Error probability analysis in quantum tomography: a tool for evaluating experiments

Takanori Sugiyama,[1,*] Peter S. Turner,[1,†] and Mio Murao[1,2,‡]

[1]*Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan 113-0033.*

[2]*Institute for Nano Quantum Information Electronics, The University of Tokyo, 4-6-1 Komaba, Meguro-ku,Tokyo , Japan 153-8505.*

(Dated: September 14, 2010)

We expand the scope of the statistical notion of error probability, *i.e.*, how often large deviations are observed in an experiment, in order to make it directly applicable to quantum tomography. We verify that the error probability can decrease at most exponentially in the number of trials, derive the explicit rate that bounds this decrease, and show that a maximum likelihood estimator achieves this bound. We also show that the statistical notion of identifiability coincides with the tomographic notion of informational completeness. Our result implies that two quantum tomographic apparatuses that have the same risk function, (e.g. variance), can have different error probability, and thus by combining these two approaches we can evaluate, in a reconstruction independent way, the performance of such experiments more discerningly.

PACS numbers: 03.65.Wj, 03.67.-a, 02.50.Tt, 06.20.Dk

I. INTRODUCTION

Many applications that make use of "quantumness" in order to outperform their classical counterparts have recently been proposed, especially in the field of quantum information. One of the main reasons for this increase has been the dramatic development of experimental technologies, and many of the proposals have already given rise to experimentally realizable applications [1]. To confirm whether or not an apparatus constructed for an application works well, we need to compare its performance to a theoretical model. The standard method used for a thorough such comparison is called quantum tomography [2]. This paper is concerned mainly with the question of how to evaluate measurement apparatuses used in quantum tomography.

The theory of quantum tomography consists of experimental design methods and reconstruction schemes. Known parts of the experimental apparatus in a quantum tomographic experiment, (or at least those parts assumed to be known), are together called the *tester*. Experimental design methods are concerned with how good (or bad) the tester is for estimating the mathematical representation of the tomographic object (*e.g.* a quantum state, or a process). Usually the goodness of the tester is evaluated by the error of the estimation result from its experimental data set. In real experiments, we cannot perform an infinite number trials – we need to estimate the true tomographic object from a finite number. This estimation procedure is called an estimator in statistical estimation theory and a reconstruction scheme in quantum tomography. The error of the estimation result depends upon the reconstruction scheme, and when evaluating a tester's performance, we usually focus upon the error in the case where the best reconstruction scheme is used.

Evaluating estimation errors on the reconstructed object is a problem of statistical estimation theory. There are two main approaches; one is to use a *risk function* and the other is to use *error probability*. We measure the difference between the true object and the estimate by a *loss function*. A risk function is the average value of the loss function. As the number of independent, identically distributed (iid) trials increases, it is known that the error, given by the risk function, of any unbiased estimator can decrease by at most the Cramér-Rao bound, and a maximum likelihood estimator achieves the bound asymptotically [3]. The application of the Cramér-Rao inequality to quantum tomography is studied in [4–7]. On the other hand, an error probability is the probability that large deviations of the loss function are observed. It has been shown that the error probability can decrease at most exponentially [8], and under some conditions, the bound is achieved (asymptotically) by a maximum likelihood estimator [9]. However, the explicit form of this bound has not been shown except for the case where the estimation setting can be reduced to one parameter estimation or the loss function is a Euclidean norm [10–13]. In general, the estimated object has multiple parameters, and the choice of the loss function depends upon the purpose of the experiment. A mean squared error can be unsuitable for some situations, especially those arising in quantum tomography. In order to be more useful in practice, the explicit form of the bound is needed in more generality.

In this paper, by using Sanov's theorem [14, 15] from large deviation theory, we derive the error probability inequality bounding general loss functions on a finite multiparameter space. We prove that a maximum likelihood estimator achieves the equality under some conditions – that are satisfied in quantum tomography – and give the explicit form of the lower bound. Our result indi-

* sugiyama@eve.phys.s.u-tokyo.ac.jp
† turner@phys.s.u-tokyo.ac.jp
‡ murao@phys.s.u-tokyo.ac.jp

cates that two testers with same value of their risk functions can be different from an error probability viewpoint, which allows for more discerning comparisons of testers in quantum tomography. We also show that the required conditions for our inequality hold not only for tomography of quantum states, but also for that of quantum instruments [16], which includes process and measurement tomography as specific cases.

In section II, we overview the theory of quantum tomography using state tomography as an example. In section III, we review classical statistical estimation theory, introducing the necessary aspects of error probability theory. In section IV, we give the main theorem and some analysis – the proof of the theorem is given in the Appendix. In section V, we discuss some open problems, and conclude with a summary in section VI.

II. OVERVIEW OF QUANTUM TOMOGRAPHY

Quantum tomography is classified by the tomographic object to be reconstructed: state tomography [17–19] treats density operators, which describe states of quantum systems; process tomography [20–24] treats linear, trace-preserving, and completely positive maps, which describe deterministic state transitions; POVM tomography [25, 26] treats positive operator-valued measures (POVMs), which are sets of positive-semidefinite operators describing the probability distributions obtained by measurements; instrument tomography [16] treats quantum instruments, which are sets of linear, trace-decreasing, completely positive maps describing both probability distributions and state transitions caused by measurements. Here we briefly overview the theory of quantum state tomography, for simplicity.

The purpose of quantum state tomography is to identify the density operator characterizing the state of the quantum system of interest. Let $\mathcal{H}$ and $\mathcal{S}(\mathcal{H})$ denote the Hilbert space corresponding to the system and the set of all density operators on $\mathcal{H}$, respectively. We assume that the dimension $d = \dim \mathcal{H}$ is finite. A density operator $\hat{\rho}$ on $\mathcal{H}$ can be linearly and bijectively parametrized by $d^2 - 1$ independent real variables [27, 28], *i.e.* $\hat{\rho} = \hat{\rho}(\boldsymbol{s})$, where $\boldsymbol{s}$ is in $\mathbf{R}^{d^2-1}$. Let us define the set of all parameters $S := \{\boldsymbol{s} \in \mathbf{R}^{d^2-1} | \ \hat{\rho}(\boldsymbol{s}) \in \mathcal{S}(\mathcal{H})\}$. Identifying the true density operator $\hat{\rho} \in \mathcal{S}(\mathcal{H})$ is equivalent to identifying the true parameter $\boldsymbol{s} \in S$. Let $\boldsymbol{\Pi} = \{\hat{\Pi}_x\}_{x\in\Omega}$ denote the POVM characterizing the tester used in the tomographic experiment [29], where $\Omega := \{1, \ldots, M\}$. When the true density operator is $\hat{\rho}(\boldsymbol{s})$, the probability distribution $p_{\boldsymbol{s}}$ describing the tomographic experiment is given by

$$p_{\boldsymbol{s}}(x) = \mathrm{Tr}[\hat{\rho}(\boldsymbol{s})\hat{\Pi}_x], \ x \in \Omega, \tag{1}$$

where Tr denotes the trace operation with respect to $\mathcal{H}$. (Note that in subsection IV C a different trace operation, tr, is introduced.) Suppose that we perform N measurement trials and obtain a data set $\boldsymbol{x}^N = (x_1, \ldots, x_N)$, where $x_i \in \Omega$ is the outcome observed in the i-th trial. Let N_x denote the number of times that outcome x occurs in $\boldsymbol{x}^N$, then $f_N(x) := N_x/N$ is the relative frequency of x for the data set $\boldsymbol{x}^N$. In the limit of $N = \infty$, the relative frequency is equal to the true probability $p_{\boldsymbol{s}}(x)$. A tester is called *informationally complete* if $\mathrm{Tr}[\hat{\rho}(\boldsymbol{s})\hat{\Pi}_x)] = \mathrm{Tr}[\hat{\rho}\hat{\Pi}_x]$ has a unique solution $\hat{\rho}$ for arbitrary $\hat{\rho}(\boldsymbol{s}) \in \mathcal{S}(\mathcal{H})$ [30]. This condition is equivalent to that of the tester POVM $\boldsymbol{\Pi}$ being a basis for the set of all Hermitian matrices on $\mathcal{H}$. For finite N, the relative frequency and true probability are generally not the same, *i.e.*, there is unavoidable statistical error, and we need to choose an estimation procedure that takes the experimental result $\boldsymbol{x}^N$ to a density operator, *i.e.*, a reconstruction scheme.

