n,\check{\theta}} \min_{M \in \mathcal{M}_n(\check{\theta})} \operatorname{tr} GV_\theta[M,\check{\theta}],$$

we can conclude that the size of an optimal measurement for estimating a d-dimensional quantum statistical model on a Hilbert space $\mathcal{H}$ can be taken to be at most $n_0 := (\dim \mathcal{H})^2 + d(d+1)$. In particular,

$$\inf_{(M,\check{\theta}):\,\text{LUE at } \theta} \operatorname{tr} GV_\theta[M,\check{\theta}] = \inf_{\check{\theta}} \min_{M \in \mathcal{M}_{n_0}(\check{\theta})} \operatorname{tr} GV_\theta[M,\check{\theta}].$$

Some remarks are in order. By considering a rank-one refinement of POVMs, we can confine ourselves, without loss of generality, to measurements $M = (M(x))_{x=1}^n$ that satisfy $\operatorname{rank} M(x) \leq 1$ for all $x = 1, \ldots, n$. Let us introduce

$$\mathcal{M}_n^{(e)}(\check{\theta}) = \{M \in \mathcal{M}_n(\check{\theta})\ ;\ \operatorname{rank} M(x) \leq 1 \text{ for all } x = 1, \ldots, n\},$$

which is not a convex subset of $\mathcal{M}_n(\check{\theta})$ but is an *extremal* subset of $\mathcal{M}_n(\check{\theta})$, see [9].

Lemma 6

$$\partial_e \mathcal{M}_n(\check{\theta}) \cap \mathcal{M}_n^{(e)}(\check{\theta}) = \left\{M \in \mathcal{M}_n^{(e)}(\check{\theta})\ ;\ M \textit{ is linearly quasi-independent with respect to } \check{\theta}\right\}.$$

Proof Left-hand side (LHS) $\subset$ right-hand side (RHS) is obvious from Lemma 2. We show that LHS $\supset$ RHS. Let M be an arbitrary element in RHS, and suppose that M is written as $M = \frac{1}{2}N^{(1)} + \frac{1}{2}N^{(2)}$ where $N^{(1)}, N^{(2)} \in \mathcal{M}_n(\check{\theta})$. To prove $M \in \partial_e \mathcal{M}_n(\check{\theta})$, it sufficies to show that $N^{(1)} = N^{(2)}$.

Since $0 \le N^{(k)}(x) \le 2M(x)$ and $\operatorname{rank} M(x) \le 1$ for all $x \in \mathcal{X}_n$ and $k \in \{1,2\}$, there is a constant $c^{(k)}(x)$ such that $N^{(k)}(x) = c^{(k)}(x)M(x)$. Substituting this relation into

$$\sum_x N^{(k)}(x) = I, \quad \sum_x \check{\theta}^i(x) \operatorname{Tr} \rho_\theta N^{(k)}(x) = \theta^i, \quad \sum_x \check{\theta}^i(x) \operatorname{Tr} \frac{\partial \rho_\theta}{\partial \theta^j} N^{(k)}(x) = \delta^i_j,$$

we have

$$\begin{aligned}
&\sum_x \left(c^{(1)}(x) - c^{(2)}(x)\right) M(x) = 0, \\
&\sum_x \left(c^{(1)}(x) - c^{(2)}(x)\right) \check{\theta}^i(x) \operatorname{Tr} \rho_\theta M(x) = 0, \\
&\sum_x \left(c^{(1)}(x) - c^{(2)}(x)\right) \check{\theta}^i(x) \operatorname{Tr} \frac{\partial \rho_\theta}{\partial \theta^j} M(x) = 0.
\end{aligned}$$

Since M is assumed linearly quasi-independent with respect to $\check{\theta}$, we have $c^{(1)}(x) - c^{(2)}(x) = 0$ for all $x \in \operatorname{supp}(M)$. This implies that $N^{(1)} = N^{(2)}$. □

Corollary 7 *For all $M \in \mathcal{M}_n^{(e)}(\check{\theta})$, there is an $N \in \partial_e \mathcal{M}_n(\check{\theta}) \cap \mathcal{M}_n^{(e)}(\check{\theta})$ such that*

$$\operatorname{tr} GV_\theta[M, \check{\theta}] \ge \operatorname{tr} GV_\theta[N, \check{\theta}].$$

In particular, $|\operatorname{supp}(N)| \le (\dim \mathcal{H})^2 + d(d+1)$.

3 Strong consistency and asymptotic efficiency of MLE

We next observe that Nagaoka's quantum adaptive estimation scheme presented in Section 1 is a variant of the so-called sequential design problem: a *sequential design problem* allows at each stage an experiment E to be taken from an experiment space $\mathcal{E}$. The observed data $x_n \in \mathcal{X}$ at time n has probability density $f(x_n; \theta, E_n)$, with respect to some σ-finite measure μ on $\mathcal{X}$, which depends on both the parameter $\theta \in \Theta$ and the experiment $E_n \in \mathcal{E}$ selected at stage n. It is assumed that E_n is measurable with respect to the natural filtration $\mathcal{F}_{n-1} := \sigma(X_1, \ldots, X_{n-1})$, that is, E_n is chosen according to the information of the past data $X_1, \ldots, X_{n-1}$. The likelihood function is therefore given by

$$L_n(\theta) := \prod_{i=1}^n f(x_i; \theta, E_i).$$

Let $\hat{\theta}_n$ denote the MLE of θ.

In our quantum estimation problem, according to Section 2, we can assume that $\mathcal{X}$ is a finite set (with μ the counting measure), and that Θ and $\mathcal{E}$ are both compact. We further assume the following regularity conditions:

(RC1) $f(x; \theta, e)$ is positive for all (x, θ, e), and is continuous in (x, θ, e),

(RC2) $\mu(\{x;\ f(x;\theta,e) \ne f(x;\theta',e)\}) > 0$ for any $\theta \ne \theta'$ and $e \in \mathcal{E}$.