Reconstruction schemes are concerned with how best to derive the mathematical representation of the tomographic object from the obtained experimental data, and are called *estimators* in statistical estimation theory, where the analysis of the estimation precision (or estimation error) is very important. In actual experiments, there are two sources of imprecision: statistical errors and systematic errors. As mentioned above, statistical error is caused by the finiteness of the total number of measurement trials, and is unavoidable in principle. Systematic error is caused by our lack of knowledge about the tester, that is, the difference between the true tester and what we believe to be the true tester. Usually, the effect of the systematic error is approximated by introducing a model, and is assumed to be known. Therefore, the analysis of the estimation error is usually reduced to that of the statistical error. To date at least five reconstruction schemes have been proposed, namely linear [16, 17, 20, 21, 25], maximum likelihood [18, 23, 26], Bayesian [31–33], maximum entropy [34], and norm minimization [35]. The effect of statistical errors on the reconstructed object depends upon the scheme used, hence the main problem is how to quantify the effect of the statistical error on the reconstructed object, and how to do so as rigorously as possible.

It is natural to consider a linear reconstruction scheme, which demands that we find a $d \times d$ matrix $\hat{\rho}^{\mathrm{l}}$ satisfying

$$\mathrm{Tr}[\hat{\rho}^{\mathrm{l}}\hat{\Pi}_x] = f_N(x), \ x \in \Omega. \tag{2}$$

However, Eq.(2) does not always have a solution, and even when it does, although the solution is Hermitian and normalized, it is not guaranteed that $\hat{\rho}^{\mathrm{l}}$ is positive semidefinite. A maximum likelihood reconstruction scheme addresses these problems. The estimated matrix $\hat{\rho}^{\mathrm{ml}}$ is defined as

$$\hat{\rho}^{\mathrm{ml}} := \mathrm{argmax}_{\hat{\rho}\in\mathcal{S}(\mathcal{H})} \prod_{i=1}^{N} \mathrm{Tr}[\hat{\rho}\hat{\Pi}_{x_i}]. \tag{3}$$

It can be shown that when $\hat{\rho}^{\mathrm{l}} \in \mathcal{S}(\mathcal{H})$, $\hat{\rho}^{\mathrm{l}} = \hat{\rho}^{\mathrm{ml}}$ holds. We will concern ourselves with maximum likelihood reconstructions here, as we will see that they are optimal in the sense that they can saturate the bounds we are considering.

III. OVERVIEW OF CLASSICAL STATISTICAL ESTIMATION

In this section, we introduce the notation and terminology of classical statistical estimation theory that we use to arrive at our main results. We also review the necessary aspects of error probability. For the reader familiar with quantum estimation theory, we justify our use of classical estimation theory in subsection V D.

Let $(\Omega, \mathscr{B}, P)$ be a probability space, denoting a sample space, a Borel algebra of subsets of the sample space, and a measure that assigns probabilities to those subsets, respectively (the Borel structure simply assures that combinations of subsets of events get assigned probabilities in a sensible fashion, *e.g.* $P(Y \cup Z) = P(Y) + P(Z)$ for disjoint Y and Z). Define the N-fold direct product $\Omega^N := \Omega \times \cdots \times \Omega$ as the space of sequences of events. Let $\boldsymbol{x}^N = \{x_1, \cdots, x_N\}, x_i \in \Omega$, be a sequence of iid observations of the sample space Ω. We assume that the sample space is finite, (see subsection V B for a discussion of infinite spaces). Suppose that the probability space admits a statistical model $\mathcal{P}_\Theta = \{P_\theta; \theta \in \Theta\}$ that assigns a valid probability measure to each parameter θ in Θ which is an subset of the k-dimensional Euclidean space $\mathbf{R}^k$, the closure $\bar{\Theta}$ is compact, and the interior Θ^o is open. The quantum state parameter space S from the last section is an example of such a Θ, where the statistical model is given by Eq.(1). We assume that each measure P_θ has a probability distribution $\{p_\theta(x)\}_{x\in\Omega}$ satisfying $P_\theta(Y) = \sum_{x \in Y} p_\theta(x)$ where $Y \in \mathscr{B}$. A probability measure P_θ and the probability distribution p_θ have a one-to-one correspondence for any $\theta \in \Theta$, and we do not distinguish between $\mathcal{P}_\Theta$ and $\{p_\theta; \theta \in \Theta\}$. Let $\mathcal{P}(\Omega)$ denote the set of all probability distributions with the sample space Ω, then $\mathcal{P}_\Theta \subseteq \mathcal{P}(\Omega)$. Let $P_\theta^{(N)}$ denote the N-fold product probability measure $P_\theta \times \cdots \times P_\theta$.

Let g denote a map from the parameter space Θ to a metric space Γ. An estimator of $g(\theta)$ is a set of maps $\varphi = \{\varphi_1, \varphi_2, \ldots\}$, (one for each number of trials N), from observation results $\boldsymbol{x}^N$ to Γ. Each $\varphi_N(\boldsymbol{x}^N)$ is called the estimate of $\boldsymbol{x}^N$. A maximum likelihood estimator $\theta^{\mathrm{ml}} = \{\theta_1^{\mathrm{ml}}, \theta_2^{\mathrm{ml}}, \ldots\}$ of θ is defined as

$$\theta_N^{\mathrm{ml}}(\boldsymbol{x}^N) = \mathrm{argsup}_{\theta\in\Theta} P_\theta^{(N)}(\{\boldsymbol{x}^N\}). \tag{4}$$

A map D from $\Gamma \times \Gamma$ to $\mathbf{R}$ is called a loss function on Γ when D satisfies the following two conditions: (i) $\forall a, b \in \Gamma$, $D(a,b) \geq 0$, (ii) $\forall a \in \Gamma$, $D(a,a) = 0$. We introduce three additional conditions: (iii) $\forall a, b \in \Gamma$, $D(a,b) = D(b,a)$, (iv) $\forall a,b,c \in \Gamma$, $D(a,b) \leq D(a,c) + D(c,b)$, (v) $\forall a, b \in \Gamma$, $D(a,b) = 0 \Rightarrow a = b$. A loss function satisfying conditions (iii) and (iv) is called a semi-distance, and a semi-distance satisfying condition (v) is called a distance. For example, let us define a function g from Θ to $\mathbf{R}$ as $g(\theta) = \|\theta\|, \theta \in \Theta$, where $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^k$. Then $|g(\theta) - g(\theta')|$ is a semi-distance on Θ $(\theta, \theta' \in \Theta)$ and $|a - b|$ is a distance on $\mathbf{R}$ $(a, b \in \mathbf{R})$. In general, a loss function is not necessarily a distance. A loss function satisfying condition (v) is called a pseudo-distance [36]. The Kullback-Leibler divergence introduced below is an example of pseudo-distance that is not also a distance. If a loss function D on $\mathbf{R}^k$ is sufficiently smooth and it can be approximated by the Hesse matrix H_a up to second order, then H_a is positive semidefinite for all $a \in \mathbf{R}^k$ from condition (i), and if the loss function D is a pseudo-distance, then H_a is positive definite for all $a \in \mathbf{R}^k$.

There are at least two methods to evaluate an estimation error by using loss function. One is a method using risk functions. An N-trial risk function $\bar{D}^{(N)}$ is defined as the expectation value of the loss function between an estimate and the true object, given by θ;

$$\bar{D}^{(N)} := E_\theta^{(N)}[D(\varphi_N(\boldsymbol{x}^N), g(\theta))], \tag{5}$$

where $E_\theta^{(N)}[f(\boldsymbol{x}^N)] = \sum_{\boldsymbol{x}^N \in \Omega^N} p_\theta(\boldsymbol{x}^N) f(\boldsymbol{x}^N)$ is the expectation value of a function f on Ω^N. When $\Gamma = \mathbf{R}^l$, for any unbiased estimator (φ satisfying $E_\theta^{(N)}[\varphi_N(\boldsymbol{x}^N)] = g(\theta)$ for any N and $\theta \in \Theta$),

$$\begin{aligned} E_\theta^{(N)}&[(\varphi_N(\boldsymbol{x}^N) - g(\theta))(\varphi_N(\boldsymbol{x}^N) - g(\theta))^T] \\ &\geq \frac{1}{N} \frac{\partial g}{\partial \theta}^T F_\theta^{-1} \frac{\partial g}{\partial \theta} \end{aligned} \tag{6}$$

holds under some regularity conditions, where $\left(\frac{\partial g}{\partial \theta}\right)_{\alpha\beta} := \frac{\partial g_\beta}{\partial \theta_\alpha}$ $(\alpha = 1, \ldots, k; \beta = 1, \ldots, l)$ is the Jacobian and F_θ^{-1} is the generalized inverse of the Fisher matrix $F_\theta := \sum_{x\in\Omega} p_\theta(x)[\nabla_\theta \log p_\theta(x)][\nabla_\theta \log p_\theta(x)]^T$. Asymptotically a maximum likelihood estimator achieves the equality under some conditions [3].

The other is a method using error probabilities. We call

$$\begin{aligned} P_\epsilon^{(N)}(\theta) :&= P_\theta^{(N)}(D(\varphi_N(\boldsymbol{x}^N), g(\theta)) > \epsilon) && (7) \\ &= P_\theta^{(N)}(\{\boldsymbol{x}^N \in \Omega^N; D(\varphi_N(\boldsymbol{x}^N), g(\theta)) > \epsilon\}) && (8) \end{aligned}$$

an error probability with a threshold $\epsilon > 0$. An estimator φ is called (*weakly*) *consistent* in the loss function D if

$$P_\epsilon^{(N)}(\theta) \to 0 \text{ as } N \to \infty \tag{9}$$

holds for any $\epsilon > 0$. The conditions under which a maximum likelihood estimator is consistent includes the *identifiability condition* [37] on a statistical model $\mathcal{P}_\Theta$: for any $\theta \in \Theta^o$ and $\theta' \in \Theta$, if $\theta \neq \theta'$, then there exists at least one outcome $x \in \Omega$ satisfying $p_\theta(x) \neq p_{\theta'}(x)$ [10, 11]. Let us define

$$R_\epsilon(\theta) := \inf_{\theta' \in \Theta} \{K(p_{\theta'} \| p_\theta); D(g(\theta'), g(\theta)) > \epsilon\}, \tag{10}$$

where $K(q\|p) = \sum_{x\in\Omega} q(x) \log \frac{q(x)}{p(x)}$ is called the Kullback-Leibler divergence (also known as the relative entropy). When g is injective and D is a distance, for any weakly consistent estimator in D,

$$\underline{\lim}_{N\to\infty} \frac{1}{N} \log P_\epsilon^{(N)}(\theta) \geq -R_\epsilon(\theta) \tag{11}$$

holds [8]. It is known that in general the lower bound of Eq.(11) is not attainable by any estimate [38]. If we consider the limit $\epsilon \to 0$, under some conditions (including the identifiability condition), a maximum likelihood estimator achieves the equality, that is,

$$\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{1}{\epsilon^u N}\log P_\epsilon^{(N)}(\theta) = -r(\theta), \tag{12}$$

where u is a real number suitable for D and $r(\theta) := \lim_{\epsilon\to 0}\frac{R_\epsilon(\theta)}{\epsilon^u}$. The explicit forms of the rate are known for two specific cases. The first is the case where $\Gamma = \mathbf{R}$ and D is the absolute value. In this case, the order u is 2 and the explicit form of the lower bound is known to be [10, 11]

$$r(\theta) = \frac{1}{2\nabla_\theta g(\theta)\cdot F_\theta^{-1}\nabla_\theta g(\theta)}. \tag{13}$$

The second is the case where $\Gamma = \mathbf{R}^k$, D is the Euclidean distance on $\mathbf{R}^k$, and the order u is again 2; the explicit form is [13]

$$r(\theta) = \frac{1}{2}\inf_{\boldsymbol{a}\in\mathbf{R}^k;\|\boldsymbol{a}\|=1}\boldsymbol{a}\cdot F_\theta\boldsymbol{a}. \tag{14}$$

For more general Γ or D, however, the explicit form of the lower bound is not known. Quantum state tomography corresponds to the case where $\Gamma = \mathbf{R}^{d^2-1}$, and the standard loss function is the square of the fidelity distance $D_F(\hat\rho,\hat\rho') := 1 - \mathrm{Tr}[\sqrt{\sqrt{\hat\rho}\hat\rho'\sqrt{\hat\rho}}]^2$ or the square of the trace distance $D_T(\hat\rho,\hat\rho') := \mathrm{Tr}[|\hat\rho - \hat\rho'|]^2$. In this paper, we extend the above results to multiparameter spaces and more general loss functions such as these that are directly applicable to quantum tomography, and give the explicit form of the lower bound. We also give quantum tomography conditions equivalent to the identifiability condition in classical estimation theory.

IV. MAIN RESULT AND ANALYSIS

A. Main theorem

For simplicity we consider quantum state tomography. Suppose that we use a loss function D on $\mathcal{S}(\mathcal{H})$. Let us define a loss function Δ on S as $\Delta(\boldsymbol{s},\boldsymbol{s}') := D(\hat\rho(\boldsymbol{s}),\hat\rho(\boldsymbol{s}'))$ $\forall\boldsymbol{s},\boldsymbol{s}'\in S$. Assume that Δ is sufficiently smooth. Let S^o denote the interior of S. We define a same point Hesse matrix $H_{\boldsymbol{s}}$ for a two variable function f on $S\times S$ as $\nabla_{\boldsymbol{s}}\nabla_{\boldsymbol{s}}f(\boldsymbol{s},\boldsymbol{s}')|_{\boldsymbol{s}=\boldsymbol{s}'} = [\partial_{s^\alpha}\partial_{s^\beta}f(\boldsymbol{s},\boldsymbol{s}')|_{\boldsymbol{s}=\boldsymbol{s}'}]$.