Under these mild conditions, we can obtain the following strong consistency.

Theorem 8 *Suppose that the true value of the parameter, θ_0, is an interior point of Θ. Then*

$$\hat{\theta}_n \longrightarrow \theta_0, \qquad P_{\theta_0}\text{-a.s.} \tag{9}$$

Chernoff [5] proved strong consistency of the MLE when Θ and $\mathcal{E}$ are both finite sets (but $\mathcal{X}$ is arbitrary). McCormick et al. [6] extended Chernoff's method to the case of compact Θ and $\mathcal{E}$. Datta [7] further studied the case when Θ is a precompact metric space and $\mathcal{X}$ and $\mathcal{E}$ are arbitrary, and proved weak consistency of the MLE. Although Theorem 8 is subsumed by McCormick et al. [6], we will give a simplified proof in Section 4 for the reader's convenience.

In order to establish asymptotic efficiency, we further assume the following smoothness condition:

(RC3) $f(x;\theta,e)$ is thrice differentiable in θ in the interior of Θ, and the derivatives

$$\frac{\partial}{\partial\theta^i} f(x;\theta,e), \quad \frac{\partial^2}{\partial\theta^i\partial\theta^j} f(x;\theta,e), \quad \frac{\partial^3}{\partial\theta^i\partial\theta^j\partial\theta^k} f(x;\theta,e)$$

are all continuous in (x,θ,e).

Then we have the following

Theorem 9 *Suppose that the true value of the parameter, θ_0, is an interior point of Θ, and that the nth experiment E_n ($n \geq 2$) is chosen according to the MLE $\hat{\theta}_{n-1}$ as*

$$E_n(X_1,\ldots,X_{n-1}) = \hat{E}\left(\hat{\theta}_{n-1}(X_1,\ldots,X_{n-1})\right),$$

where $\hat{E}:\Theta\to\mathcal{E}$ is a continuous function. Then

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \longrightarrow N\left(0, J(\theta_0|\hat{E}(\theta_0))^{-1}\right), \qquad \textit{in distribution}, \tag{10}$$

where $J(\theta|e)$ is the Fisher information matrix at θ under the experiment e, and its (i,j)th entry is

$$J(\theta|e)_{ij} := E_\theta\left[\left(\frac{\partial}{\partial\theta^i}\log f(X;\theta,e)\right)\left(\frac{\partial}{\partial\theta^j}\log f(X;\theta,e)\right)\right].$$

Chaudhuri and Mykland [8] studied an adaptive two stage design for a general nonlinear regression problem, and obtained a consistent sequence of roots of the likelihood equation (local maximizers of the likelihood) and its asymptotic normality under the criterion of "D-optimality." In contrast to their results, this paper deals with a true maximizer of the likelihood function, and with the asymptotic efficiency on condition that the optimal experiment $\hat{E}(\theta)\in\mathcal{E}$ for each $\theta\in\Theta$ is known in advance. The proof of Theorem 9, which is a modification of that in [8], is given in Section 5.

4 Proof of Theorem 8

We prove Theorem 8 in a series of lemmas. Let

$$R(\theta_0,\theta,e) := \log\frac{f(X;\theta,e)}{f(X;\theta_0,e)}$$

and, for $\varepsilon > 0$, let

$$R^{\varepsilon}(\theta_0, \theta, e) := \sup_{\theta' \in N_{\varepsilon}(\theta)} R(\theta_0, \theta', e),$$

where $N_{\varepsilon}(\theta)$ is the open ε-neighborhood of θ in Θ.

Lemma 10 *For any θ with $\theta \neq \theta_0$, it holds that*

$$\sup_{e \in \mathcal{E}} E_{\theta_0}\left[e^{R(\theta_0, \theta, e)/2}\right] < 1. \tag{11}$$

Proof Since the function

$$g(e) := E_{\theta_0}\left[e^{R(\theta_0, \theta, e)/2}\right] = \sum_{x \in \mathcal{X}} \sqrt{f(x; \theta, e)}\, \sqrt{f(x; \theta_0, e)}$$

is continuous on the compact set $\mathcal{E}$, there is a point $e_0 \in \mathcal{E}$ such that $g(e) \leq g(e_0)$ for all $e \in \mathcal{E}$. For this e_0, we have by the identifiability condition (RC2) that there is a $\gamma > 0$ such that the set

$$A_{\gamma} := \left\{x;\ \left|\sqrt{f(x; \theta, e_0)} - \sqrt{f(x; \theta_0, e_0)}\right| > \gamma\right\}$$

has positive μ-measure (cardinality). Since, for $x \in A_{\gamma}$,

$$\sqrt{f(x; \theta, e_0)}\, \sqrt{f(x; \theta_0, e_0)} < \frac{1}{2}\left\{f(x; \theta, e_0) + f(x; \theta_0, e_0)\right\} - \frac{\gamma^2}{2},$$

we have

$$\begin{aligned} g(e_0) &= \left(\sum_{x \in A_{\gamma}} + \sum_{x \in A_{\gamma}^c}\right) \sqrt{f(x; \theta, e_0)}\, \sqrt{f(x; \theta_0, e_0)} \\ &< \sum_{x \in A_{\gamma}} \left(\frac{1}{2}\left\{f(x; \theta, e_0) + f(x; \theta_0, e_0)\right\} - \frac{\gamma^2}{2}\right) + \sum_{x \in A_{\gamma}^c} \frac{1}{2}\left\{f(x; \theta, e_0) + f(x; \theta_0, e_0)\right\} \\ &= 1 - \frac{\gamma^2}{2}\, \mu(A_{\gamma}) < 1. \end{aligned}$$

□

Lemma 11 *For any θ with $\theta \neq \theta_0$, it holds that*

$$\lim_{\varepsilon \downarrow 0} \sup_{e \in \mathcal{E}} E_{\theta_0}\left[e^{R^{\varepsilon}(\theta_0, \theta, e)/2}\right] < 1. \tag{12}$$