Theorem 1 *Suppose that Δ is a pseudo-distance on S with a non-zero same point Hesse matrix $H_{\boldsymbol{s}}$. If $\boldsymbol{s}\in S^o$, for an arbitrary consistent estimator $\boldsymbol{s}^{\mathrm{est}}$, the following inequality holds:*

$$\underline{\lim}_{\epsilon\to 0}\,\underline{\lim}_{N\to\infty}\frac{1}{\epsilon N}\log P_{\boldsymbol{s}}^{(N)}(\Delta(\boldsymbol{s}_N^{\mathrm{est}},\boldsymbol{s})>\epsilon) \geq -1/\sigma_1(\sqrt{H_{\boldsymbol{s}}}F_{\boldsymbol{s}}^{-1}\sqrt{H_{\boldsymbol{s}}}), \tag{15}$$

where $\sigma_1(A)$ is the maximal eigenvalue of an Hermitian matrix A. Furthermore, when the tester is informationally complete, a maximum likelihood estimator $\boldsymbol{s}^{\mathrm{ml}}$ is consistent and achieves the equality in Eq.(15),i.e.,

$$\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{1}{\epsilon N}\log P_{\boldsymbol{s}}^{(N)}(\Delta(\boldsymbol{s}_N^{\mathrm{ml}},\boldsymbol{s})>\epsilon) = -1/\sigma_1(\sqrt{H_{\boldsymbol{s}}}F_{\boldsymbol{s}}^{-1}\sqrt{H_{\boldsymbol{s}}}) \tag{16}$$

holds.

The detailed proof of Theorem 1 appears in the Appendix – here we give an outline. The proof is divided into six parts. For parts one through five, we do not assume that the probability distributions are quantum mechanical; we only assume that they are sufficiently differentiable and that the parameter space is compact. Only in the sixth part does quantum mechanics arise. In Lemma 1, by using the same logic as the proof of Eq.(11) in [8], we show that Eq.(11) holds for any estimator consistent not only in distances but also in pseudo-distances. Lemma 2 is introduced in order to calculate the infimum in Eq.(10) directly. We use this in Lemma 3, where we obtain the explicit form of the bound on the rate, and obtain Eq.(15). Next we introduce Sanov's theorem, a large deviation theorem that, roughly speaking, gives the rate of the probability of observing a relative frequency that differs from the true probability distribution. Lemma 4 uses the compactness of the parameter space and Sanov's theorem to prove that the error probability of a maximum likelihood estimator decreases exponentially if the identifiability condition is satisfied. Then, the maximum likelihood estimator is consistent and satisfies Eq.(15). In Lemma 5, we calculate the rate of decrease of the maximum likelihood estimator directly by using Sanov's theorem and Lemma 3, and show that the rate coincides with the lower bound in Eq.(15). Hence, we obtain Eq.(16), subject to the identifiability condition. Finally, we prove that in quantum state tomography the identifiability condition is equivalent to the informational completeness of the tester, which we present as Lemma 6. Together these lemmas prove Theorem 1.

Note that in the proof we assume the compactness of the parameter space (in Lemmas 1 to 5) and the linear parametrizability of probability distributions (in Lemma 6). These assumptions hold for any quantum operator. Also, the concept of identifiability applies to the tomographic completeness of states equally well as it does to the informational completeness of measurements, which can be shown using the same logic as that of Lemma 6. Thus theorem 1 holds for all types of quantum tomography. The dimension of the parameter space k depends upon the type of quantum tomography: $k = d^2 - 1$ and $d^4 - d^2$ for state and process tomography, respectively. For POVM and instrument tomography, $k = (M-1)d^2$ and $Md^4 - d^2$ respectively, where M denotes the number of measurement outcomes.

B. Meaning of the lower bound

Theorem 1 indicates that in quantum tomography, if we have a sufficiently large data set, the error probability of any consistent estimator with a small threshold can decrease at most exponentially, and the rate is bounded by an estimator-independent function $1/\sigma_1(\sqrt{H_{\boldsymbol{s}}}F_{\boldsymbol{s}}^{-1}\sqrt{H_{\boldsymbol{s}}})$. Also, the bound is achievable by a maximum likelihood estimator. Therefore, from the error probability viewpoint, if we can perform a large number of measurement trials, a maximum likelihood reconstruction scheme is optimal. We can evaluate the performance of a given tester by the size of the maximal eigenvalue of the matrix

$$G_{\boldsymbol{s}} := \sqrt{H_{\boldsymbol{s}}}F_{\boldsymbol{s}}^{-1}\sqrt{H_{\boldsymbol{s}}}. \tag{17}$$

Testers with smaller maximal eigenvalues are better. The inverse Fisher matrix $F_{\boldsymbol{s}}^{-1}$ alone characterizes the parameter-identifiability of the tester with respect to the Euclidean distance because the Hesse matix of the square of the Euclidean distance $\Delta_E(\boldsymbol{s},\boldsymbol{s}') := \|\boldsymbol{s}-\boldsymbol{s}'\|^2$ is $2I$, and we obtain

$$\frac{1}{\sigma_1(G_{\boldsymbol{s}})} = \frac{1}{2\sigma_1(F_{\boldsymbol{s}}^{-1})} \tag{18}$$
$$= \frac{1}{2}\sigma_k(F_{\boldsymbol{s}}) \tag{19}$$
$$= \frac{1}{2}\inf_{\boldsymbol{a}\in\mathbf{R}^k;\|\boldsymbol{a}\|=1} \boldsymbol{a}\cdot F_{\boldsymbol{s}}\boldsymbol{a}, \tag{20}$$

where $\sigma_k(A)$ is the minimal eigenvalue of an Hermitian matrix A. This result coincides with the known result of Eq.(14). The loss function Δ characterizes the purpose of the estimation (what we want to know), and the same point Hesse matrix $H_{\boldsymbol{s}}$ modifies the inverse Fisher matrix from the Euclidean distance to the loss function Δ on S. Therefore the matrix $G_{\boldsymbol{s}}$ characterizes the parameter-identifiability of the tester with a modification according to our estimation purpose.

C. Relation to risk functions

If we assume sufficient smoothness of a loss function Δ on S and informational completeness on the tester, a generalized Cramér-Rao inequality can be derived, *i.e.*, for any unbiased estimator, the following inequality holds:

$$\bar{\Delta}^{(N)} \geq \frac{\mathrm{tr}[H_{\boldsymbol{s}}F_{\boldsymbol{s}}^{-1}]}{2N} + o(\frac{1}{N}), \tag{21}$$

where tr denotes the trace operation with respect to the parameter space [4]. Eq.(21) indicates that for sufficiently large N, the risk function can decrease at most inverse-proportionally to N, and the rate is characterized by $\mathrm{tr}[H_{\boldsymbol{s}}F_{\boldsymbol{s}}^{-1}]$. We can rewrite this as

$$\mathrm{tr}[H_{\boldsymbol{s}}F_{\boldsymbol{s}}^{-1}] = \mathrm{tr}[\sqrt{H_{\boldsymbol{s}}}F_{\boldsymbol{s}}^{-1}\sqrt{H_{\boldsymbol{s}}}] \tag{22}$$
$$= \textstyle\sum_{\alpha=1}^{k}\sigma_\alpha(G_{\boldsymbol{s}}), \tag{23}$$

where $\sigma_\alpha(A)$ is the α-th eigenvalue of a symmetric $k\times k$ matrix A arranged in decreasing order. Therefore, the rates of decrease of error probability and risk function are both characterized by, respectively, the maximal eigenvalue and the sum of all the eigenvalues of a common matrix $G_{\boldsymbol{s}}$. The rates' properties depend upon the choice of the loss function. For example, when we choose the Kullback-Leibler divergence, *i.e.*, $\Delta(\boldsymbol{s},\boldsymbol{s}') = K(p_{\boldsymbol{s}}\|p_{\boldsymbol{s}'})$, we obtain $H_{\boldsymbol{s}} = F_{\boldsymbol{s}}$ and therefore $\sigma_1(G_{\boldsymbol{s}}) = 1$ and $\sum_{a=1}^{k}\sigma_a(G_{\boldsymbol{s}}) = k$. In this case the rates of decrease do not depend upon the true parameter or the tester, but in general the rates depend upon both.

The Cramér-Rao inequality holds only for unbiased estimators, and the bound can be broken by biased estimators. On the other hand, the error probability inequality holds for any consistent estimator. A maximum likelihood estimator is consistent under some conditions (including the identifiability condition), and is not unbiased in general but achieves the lower bound of Eq.(21) asymptotically. When we use a maximum likelihood reconstruction scheme in quantum tomography, the performance of the tester is evaluated by $\sum_{\alpha=1}^{k}\sigma_\alpha(G_{\boldsymbol{s}})$ from the risk function viewpoint. When we have two testers with the same value of $\sum_{\alpha=1}^{k}\sigma_\alpha(G_{\boldsymbol{s}})$ at a $\boldsymbol{s}\in S$, their performances are equivalent in the risk function sense, but if the maximal eigenvalues $\sigma_1(G_{\boldsymbol{s}})$ are different, their error probability performances are different. Thus we can evaluate the performance of testers more discerningly by considering error probabilities than we can by considering only risk functions, using the same set of eigenvalues – that of the matrix $G_{\boldsymbol{s}}$.

D. Extension to more general quantum estimation problem

A loss function used in quantum state tomography is usually a distance on S (or $\mathcal{S}(\mathcal{H})$). This is because the purpose of quantum state tomography is to identify the true parameter (or true density operator). There are, however, cases where exact identifiability is not required, for example, estimations of the average value of an Hermitian operator, the purity of an unknown state, or the value of an entanglement measure. These examples correspond to the case where g is a map from S to $\mathbf{R}$. More generally, we can consider $g: S\to\mathbf{R}^l, l\leq k = d^2-1$. Theorem 1 can be generalized to this case by modifying the identifiability condition (see Appendix) and changing the meaning of the superscript -1 from the inverse matrix to the generalized inverse matrix. Specifically, when $l=1$ and the loss function Δ is the squared absolute value, *i.e.*, $g: S\to\mathbf{R}$ and $\Delta(\boldsymbol{s},\boldsymbol{s}') = |g(\boldsymbol{s})-g(\boldsymbol{s}')|^2$, we can obtain

$$H_{\boldsymbol{s}} = 2(\nabla_{\boldsymbol{s}}g)(\nabla_{\boldsymbol{s}}g)^T, \tag{24}$$
$$\frac{1}{\sigma_1(G_{\boldsymbol{s}})} = \frac{1}{2\nabla_{\boldsymbol{s}}g\cdot F_{\boldsymbol{s}}^{-1}\nabla_{\boldsymbol{s}}g}. \tag{25}$$

This result exactly coincides with the known result of Eq.(13).

When the parameter space is 1-dimensional, the rates of decrease of the two evaluation methods are characterized by the same function, but when the parameter space is more than 2-dimensional, the rates can be characterized differently. The most simple tomographic object, a 1-qubit state, has a 3-dimensional parameter space, therefore even in the simplest type of quantum tomography, if two given testers have the same rate of decrease of a risk function, their rates of decrease of error probability can be different, *i.e.*, the testers can have different quantum tomographic performance.

V. DISCUSSION

A. Evaluating tester performance

Our result shows that when the true parameter is $\boldsymbol{s}$, the rate of decrease of the error probability is characterized by $\sigma_1(G_{\boldsymbol{s}})$. In real experiments of course, we do not know the true parameter, which is the reason we perform tomography in the first place. We explain three approaches to evaluating tester performance below.

The first is to use a parameter which we expect as the true parameter. In many experiments, quantum state tomography is performed not for estimating a state but for proving an experimental realization of a specific quantum state, for example, a maximally entangled state. By using the parameter corresponding to the quantum state we want to realize, we can evaluate the tester's performance in achieving that state. Of course the disadvantage of this method is that this evaluation result can be different from the true performance in the experiment, because the true parameter can be different from the parameter which we expect.

The second is to consider the average performance. Let μ denote a measure on the parameter space S. We define the average performance of the error probability with respect to a measure μ as

$$\int_S d\mu(\boldsymbol{s})\sigma_1(G_{\boldsymbol{s}}). \tag{26}$$

In this approach, a tester with smaller average rate of decrease is better. The average performance can be calculated without knowing the true parameter, but of course it is not guaranteed that the average value is equivalent to the true performance in the experiment. Since this evaluation results depend upon the choice of the measure μ, we need to ascertain the validity of the choice.

The third is to consider the worst case performance. We define the worst case performance of a tester as

$$\max_{\boldsymbol{s}\in S}\sigma_1(G_{\boldsymbol{s}}). \tag{27}$$

This can be calculated without the true parameter, and it is guaranteed that the true performance is necessarily better or equal to the value. The disadvantage of this method is that we might evaluate the tester's performance much lower than the true performance in the experiment.

B. Extension to infinite sample space

Theorem 1 holds for a finite sample space. For a specific case ($g : S \to \mathbf{R}$ and Δ is the squared absolute value), it is known that Eq.(11) also holds for infinite sample space under some regularity conditions [10, 11]. We can prove that Theorem 1 holds for infinite sample space under some conditions by combining the proof in [10, 11] with Sanov's theorem and using the linear parametrizability of probability distributions in quantum mechanics. Therefore, Theorem 1 holds not only for finite, but also infinite sample spaces. However, any real experiments will have finite detector resolution, and so finite sample spaces suffice.

C. Effect of parameter space boundary

In Theorem 1, the true parameter is limited to the interior S^o. Hence it cannot be applied to parameters on the boundary $\partial S := S\setminus S^o$ which corresponds to the set of all non full rank density operators, including all pure states. This limitation can be overlooked by invoking decoherence: in real experiments the system of interest is uncontrollably affected by the enviroment, leading to full rank states parametrized in the interior. The reason behind the limitation is very technical, stemming from the fact that in our proof we assume the invariance of the support of probability distribution, differentibility, and openness at each point of the parameter space. Such regularity conditions are assumed in standard classical statistical estimation theory. Statistical models that do not satisfy the regularity conditions are called non-regular, and it is known that they can behave very differently from regular statistical models [39]. The analysis of risk functions and error probabilities at ∂S is an open problem.

D. Relation to quantum estimation theory

Our goal in this paper has been to adapt some classical statistical estimation theory to quantum tomography (or, speaking mathematically, to large deviations with more general loss functions on a multiparameter space). It should be pointed out that this is *not* the same thing as quantum statistical estimation theory [40], which can be viewed as the generalisation of classical statistical estimation theory to density matrices instead of probability distributions. There the focus is strictly on quantum states, and one of the main topics is to derive similar estimation bounds for all experimental setups, including collective or adaptive schemes (see [29]). Since our focus here is to evaluate the performance of a fixed experimental apparatus (tester), quantum statistical estimation is ill suited to our purpose. Furthermore, any quantum statistical estimation results that can be made for state estimation would have to be drastically modified for anything other than state tomography. Here we are interested in more general types of quantum tomography.

VI. SUMMARY

In this paper, we proved a large deviation inequality for consistent estimators in quantum tomography by using classical statistical estimation techniques. The inequality shows that, under some conditions, the error probability of any consistent estimator can decrease at most exponentially with respect to the total number of measurement trials, and there is a bound of the rate of decrease which is achievable by a maximum likelihood estimator under the informational completeness of the tester. We also derived the explicit form of the bound and proved that known quantum tomography conditions are equivalent to the identifiability condition in classical estimation theory.