Proof For

$$g^{\varepsilon}(e) := E_{\theta_0}\left[e^{R^{\varepsilon}(\theta_0, \theta, e)/2}\right] = \sum_{x \in \mathcal{X}} \sup_{\theta' \in N_{\varepsilon}(\theta)} \sqrt{f(x; \theta', e)}\, \sqrt{f(x; \theta_0, e)},$$

we have (i) $g^\varepsilon(e) \downarrow g(e)$ as $\varepsilon \downarrow 0$ for each $e \in \mathcal{E}$, and (ii) g^ε is continuous on $\mathcal{E}$ for all $\varepsilon > 0$ (see Appendix A). Therefore, by Dini's theorem, the convergence is uniform in e. That is, for any $\delta > 0$, there is an $\varepsilon_0 > 0$ such that

$$g(e) \le g^\varepsilon(e) < g(e) + \delta$$

for all $e \in \mathcal{E}$ and $\varepsilon \in (0, \varepsilon_0)$. As a consequence

$$\sup_{e\in\mathcal{E}} g(e) \le \lim_{\varepsilon\downarrow 0} \sup_{e\in\mathcal{E}} g^\varepsilon(e) \le \sup_{e\in\mathcal{E}} g(e) + \delta.$$

Since $\delta > 0$ is arbitrary, we have

$$\lim_{\varepsilon\downarrow 0} \sup_{e\in\mathcal{E}} g^\varepsilon(e) = \sup_{e\in\mathcal{E}} g(e).$$

Thus (12) follows from (11). □

Lemma 12 *For any θ with $\theta \neq \theta_0$, there exist $\varepsilon > 0$ and $b > 0$ such that for any $n \in \mathbb{N}$ and any strategy $\{E_i\}_{1\le i\le n} \subset \mathcal{E}$, we have*

$$P_{\theta_0}\left[\sum_{i=1}^n R^\varepsilon(\theta_0, \theta, E_i) > 0\right] \le e^{-bn}. \tag{13}$$

Proof According to Lemma 11, there exist $\varepsilon > 0$ and $b > 0$ such that

$$\sup_{e\in\mathcal{E}} E_{\theta_0}\left[e^{R^\varepsilon(\theta_0,\theta,e)/2}\right] = e^{-b}. \tag{14}$$

For each n, let

$$M_n := \prod_{i=1}^n e^{R^\varepsilon(\theta_0,\theta,E_i)/2}.$$

The stochastic process $\{M_n\}_n$ enjoys the following supermartingale property [11]:

$$\begin{aligned} E_{\theta_0}\left[M_n|\mathcal{F}_{n-1}\right] &= E_{\theta_0}\left[M_{n-1}\, e^{R^\varepsilon(\theta_0,\theta,E_n)/2} \,\middle|\, \mathcal{F}_{n-1}\right] \\ &= M_{n-1}\, E_{\theta_0}\left[e^{R^\varepsilon(\theta_0,\theta,E_n)/2} \,\middle|\, \mathcal{F}_{n-1}\right] \\ &\le e^{-b}\, M_{n-1}, \qquad P_{\theta_0}\text{-a.s.} \end{aligned}$$

Here, the second equality follows from the fact that M_{n-1} is $\mathcal{F}_{n-1}$-measurable [12], and the inequality follows from (14). Then by induction

$$E_{\theta_0}\left[M_n\right] = E_{\theta_0}\left[M_n|\mathcal{F}_{n-1}|\mathcal{F}_{n-2}|\cdots|\mathcal{F}_0\right] \le e^{-bn},$$

and by Markov's inequality

$$P_{\theta_0}\left[\sum_{i=1}^n R^\varepsilon(\theta_0, \theta, E_i) > 0\right] = P_{\theta_0}[M_n > 1] \le E_{\theta_0}[M_n] \le e^{-bn}.$$

□

Proof of Theorem 8. Fix $a > 0$ arbitrarily. Since $N_a(\theta_0)^c$ is a compact subset of Θ, for some finite subset $\{\theta_1, \ldots, \theta_r\} \subset N_a(\theta_0)^c$, we have

$$N_a(\theta_0)^c \subset \bigcup_{\lambda=1}^{r} N_{\varepsilon_\lambda}(\theta_\lambda),$$

where the ε_λ are chosen so that (13) holds with $\theta = \theta_\lambda$ and $b = b_\lambda$, $\lambda = 1, \ldots, r$. Then

$$\begin{aligned}
P_{\theta_0}\left[|\hat{\theta}_n - \theta_0| \geq a\right] &= P_{\theta_0}\left[\max_{\theta \in N_a(\theta_0)^c} \sum_{i=1}^{n} R(\theta_0, \theta, E_i) > 0\right] \\
&\leq P_{\theta_0}\left[\max_{1 \leq \lambda \leq r} \sup_{\theta \in N_{\varepsilon_\lambda}(\theta_\lambda)} \sum_{i=1}^{n} R(\theta_0, \theta, E_i) > 0\right] \\
&\leq P_{\theta_0}\left[\max_{1 \leq \lambda \leq r} \sum_{i=1}^{n} R^{\varepsilon_\lambda}(\theta_0, \theta_\lambda, E_i) > 0\right] \\
&\leq \sum_{\lambda=1}^{r} P_{\theta_0}\left[\sum_{i=1}^{n} R^{\varepsilon_\lambda}(\theta_0, \theta_\lambda, E_i) > 0\right] \\
&\leq \sum_{\lambda=1}^{r} e^{-b_\lambda n} \\
&\leq r\, e^{-b_0 n},
\end{aligned}$$

where $b_0 := \min_{1 \leq \lambda \leq r} b_\lambda$ is a positive constant. Then we have by the Borel-Cantelli lemma that

$$P_{\theta_0}\left[|\hat{\theta}_n - \theta_0| \geq a, \text{ i.o.}\right] = 0.$$