From our results, it is shown that a risk function and error probability measured by the same loss function are characterized by a common matrix, the inverse Fisher matrix modified by the loss function. The rate of decrease (with respect to the number of trials) of the risk function is characterized by the sum of the eigenvalues of this matrix, and that of the error probability by the maximal eigenvalue. The Cramér-Rao inequality, which is a known risk function inequality, holds only for unbiased estimators, and the bound can be broken by biased estimators. On the other hand, the error probability inequality holds for any consistent estimator which gives us the true object in the limit of infinite trials. Therefore, the lower bound of the error probability characterizes the performance of the given apparatus, independently of the choice of estimator. The explicit form of the bound makes it possible to quantify the performance of the apparatus for the estimation purpose in the error probability sense. By combining our error probability approach with a risk function approach, we can evaluate the performance more discerningly than we can by considering only risk functions.

Appendix: Proof of main theorem

We give the detailed proof of Theorem 1, using classical statistical estimation theory. We divide the proof into six parts in order to clarify the role of each condition, as well as to isolate the role of quantum mechanics in the main result.

1. Six lemmas

We first consider the setup described in section III, that is, we do not assume the statistical model given by Eq.(1). Suppose that the parameter space Θ is a closed compact subset of $\mathbf{R}^k$. Let $\partial\Theta$ denote the boundary of Θ, that is, $\partial\Theta := \Theta \setminus \Theta^o$ and assume that Θ^o is open and nonempty. We also assume that $p_\theta(x)$ is a thrice differentiable function with respect to $\theta \in \Theta$ for any $x \in \Omega$. Note that these assumptions are satisfied in quantum mechanics for finite dimensional systems.

First, we prove that Eq.(11) holds for any estimator consistent not only in distances, but also in pseudo-distances.

Lemma 1 *Suppose that Δ is a pseudo-distance on Θ. If $\theta \in \Theta^o$, for an arbitrary consistent estimator $\theta^{\rm est}$ in Δ, the following inequality holds:*

$$\begin{aligned}&\underline{\lim}_{N\to\infty} \frac{1}{N} \log P_\theta^{(N)}(\Delta(\theta_N^{\rm est}, \theta) > \epsilon) \\ &\geq - \inf_{\theta' \in \Theta} \{K(p_{\theta'} \| p_\theta); \Delta(\theta', \theta) > \epsilon\} \end{aligned} \tag{A.1}$$

Proof: This is a straightforward generalizations of the proof in [8], so we omit it here. □

From Lemma 1, we obtain

$$\begin{aligned}&\underline{\lim}_{\epsilon\to 0} \underline{\lim}_{N\to\infty} \frac{1}{\epsilon N} \log P_\theta^{(N)}(\Delta(\theta_N^{\rm est}, \theta) > \epsilon) \\ &\geq - \overline{\lim}_{\epsilon\to 0} \frac{1}{\epsilon} \inf_{\theta' \in \Theta} \{K(p_{\theta'} \| p_\theta); \Delta(\theta', \theta) > \epsilon\}. \end{aligned} \tag{A.2}$$

Second, we introduce a lemma for calculating the R.H.S. of Eq.(A.2).

Lemma 2 *Let A and B be $k \times k$ real, positive-semidefinite matrices. If* ${\rm supp}A \supseteq {\rm supp}B$ *holds, then*

$$\inf_{\boldsymbol{a} \in {\rm supp}B} \{\frac{\boldsymbol{a}\cdot A\boldsymbol{a}}{\boldsymbol{a}\cdot B\boldsymbol{a}}\} = \frac{1}{\sigma_1(\sqrt{B}A^{-1}\sqrt{B})}. \tag{A.3}$$

holds.

Proof: We define C^{-1} as the generalized inverse of a matrix C. Let ${\rm supp}C$ denote the support of C and I_C denote the identity matrix on ${\rm supp}C$, then $C^{-1}C = CC^{-1} = I_C$. Let us define $\boldsymbol{b} := \sqrt{B}\boldsymbol{a}/\|\sqrt{B}\boldsymbol{a}\|$. We can assume that $\boldsymbol{a} \in {\rm supp}B$ *i.e.* $\boldsymbol{a} = I_B\boldsymbol{a}$ holds. Then,

$$\inf_{\boldsymbol{a}} \{\frac{\boldsymbol{a}\cdot A\boldsymbol{a}}{\boldsymbol{a}\cdot B\boldsymbol{a}}\} = \inf_{\boldsymbol{b};\|\boldsymbol{b}\|=1} \{\boldsymbol{b}\cdot \sqrt{B}^{-1} A \sqrt{B}^{-1}\boldsymbol{b}\} \tag{A.4}$$

$$= \sigma_{\min}(\sqrt{B}^{-1} A \sqrt{B}^{-1}) \tag{A.5}$$

$$= 1/\sigma_1([\sqrt{B}^{-1} A \sqrt{B}^{-1}]^{-1}) \tag{A.6}$$

$$= 1/\sigma_1(\sqrt{B}A^{-1}\sqrt{B}), \tag{A.7}$$

where $\sigma_{\min}(C)$ is the minimal non-zero eigenvalue of C and coincides with $\sigma_k(C)$ if C is full rank. □

Third, we calculate the infimum on the R.H.S. of Eq.(A.2).

Lemma 3 *Suppose that Δ is a sufficient smooth pseudo-distance with a non-zero same point Hesse matrix H_θ. Then*

$$\overline{\lim_{\epsilon\to 0}} \frac{1}{\epsilon} \inf_{\theta'\in\Theta} \{K(p_{\theta'}\|p_\theta); \Delta(\theta',\theta) > \epsilon\} = \frac{1}{\sigma_1\big(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta}\big)}. \tag{A.8}$$

holds.

Proof: Let us define $B(\theta',\theta) := 2\frac{\Delta(\theta',\theta)}{\|\theta'-\theta\|^2}$. Then

$$B(\theta',\theta) = \frac{(\theta'-\theta)}{\|\theta'-\theta\|} \cdot H_\theta \frac{(\theta'-\theta)}{\|\theta'-\theta\|} + O(\|\theta'-\theta\|), \tag{A.9}$$

and the first term is independent of $\|\theta'-\theta\|$. Then, for sufficiently small ϵ,

$$\begin{aligned}
&\frac{1}{\epsilon} \inf_{\theta'\in\Theta} \{K(p_{\theta'}\|p_\theta); \Delta(\theta',\theta) > \epsilon\} \\
&= \frac{1}{\epsilon} \inf_{\theta'\in\Theta} \{K(p_{\theta'}\|p_\theta); \|\theta'-\theta\| > \sqrt{\frac{2\epsilon}{B(\theta',\theta)}}\} \\
&= \frac{1}{\epsilon} \inf_{\theta'\in\Theta} \{\frac{1}{2}(\theta'-\theta)F_\theta(\theta'-\theta) + O(\|\theta'-\theta\|^3); \\
&\qquad \|\theta'-\theta\| > \sqrt{\frac{2\epsilon}{B(\theta',\theta)}}\} \qquad (A.10) \\
&= \inf_{\boldsymbol{a}\in \mathrm{supp} H_\theta} \{\frac{\boldsymbol{a}\cdot F_\theta \boldsymbol{a}}{\boldsymbol{a}\cdot H_\theta \boldsymbol{a}}; \|\boldsymbol{a}\| = 1\} \qquad (A.11) \\
&= \frac{1}{\sigma_1\big(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta}\big)}, \qquad (A.12)
\end{aligned}$$

where we used Lemma 2 in the last line. Note that E.(A.8) holds not only for the linit superior $\overline{\lim}_{\epsilon\to 0}$, but also for the limit inferior $\underline{\lim}_{\epsilon\to 0}$. □