Since this holds for any $a > 0$, the strong consistency (9) is established. □

5 Proof of Theorem 9

According to the Cramér-Wold device (see, for example, [13, pp. 48-49]), it suffices to treat the case when θ is a one-dimensional parameter. For notational simplicity, let

$$\ell(\theta|x, e) := \log\, f(x; \theta, e),$$

and let

$$\ell_n(\theta) := \sum_{i=1}^{n} \ell(\theta|X_i, E_i).$$

By assumption (RC3), we have the 2nd order Taylor expansion of $\ell_n'(\theta)$ around θ_0 as follows:

$$\ell_n'(\theta) = \ell_n'(\theta_0) + \ell_n''(\theta_0)\,(\theta - \theta_0) + \frac{1}{2}\,\ell_n'''(\theta^*)\,(\theta - \theta_0)^2,$$

where θ^* is a point satisfying $|\theta^* - \theta_0| < |\theta - \theta_0|$. Since LHS vanishes at $\theta = \hat{\theta}_n$, we have

$$0 = \frac{\ell_n'(\theta_0)}{\sqrt{n}} + \sqrt{n}\,(\hat{\theta}_n - \theta_0)\left\{\frac{\ell_n''(\theta_0)}{n} + \frac{\ell_n'''(\theta^*)}{2n}\,(\hat{\theta}_n - \theta_0)\right\},$$

or

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{\dfrac{\ell_n'(\theta_0)}{\sqrt{n}}}{-\dfrac{\ell_n''(\theta_0)}{n} - \dfrac{\ell_n'''(\theta^*)}{2n}\,(\hat{\theta}_n - \theta_0)}. \tag{15}$$

Since the MLE $\hat{\theta}_n$ exhibits the strong consistency $\hat{\theta}_n \to \theta_0$ (Theorem 8), it suffices to show that

$$\frac{\ell_n'(\theta_0)}{\sqrt{n}} \longrightarrow N\left(0, J(\theta_0|\hat{E}(\theta_0))\right), \qquad \text{in distribution,} \tag{16}$$

that

$$\frac{\ell_n''(\theta_0)}{n} \longrightarrow -J(\theta_0|\hat{E}(\theta_0)), \qquad \text{in probability,} \tag{17}$$

and that for sufficiently large n,

$$\frac{\ell_n'''(\theta^*)}{n} \quad \text{is bounded in probability.} \tag{18}$$

The desired result $\sqrt{n}\,(\hat{\theta}_n - \theta_0) \to N\left(0, J(\theta_0|\hat{E}(\theta_0))^{-1}\right)$ follows immediately from (15).

Proof of (18). According to (RC3), for sufficiently small $\delta > 0$, there is a number M such that

$$\sup_{(\theta,x,e)\in N_\delta(\theta_0)\times\mathcal{X}\times\mathcal{E}} |\ell'''(\theta|x,e)| \le M.$$

Thus for sufficiently large n, it holds that $\hat{\theta}_n \in N_\delta(\theta_0)$, P_{θ_0}-a.s., and that

$$\left|\frac{\ell_n'''(\theta^*)}{n}\right| = \frac{1}{n}\left|\sum_{i=1}^n \ell'''(\theta^*|X_i,E_i)\right| \le \frac{1}{n}\sum_{i=1}^n |\ell'''(\theta^*|X_i,E_i)| \le M, \qquad P_{\theta_0}\text{-a.s.}$$

□

Proof of (17). According to (RC3), there is a number M such that

$$\sup_{(x,e)\in\mathcal{X}\times\mathcal{E}} |\ell''(\theta_0|x,e)| \le M.$$

Let Y be a random variable (r.v.) which takes the constant value M. Then $E_{\theta_0}[|Y|\log^+|Y|] = M\log^+ M < \infty$, and

$$P_{\theta_0}(|\ell''(\theta_0|X_n,E_n)| > y) \le P_{\theta_0}(|Y| > y)$$

for each $y \geq 0$ and $n \geq 1$. As a consequence, we have from Theorem 14 in Appendix B that

$$\frac{1}{n}\sum_{i=1}^{n}\{\ell''(\theta_0|X_i,E_i) - E_{\theta_0}\left[\ell''(\theta_0|X_i,E_i)|\,\mathcal{F}_{i-1}\right]\} \longrightarrow 0, \qquad P_{\theta_0}\text{-a.s.} \tag{19}$$

For an interior point $\theta \in \Theta$, let

$$I(\theta|E_i) := E_\theta\left[\ell'(\theta|X_i,E_i)^2\big|\,\mathcal{F}_{i-1}\right] = -E_\theta\left[\ell''(\theta|X_i,E_i)|\,\mathcal{F}_{i-1}\right], \quad P_{\theta_0}\text{-a.s.} \tag{20}$$

be the *conditional Fisher information* [14, p. 157] under the experiment E_i. Since $E_i = \hat{E}(\hat{\theta}_{i-1})$ for $i \geq 2$ and $\hat{\theta}_{i-1} \to \theta_0$, P_{θ_0}-a.s., it follows from (RC1) and (RC3) that

$$I(\theta_0|E_i) \longrightarrow J(\theta_0|\hat{E}(\theta_0)), \qquad P_{\theta_0}\text{-a.s.} \tag{21}$$

Combining (19)-(21), and using Cesàro's lemma, we have

$$\frac{1}{n}\sum_{i=1}^{n}\ell''(\theta_0|X_i,E_i) \longrightarrow -J(\theta_0|\hat{E}(\theta_0)), \qquad P_{\theta_0}\text{-a.s.}$$

□

Proof of (16). For an interior point $\theta \in \Theta$, $n \geq 1$, and $1 \leq i \leq n$, let

$$S_{ni}(\theta) := \frac{1}{\sqrt{n}}\,\ell_i'(\theta) = \frac{1}{\sqrt{n}}\sum_{j=1}^{i}\ell'(\theta|X_j,E_j),$$

and let $\mathcal{F}_{ni} := \sigma(X_1,\ldots,X_i)$. Then $\{S_{ni}(\theta)\}_{1\leq i\leq n}$ is a martingale array (see Appendix C) with the martingale difference

$$Y_{ni}(\theta) := S_{ni}(\theta) - S_{n,i-1}(\theta) = \frac{1}{\sqrt{n}}\,\ell'(\theta|X_i,E_i).$$

In fact,

$$E_\theta[S_{ni}(\theta)\,|\,\mathcal{F}_{n,i-1}] = S_{n,i-1}(\theta) + \frac{1}{\sqrt{n}}\,E_\theta[\ell'(\theta|X_i,E_i)\,|\,\mathcal{F}_{n,i-1}] = S_{n,i-1}(\theta), \quad P_{\theta_0}\text{-a.s.}$$

and the square integrability is verified as follows.

$$\begin{aligned} E_\theta\left[S_{ni}(\theta)^2\right] &= \frac{1}{n}\sum_{j=1}^{i}E_\theta\left[\ell'(\theta|X_j,E_j)^2\right] + \frac{2}{n}\sum_{j<k}^{i}E_\theta\left[\ell'(\theta|X_j,E_j)\,\ell'(\theta|X_k,E_k)\right] \\ &= \frac{1}{n}\sum_{j=1}^{i}E_\theta\left[\ell'(\theta|X_j,E_j)^2\right] \\ &\leq M(\theta)^2. \end{aligned}$$

Here

$$M(\theta) := \sup_{(x,e)\in\mathcal{X}\times\mathcal{E}}|\ell'(\theta|x,e)|,$$

and the following relation is used in the second equality: for $j < k$

$$E_\theta\left[\ell'(\theta|X_j,E_j)\,\ell'(\theta|X_k,E_k)\,|\,\mathcal{F}_{n,k-1}\right] = \ell'(\theta|X_j,E_j)\;E_\theta[\ell'(\theta|X_k,E_k)\,|\,\mathcal{F}_{n,k-1}] = 0, \quad P_{\theta_0}\text{-a.s.}$$

Let us verify the conditions that ensure the martingale central limit theorem (Theorem 15 in Appendix C). Clearly the σ-algebras $\{\mathcal{F}_{ni}\}$ are nested. The conditional variance is

$$V_{ni}^2(\theta) := \sum_{j=1}^{i} E_\theta\left[\,Y_{nj}(\theta)^2\,\middle|\,\mathcal{F}_{n,j-1}\right] = \frac{1}{n}\sum_{j=1}^{i} E_\theta\left[\,\ell'(\theta|X_j,E_j)^2\,\middle|\,\mathcal{F}_{n,j-1}\right] = \frac{1}{n}\sum_{j=1}^{i} I(\theta|E_j).$$

Thus, by (21),

$$V_{nn}^2(\theta_0) = \frac{1}{n}\sum_{j=1}^{n} I(\theta_0|E_j) \longrightarrow J(\theta_0|\hat{E}(\theta_0)), \quad P_{\theta_0}\text{-a.s.}$$

On the other hand, for any $\varepsilon > 0$, we have from the conditional Markov inequality that

$$\begin{aligned} E_{\theta_0}\left[\,Y_{ni}(\theta_0)^2\,I(|Y_{ni}(\theta_0)| > \varepsilon)\big|\,\mathcal{F}_{n,i-1}\right] &\le \frac{M(\theta_0)^2}{n}\,P_{\theta_0}\left[\,|Y_{ni}(\theta_0)| > \varepsilon\,|\,\mathcal{F}_{n,i-1}\right] \\ &\le \frac{M(\theta_0)^2}{n}\,\frac{1}{\varepsilon^2}E_{\theta_0}\left[\,Y_{ni}(\theta_0)^2\,\middle|\,\mathcal{F}_{n,i-1}\right] \\ &\le \frac{M(\theta_0)^4}{n^2\varepsilon^2}, \quad P_{\theta_0}\text{-a.s.} \end{aligned}$$

This assures the conditional Lindeberg condition:

$$\sum_{i=1}^{n} E_{\theta_0}\left[\,Y_{ni}^2(\theta_0)\,I(|Y_{ni}(\theta_0)| > \varepsilon)\big|\,\mathcal{F}_{n,i-1}\right] \le \frac{M(\theta_0)^4}{n\varepsilon^2} \longrightarrow 0, \quad P_{\theta_0}\text{-a.s.}$$

As a consequence, the desired result

$$S_{nn}(\theta_0) \longrightarrow N\left(0, J(\theta_0|\hat{E}(\theta_0))\right), \qquad \text{in distribution}$$

follows immediately from the martingale central limit theorem. □

6 Adaptive quantum estimation problems

Having established strong consistency and asymptotic efficiency of MLE, let us investigate the operational meaning of the Cramér-Rao type approach in the context of adaptive quantum estimation. Suppose that, by prior investigation of the quantum statistical model $\mathcal{S}$, we have the list of optimal LUEs $\left(M(\,\cdot\,;\theta), \check{\theta}(\,\cdot\,;\theta)\right)$ for each $\theta \in \Theta$. The basic idea of Nagaoka's adaptive quantum estimation scheme was to use the LUE $\left(M(\,\cdot\,;\hat{\theta}_{n-1}), \check{\theta}(\,\cdot\,;\hat{\theta}_{n-1})\right)$ at stage $n\,(\ge 2)$ which is chosen according to the MLE $\hat{\theta}_{n-1}$ obtained at the previous stage. Because of the strong consistency of MLE (Theorem 8), it is expected that this sequence of LUEs would work well. We assume the following regularity condition:

(RC4) $\big(M(x;\theta),\check{\theta}(x;\theta)\big)$ is continuous in (x,θ).