From Lemma 1 and Lemma 3, we obtain the following inequality for any estimator consistent in a sufficiently smooth pseudo-distance with the Hesse matrix H_θ:

$$\underline{\lim_{\epsilon\to 0}} \underline{\lim_{N\to\infty}} \frac{1}{\epsilon N} \log P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{est}},\theta) > \epsilon) \geq -\frac{1}{\sigma_1\big(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta}\big)}. \tag{A.13}$$

Fourth, we prove that if the identifiability condition is satisfied, then a maximum likelihood estimator is consistent in the pseudo-distance Δ. In preparation, we introduce empirical measures. Given a finite sequence $\boldsymbol{x}^N = \{x_1,\ldots,x_N\}$ and $Y \in \mathscr{B}$, the empirical measure $L_N^{\boldsymbol{x}^N}$ induced by the sequence is defined as

$$L_N^{\boldsymbol{x}^N}(Y) := \sum_{y\in Y} \frac{1}{N} \sum_{i=1}^N \delta_{y,x_i}, \tag{A.14}$$

where $\delta_{y,x}$ is Kronecker's delta. Then the value of the empirical measure on an elemental set $\{x\} \in \mathscr{B}$ is equivalent to the relative frequency of x for the data $\boldsymbol{x}^N$, *i.e.*, $f_N(x) = L_N^{\boldsymbol{x}^N}(\{x\})$. We identify $L_N^{\boldsymbol{x}^N}$ and f_N below.

Now we introduce Sanov's theorem for empirical measures. Let P_p denote a probability measure on $\mathscr{B}$ with a probability distribution p. When $p \in \mathcal{P}_\Theta$, we have $P_p = P_\theta$. We use a notation $P_p^{(N)}(L_N^{X^N} \in \mathcal{A}) := P_p^{(N)}(\{\boldsymbol{x}^N \in \Omega^N; L_N^{\boldsymbol{x}^N} \in \mathcal{A}\})$, where $\mathcal{A}$ is a given set of probability distributions.

Theorem (Sanov) *For every set $\mathcal{A}$ of probability distributions in $\mathcal{P}(\Omega)$,*

$$\begin{aligned}
-\inf_{p'\in\mathcal{A}^o} K(p'\|p) &\leq \underline{\lim_{N\to\infty}} \frac{1}{N} \log P_p^{(N)}(L_N^{X^N} \in \mathcal{A}) \qquad (A.15) \\
&\leq \overline{\lim_{N\to\infty}} \frac{1}{N} \log P_p^{(N)}(L_N^{X^N} \in \mathcal{A}) \qquad (A.16) \\
&\leq -\inf_{p'\in\mathcal{A}} K(p'\|p), \qquad (A.17)
\end{aligned}$$

where $\mathcal{A}^o$ is the interior of $\mathcal{A}$ considered as a subset of $\mathcal{P}(\Omega)$ and K is the Kullback-Leibler divergence [14, 15].

We are now in a position to prove the following lemma.

Lemma 4 *If the identifiability condition is satisfied, then*

$$\lim_{N\to\infty} P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{ml}},\theta) > \epsilon) = 0 \tag{A.18}$$

holds for any $\epsilon > 0$. That is, a maximum likelihood estimator is consistent in a pseudo-distance Δ on Θ.

Proof: A maximum likelihood estimate θ_N^{ml} can be redefined by using the Kullback-Leibler divergence and the relative frequency as follows:

$$\begin{aligned}
\theta_N^{\mathrm{ml}} &:= \mathrm{argmax}_{\theta\in\Theta} \textstyle\prod_{i=1}^N p_\theta(x_i) \qquad (A.19) \\
&= \mathrm{argmin}_{\theta\in\Theta} K(f_N\|p_\theta). \qquad (A.20)
\end{aligned}$$

Let us define

$$\theta_p := \mathrm{argmin}_{\theta\in\Theta} K(p\|p_\theta). \tag{A.21}$$

Then $\theta_N^{\mathrm{ml}} = \theta_{f_N}$. When analyzing a maximum likelihood estimate θ_N^{ml}, we need to be careful to check whether θ_N^{ml} is included in Θ^o or $\partial\Theta$. Let us introduce four sets of probability distributions $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, and $\mathcal{D}_{\theta,\epsilon}$ as

$$\begin{aligned}
\mathcal{A}_1 &:= \{p \in \mathcal{P}_\Theta; \theta_p \in \Theta^o\}, \qquad (A.22) \\
\mathcal{A}_2 &:= \{p \in \mathcal{P}_\Theta; \theta_p \in \partial\Theta\}, \qquad (A.23) \\
\mathcal{A}_3 &:= \mathcal{P}(\Omega) \setminus \mathcal{P}_\Theta, \qquad (A.24) \\
\mathcal{D}_{\theta,\epsilon} &:= \{p \in \mathcal{P}(\Omega);\ \Delta(\theta_p,\theta) > \epsilon\}. \qquad (A.25)
\end{aligned}$$

If $f_N \in \mathcal{A}_1 \cup \mathcal{A}_2 (= \mathcal{P}_\Theta)$, then $p_{\theta_N^{\mathrm{ml}}} = f_N$. If $f_N \in \mathcal{A}_3$, then $p_{\theta_N^{\mathrm{ml}}} \in \mathcal{A}_2$ and $p_{\theta_N^{\mathrm{ml}}} \neq f_N$. Since $\mathcal{P}(\Omega) = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$ and these sets are disjoint, we can rewrite the error probability as

$$\begin{aligned}
P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{ml}},\theta) > \epsilon) &= P_\theta^{(N)}(f_N \in \mathcal{D}_{\theta,\epsilon}) \qquad (A.26) \\
&= P_\theta^{(N)}(f_N \in \mathcal{A}_1 \cap \mathcal{D}_{\theta,\epsilon}) \\
&\quad + P_\theta^{(N)}(f_N \in \mathcal{A}_2 \cap \mathcal{D}_{\theta,\epsilon}) \\
&\quad + P_\theta^{(N)}(f_N \in \mathcal{A}_3 \cap \mathcal{D}_{\theta,\epsilon}). \qquad (A.27)
\end{aligned}$$

Because Θ is compact and Θ^o is not empty, from Sanov's theorem, we can obtain

$$\lim_{n\to\infty}\frac{1}{N}\log P_\theta^{(N)}(f_N\in\mathcal{A}_j\cap\mathcal{D}_{\theta,\epsilon}) = -\inf_{p\in\mathcal{A}_j\cap\mathcal{D}_{\theta,\epsilon}} K(p\|p_\theta),\ j=1,2,3. \tag{A.28}$$

From the identifiability condition,

$$\inf_{p\in\mathcal{A}_j\cap\mathcal{D}_{\theta,\epsilon}} K(p\|p_\theta) > 0,\ j=1,2,3. \tag{A.29}$$

Therefore, for sufficiently large N, there exists ν, $0<\nu<1$, such that

$$P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{ml}},\theta)>\epsilon)<\nu^N \tag{A.30}$$

holds for any $\epsilon>0$. So, a maximum likelihood estimator is consistent in Δ under the identifiability condition. $\square$

Fifth, we prove that if the identifiability condition is satisfied, a maximum likelihood estimator achieves the equality in Eq.(A.13).

Lemma 5 *Suppose that Δ is a sufficiently smooth pseudo-distance on Θ with a non-zero same point Hesse matrix H_θ. If the identifiability condition is satisfied, then*

$$\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{1}{\epsilon N}\log P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{ml}},\theta)>\epsilon) = -\frac{1}{\sigma_1(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta})} \tag{A.31}$$

holds.