We first treat the case when the SLD lower bound in (4) is achievable. Note that, by the optimality of LUEs, the estimator $\big(M(\,\cdot\,;\theta_0),\check{\theta}(\,\cdot\,;\theta_0)\big)$ achieves the lower bound $(J_{\theta_0})^{-1}$ in this case. Because of (RC4) and the strong consistency of $\hat{\theta}_n$, the sequence $M(\,\cdot\,;\hat{\theta}_{n-1})$ of measurements converges almost surely to $M(\,\cdot\,;\theta_0)$. Thus the limiting Fisher information matrix $J(\theta_0|\hat{E}(\theta_0))$ in Theorem 9 is almost surely identical to the SLD Fisher information matrix J_{θ_0}, and the asymptotic efficiency of the MLE (Theorem 9) establishes the achievability of the Cramér-Rao lower bound by means of the adaptive estimation scheme. Note that there is no need for using the estimating functions $\check{\theta}(\,\cdot\,;\hat{\theta}_{n-1})$ in this case.

We next treat the case when the SLD lower bound is not necessarily achievable, and study the problem of minimizing the scalar quantity $\mathrm{tr}\, GV_\theta[M,\check{\theta}]$, given a weight $G>0$. Note that the best LUE $\big(M(\,\cdot\,;\theta),\check{\theta}(\,\cdot\,;\theta)\big)$ at each $\theta\in\Theta$ is the minimizer of $\mathrm{tr}\, GV_\theta[M,\check{\theta}]$ in this case. It then follows from the strong consistency (Theorem 8) and continuity (RC1) (RC4) that the sequence $\mathrm{tr}\, GV_{\theta_0}[M(\,\cdot\,;\hat{\theta}_{n-1}),\check{\theta}(\,\cdot\,;\hat{\theta}_{n-1})]$ converges to

$$\mathrm{tr}\, GV_{\theta_0}[M(\,\cdot\,;\theta_0),\check{\theta}(\,\cdot\,;\theta_0)] = \min_{(M,\check{\theta}):\,\text{LUE at }\theta_0} \mathrm{tr}\, GV_{\theta_0}[M,\check{\theta}].$$

This quantity is sometimes referred to as the *most informative Cramér-Rao bound* at θ_0 [4].

Let us investigate the operational meaning of this bound. The sample covariance matrix $S_n = (s_n^{ij})_{1\le i,j\le d}$ for the above sequence of LUEs up to stage n is given by

$$\begin{aligned} s_n^{ij} &:= \frac{1}{n}\sum_{k=1}^{n}\Big(\check{\theta}^i(X_k;\hat{\theta}_{k-1})-\bar{\theta}_n^i\Big)\Big(\check{\theta}^j(X_k;\hat{\theta}_{k-1})-\bar{\theta}_n^j\Big) \\ &= \frac{1}{n}\sum_{k=1}^{n}\Big(\check{\theta}^i(X_k;\hat{\theta}_{k-1})-\theta_0^i\Big)\Big(\check{\theta}^j(X_k;\hat{\theta}_{k-1})-\theta_0^j\Big) - \big(\bar{\theta}_n^i-\theta_0^i\big)\Big(\bar{\theta}_n^j-\theta_0^j\Big), \end{aligned} \tag{22}$$

where $\bar{\theta}_n^i$ is the sample mean:

$$\bar{\theta}_n^i := \frac{1}{n}\sum_{k=1}^{n}\check{\theta}^i(X_k;\hat{\theta}_{k-1}).$$

According to (RC4), there is a number M such that

$$\max_{1\le i\le d}\sup_{(x,\theta)\in\mathcal{X}\times\Theta}|\check{\theta}^i(x;\theta)|\le M.$$

Letting $Y:=M$, we see that $E_{\theta_0}[|Y|\log^+|Y|]=M\log^+ M<\infty$, and

$$P_{\theta_0}(|\check{\theta}^i(X_k;\hat{\theta}_{k-1})|>y)\le P_{\theta_0}(|Y|>y)$$

for all $y\ge 0$, $k\ge 1$, and $1\le i\le d$. It then follows from Theorem 14, continuity (RC1) (RC4), and the local unbiasedness condition (3), that the sample mean $\bar{\theta}_n^i$ converges to

$$\lim_{k\to\infty}E_{\theta_0}\Big[\check{\theta}^i(X_k;\hat{\theta}_{k-1})\,\Big|\,\mathcal{F}_{k-1}\Big]=\theta_0^i,$$

almost surely as $n \to \infty$. As a consequence, the second term of (22) approaches zero almost surely, and, again by Theorem 14, the sample covariance s_n^{ij} converges to

$$\lim_{k\to\infty} E_{\theta_0}\left[\left(\check{\theta}^i(X_k;\hat{\theta}_{k-1})-\theta_0^i\right)\left(\check{\theta}^j(X_k;\hat{\theta}_{k-1})-\theta_0^j\right)\,\middle|\,\mathcal{F}_{k-1}\right] = v_{\theta_0}^{ij}\left[M(\,\cdot\,;\theta_0),\check{\theta}(\,\cdot\,;\theta_0)\right]$$

almost surely as $n \to \infty$. We thus conclude that

$$\operatorname{tr} GS_n \longrightarrow \min_{(M,\check{\theta})\colon \text{LUE at } \theta_0} \operatorname{tr} GV_{\theta_0}[M,\check{\theta}]$$

almost surely as $n \to \infty$.

In summary, when the true value of the parameter is θ_0, the weighted sum $\operatorname{tr} GS_n$ of sample covariances s_n^{ij} for the sequence $\left(M(\,\cdot\,;\hat{\theta}_{n-1}),\check{\theta}(\,\cdot\,;\hat{\theta}_{n-1})\right)$ of LUEs converges almost surely to the most informative Cramér-Rao bound at θ_0.

7 Concluding remarks

A crucial observation in the proof of strong consistency is the continuity of the functions $g(e)$ and $g^\varepsilon(e)$, and the convergence $g^\varepsilon(e) \downarrow g(e)$, which hold trivially when $\mathcal{X}$ is finite. For a general data space $\mathcal{X}$ equipped with a σ-finite measure μ, these properties might fail. However, with additional regularity conditions, the same argument would still work. A possible additional requirement will be as follows:

(RC5) For each $\theta \in \Theta$, $e \in \mathcal{E}$, and sufficiently small $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\int \sup_{(\theta',e')\in N_\varepsilon(\theta)\times N_\delta(e)} \sqrt{f(x;\theta',e')\,f(x;\theta_0,e')}\;\mu(dx) < \infty.$$

This condition allows us to use the dominated convergence theorem in proving the continuity of $g(e)$ and $g^\varepsilon(e)$ as well as the convergence $g^\varepsilon(e) \downarrow g(e)$. For example, for any $e_0 \in \mathcal{E}$ and sufficiently small $\varepsilon > 0$,

$$\begin{aligned}
\lim_{e\to e_0} g^\varepsilon(e) &= \lim_{e\to e_0} \int \sup_{\theta'\in N_\varepsilon(\theta)} \sqrt{f(x;\theta',e)\,f(x;\theta_0,e)}\;\mu(dx) \\
&= \int \lim_{e\to e_0} \sup_{\theta'\in N_\varepsilon(\theta)} \sqrt{f(x;\theta',e)\,f(x;\theta_0,e)}\;\mu(dx) \\
&= \int \sup_{\theta'\in N_\varepsilon(\theta)} \sqrt{f(x;\theta',e_0)\,f(x;\theta_0,e_0)}\;\mu(dx) = g^\varepsilon(e_0).
\end{aligned}$$

Note that condition (RC5) is similar in spirit to condition (S2) in Datta [7], and is slightly stronger than condition (2.4) in McCormick et al. [6]. Also note that when $\mathcal{X}$ is compact and $\mu(\mathcal{X}) < \infty$, (RC5) is always satisfied. In order to handle asymptotic efficiency for a general data space $\mathcal{X}$, further regularity conditions are also required.

Returning to adaptive quantum estimation problems, the finiteness of the sample space $\mathcal{X}$ was ensured by the finite dimensionality of $\mathcal{H}$ (Theorem 5). It is nevertheless clear from the above remarks that the asymptotic properties of adaptive quantum estimation schemes obtained in Section 6 can be extended, with some additional regularity conditions, to quantum statistical models on infinite dimensional Hilbert spaces. Such an extension will provide a rigorous foundation of adaptive estimation schemes in quantum optics [15, 16, 17].

Acknowledgment

The author is grateful to H. Nagaoka for stimulating discussions.

Appendices

A Continuity of g^ε

Since $f(x;\,\cdot\,,e)$ is continuous on the compact set Θ, we have

$$\sup_{\theta'\in N_\varepsilon(\theta)} \sqrt{f(x;\theta',e)} = \max_{\theta'\in \overline{N_\varepsilon(\theta)}} \sqrt{f(x;\theta',e)}.$$

Hence to prove the continuity of $g^\varepsilon(e)$ in e for each $\varepsilon>0$, it suffices to establish the following

Lemma 13 *Let $f : A\times B\to\mathbb{R}$ be a continuous function with A a compact set. Then the function $g(y) := \max_{x\in A} f(x,y)$ is continuous in y.*

Proof We show that the opposite claim leads to a contradiction. Suppose that $g(y)$ is discontinuous at $y_0\in B$. Then there exists an $\varepsilon>0$ and a sequence $\{y_n\}_n$ such that $y_n\to y_0$ as $n\to\infty$, and that either (i) $g(y_n)-g(y_0)>\varepsilon$ for all n, or (ii) $g(y_0)-g(y_n)>\varepsilon$ for all n.

In case (i), we have

$$\max_x f(x,y_n) > \max_x f(x,y_0)+\varepsilon \geq f(x',y_0)+\varepsilon$$

for all n and all $x'\in A$. For each n, let $x_n\in A$ be a maximizer of $f(x,y_n)$. Then we have $f(x_n,y_n) > f(x_n,y_0)+\varepsilon$ for all n. Since A is compact, there is a subsequence $\{x_{n(i)}\}_i$ which converges to a point, say x_0, in A as $i\to\infty$. For this subsequence, $f(x_{n(i)},y_{n(i)}) > f(x_{n(i)},y_0)+\varepsilon$. Taking the limit $i\to\infty$, and using the continuity of $f(x,y)$, we have $f(x_0,y_0)\geq f(x_0,y_0)+\varepsilon$. This is a contradiction.

In case (ii), on the other hand, we have

$$\max_x f(x,y_0) > \max_x f(x,y_n)+\varepsilon \geq f(x',y_n)+\varepsilon$$

for all n and all $x'\in A$. Let $x_0\in A$ be a maximizer of $f(x,y_0)$. Then $f(x_0,y_0) > f(x_0,y_n)+\varepsilon$. Taking the limit $n\to\infty$, we have $f(x_0,y_0)\geq f(x_0,y_0)+\varepsilon$. Once again, this is a contradiction. □

B Law of large numbers

Theorem 14 *Let $\{Y_n\}_n$ be a sequence of r.v. adapted to an increasing sequence $\{\mathcal{F}_n\}_n$ of σ-algebras. If there is a r.v. Y satisfying $E|Y|<\infty$, and a constant λ such that*

$$P(|Y_n|>y) \leq \lambda\, P(|Y|>y)$$

for all $n \geq 1$ and $y \geq 0$, then

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i - E[Y_i \,|\, \mathcal{F}_{i-1}]\} \longrightarrow 0, \quad \textit{in probability.}$$

If $E[\,|Y|\log^+|Y|\,] < \infty$, then the convergence in probability can be strengthened to a.s. convergence.

Here, $\log^+ x := \max\{0, \log x\}$. For the proof, see [14, p. 36, Theorem 2.19]. For example, if $\{Y_n\}_n$ is bounded, in that there is a constant $M > 0$ satisfying $|Y_n| \leq M$ for all n, then the a.s. convergence in Theorem 14 holds. (To see this, just let $Y := M$ and $\lambda = 1$.)

C Martingale central limit theorem

For each $n \geq 1$, let $\{S_{ni}\}_{1\leq i\leq k_n}$ be a zero-mean, square-integrable martingale adapted to $\{\mathcal{F}_{ni}\}_{1\leq i\leq k_n}$, and let $Y_{ni} := S_{ni} - S_{n,i-1}$ denote the martingale difference (with $S_{n0} = 0$). It is assumed that $k_n \uparrow \infty$ as $n \to \infty$. We shall call the double sequence $\{S_{ni}\}_{1\leq i\leq k_n}$ a *martingale array*, and

$$V_{ni}^2 := \sum_{j=1}^{i} E\left[\,Y_{nj}^2 \big|\, \mathcal{F}_{n,j-1}\right]$$

the *conditional variance* of S_{ni}. The following result is a martingale extension of the central limit theorem [14, p. 58, Corollary 3.1].

Theorem 15 *Suppose that the σ-algebras are nested: $\mathcal{F}_{ni} \subset \mathcal{F}_{n+1,i}$, for $n \geq 1$ and $1 \leq i \leq k_n$, and that*

$$\textit{for all } \varepsilon > 0, \quad \sum_{i=1}^{k_n} E\left[\,Y_{ni}^2\, I(|Y_{ni}| > \varepsilon)\big|\, \mathcal{F}_{n,i-1}\right] \longrightarrow 0, \quad \textit{in probability,} \tag{23}$$

and

$$V_{nk_n}^2 \longrightarrow \eta^2, \quad \textit{in probability,}$$

where η^2 is an a.s. finite r.v. Then $S_{nk_n} \to Z$ in distribution stably, where the r.v. Z has characteristic function $E\big[\exp\left(-\frac{1}{2}\eta^2 t^2\right)\big]$.

Note that convergence $Z_n \to Z$ in distribution is called *stable* if for any continuity point z of Z and any event $F \in \mathcal{F}$, the limit

$$\lim_{n\to\infty} P(\{Z_n \leq z\} \cap F) =: Q_z(F)$$

exhists and satisfies $Q_z(F) \to P(F)$ as $z \to \infty$. The condition (23) is called the *conditional Lindeberg condition*.

References

[1] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[2] C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[3] W.G. Cochran, "Experiments for nonlinear functions" (R.A. Fisher memorial lecture), Journal of the American Statistical Association, vol. 68, pp. 771-781 (1973).

[4] H. Nagaoka, "An asymptotically efficient estimator for a one-dimensional parametric model of quantum statistical operators," in Proc. Int. Symp. on Inform. Theory, p. 198 (1988); H. Nagaoka, "On the parameter estimation problem for quantum statistical models," in Proc. 12th Symp. on Inform. Theory and its Appl., pp. 577-582 (1989), reprinted in *Asymptotic Theory of Quantum Statistical Inference*, ed. M. Hayashi (World Scientific, Singapore, 2005), pp. 125-132.

[5] H. Chernoff, "Sequential design of experiments," Annals of Mathematical Statistics, vol. 30, pp. 755-777 (1959).

[6] W.P. McCormick, A.K. Mallik and J.H. Reeves, "Strong consistency of the MLE for sequential design problems," Statistics and Probability Letters, vol. 6, pp. 441-446 (1988).

[7] S. Datta, "Consistency of the mle for a general sequential design problem," Sankyā: The Indian Journal of Statistics, Ser. A, vol. 57, pp. 88-99 (1995).

[8] P. Chaudhuri and R.A. Mykland, "Nonlinear experiments: Optimal design and inference based on likelihood," Journal of the American Statistical Association, vol. 88, pp. 538-545 (1993).

[9] A. Fujiwara and H. Nagaoka, "Operational capacity and pseudoclassicality of a quantum channel," IEEE Trans. Inform. Theory, vol. 44, pp. 1071-1086 (1998).

[10] N. Dunford and J. T. Schwartz, *Linear Operators, Part I: General Theory* (Wiley, NY, 1958).

[11] D. Williams, *Probability with Martingales* (Cambridge University Press, Cambridge, 1991).

[12] The remaining conditional expectation $E_{\theta_0}\left[e^{R^{\varepsilon}(\theta_0,\theta,E_n)/2}\,\middle|\,\mathcal{F}_{n-1}\right]$ is the following $\mathcal{F}_{n-1}$-measurable random variable:

$$\sum_x \sup_{\theta'\in N_\varepsilon(\theta)} \sqrt{f(x;\theta',E_n(X_1,\ldots,X_{n-1}))}\,\sqrt{f(x;\theta_0,E_n(X_1,\ldots,X_{n-1}))}.$$

[13] P. Billingsley, *Convergence of Probability Measures* (Wiley, NY, 1968).

[14] P. Hall and C.C. Heyde, *Martingale Limit Theory and Its Application* (Academic, NY, 1980).

[15] H.M. Wiseman, "Adaptive phase measurements of optical modes: Going beyond the marginal Q distribution," Phys. Rev. Lett., vol. 75, pp. 4587-4590 (1995).

[16] G.M. D'Ariano, M.G.A. Paris, and R. Seno, "Feedback-assisted homodyne detection of phase shifts," Phys. Rev. A, vol. 54, pp. 4495-4504 (1996).

[17] M.A. Armen, J.K. Au, J.K. Stockton, A.C. Doherty, and H. Mabuchi, "Adaptive homodyne measurement of optical phase," Phys. Rev. Lett., vol. 89, 133602 (2002).