Proof: From the continuity of K and the openness of S^o, for arbitrary $\theta\in\Theta^o$, there exists $\epsilon_0>0$ such that

$$\inf_{p\in\mathcal{A}_1\cap\mathcal{D}_{\theta,\epsilon}} K(p\|p_\theta) < \inf_{p\in\mathcal{A}_j\cap\mathcal{D}_{\theta,\epsilon}} K(p\|p_\theta), \tag{A.32}$$

hold for $j=2,3$ and for any ϵ satisfying $0<\epsilon<\epsilon_0$ [41]. Hence, for sufficiently large N and sufficiently small ϵ,

$$P_\theta^{(N)}(f_N\in\mathcal{A}_1\cap\mathcal{D}_{\theta,\epsilon}) > P_\theta^{(N)}(f_N\in\mathcal{A}_j\cap\mathcal{D}_{\theta,\epsilon}), \tag{A.33}$$

hold for $j=2,3$, and we have

$$\begin{aligned}
&\overline{\lim_{\epsilon\to 0}}\lim_{N\to\infty}\frac{1}{\epsilon N}\log P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{ml}},\theta)>\epsilon)\\
&=\overline{\lim_{\epsilon\to 0}}\lim_{N\to\infty}\frac{1}{\epsilon N}\log\big[P_\theta^{(N)}(f_N\in\mathcal{A}_1\cap\mathcal{D}_{\theta,\epsilon})\\
&\qquad +P_\theta^{(N)}(f_N\in\mathcal{A}_2\cap\mathcal{D}_{\theta,\epsilon})\\
&\qquad +P_\theta^{(N)}(f_N\in\mathcal{A}_3\cap\mathcal{D}_{\theta,\epsilon})\big] &\text{(A.34)}\\
&=\overline{\lim_{\epsilon\to 0}}\lim_{N\to\infty}\frac{1}{\epsilon N}\log P_\theta^{(N)}(f_N\in\mathcal{A}_1\cap\mathcal{D}_{\theta,\epsilon}) &\text{(A.35)}\\
&=\overline{\lim_{\epsilon\to 0}}\frac{1}{\epsilon}\big[-\inf_{p\in\mathcal{A}_1\cap\mathcal{D}_{\theta,\epsilon}}K(p\|p_\theta)\big] &\text{(A.36)}\\
&=-\underline{\lim_{\epsilon\to 0}}\frac{1}{\epsilon}\inf_{\theta'\in\Theta^o}\{K(p_{\theta'}\|p_\theta);\Delta(\theta',\theta)>\epsilon\} &\text{(A.37)}\\
&=-\frac{1}{\sigma_1(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta})}, &\text{(A.38)}
\end{aligned}$$

where we used Lemma 3 in the last line. Because a maximum likelihood estimator satisfies both Eqs.(A.13) and (A.38), it achieves the equality in Eq.(A.13), and Eq.(A.31) holds. $\square$

The final lemma relates the identifiability condition in classical statistical estimation theory to informational completeness in quantum tomography. We assume now that the probability distributions are given by quantum mechanics, Eq.(1), for finite dimensional systems.

Lemma 6 *Let $\hat\rho=\hat\rho(\boldsymbol{s})$ denote a density operator parametrized by a vector $\boldsymbol{s}\in S$. We assume that the parametrization is one-to-one. Suppose that we perform quantum state tomography with a POVM $\boldsymbol{\Pi}=\{\hat\Pi_x\}_{x\in\Omega}$. Then the following statements are equivalent.*

1. *The probability distribution describing the tomographic experiment satisfies the identifiability condition.*
2. *The Fisher matrix $F_{\boldsymbol{s}}$ is full rank for any $\boldsymbol{s}\in S^o$.*
3. *The POVM is informationally complete.*

Proof: First we show that it is sufficient to prove the equivalence of the three conditions in Lemma 6 for a linear parametrization. In quantum mechanics, for a finite dimensional system, any probability distribution is linearly one-to-one parametrizable, and we can assume that the probability distribution has the form

$$p_{\boldsymbol{s}}(x)=v(x)+\boldsymbol{s}\cdot\boldsymbol{w}(x), \tag{A.39}$$

where $v(x)\in\mathbf{R}$ and $\boldsymbol{w}(x)\in\mathbf{R}^{d^2-1}$ satisfy $\sum_{x\in\Omega}p_{\boldsymbol{s}}(x)=1$ for any $\boldsymbol{s}\in S$. If the probability distribution is one-to-one (but not necessarily linearly) parametrized by a different parameter $\boldsymbol{t}\in\mathbf{R}^{d^2-1}$, then we have

$$\tilde p_{\boldsymbol{t}}(x)=p_{\boldsymbol{s}(\boldsymbol{t})}(x), \tag{A.40}$$

$$\tilde\nabla_{\boldsymbol{t}}\tilde p_{\boldsymbol{t}}(x)=\frac{\partial\boldsymbol{s}}{\partial\boldsymbol{t}}\nabla_{\boldsymbol{s}}p_{\boldsymbol{s}}(x). \tag{A.41}$$

Condition 1 for $\boldsymbol{s}$ and condition 1 for $\boldsymbol{t}$ are equivalent because both parametrizations are one-to-one. Condition 2 for $\boldsymbol{s}$ and condition 2 for $\boldsymbol{t}$ are equivalent because the Fisher matrices satisfy the equation

$$\tilde F_{\boldsymbol{t}}=\frac{\partial\boldsymbol{s}}{\partial\boldsymbol{t}}F_{\boldsymbol{s}}\frac{\partial\boldsymbol{s}}{\partial\boldsymbol{t}}^T, \tag{A.42}$$

and the Jacobian $\frac{\partial\boldsymbol{s}}{\partial\boldsymbol{t}}$ is full rank. Condition 3 is independent of state parametrization. Therefore if condition 1, 2, and 3 are equivalent for a linear parametrization, then they are also equivalent for a general parametrization.

Next we prove the equivalence of conditions 1 and 2. As in the above discussion, without loss of generality, we can assume that $\boldsymbol{s}$ is the fixed parameter such that $F_{\boldsymbol{s}}$ is diagonalized because this is a linear transformation of a general parameter. Under this assumption, condition 1 is equivalent to the condition that for any $\boldsymbol{s}\in S^o$ and for all $\alpha = 1, \ldots, d^2-1$, there exists at least one $x \in \Omega$ such that

$$\partial_\alpha p_{\boldsymbol{s}}(x) \neq 0, \tag{A.43}$$

where $\partial_\alpha := \frac{\partial}{\partial s^\alpha}$. On the other hand, the diagonal elements of the Fisher matrix are

$$F_{\boldsymbol{s},\alpha\alpha} = \sum_{x\in\Omega} \frac{(\partial_\alpha p_{\boldsymbol{s}}(x))^2}{p_{\boldsymbol{s}}(x)},\ \alpha = 1, \ldots, d^2-1. \tag{A.44}$$

Therefore the full rankness of the Fisher matrix is equivalent to Eq.(A.43), and condition 1 and condition 2 are equivalent.

Third we prove the equivalence between condition 2 and 3. We choose the generalized Bloch parametrization of density operators [27, 28]; any density operator $\hat{\rho}$ can be represented as

$$\hat{\rho}(\boldsymbol{s}) = \frac{1}{d}\hat{I} + \frac{1}{2}\boldsymbol{s}\cdot\hat{\boldsymbol{\sigma}}, \tag{A.45}$$

where $\hat{I}$ is the identity operator on $\mathcal{H}$ and $\hat{\sigma}_\alpha$ are generators of $SU(d)$ satisfying $\hat{\sigma}_\alpha = \hat{\sigma}_\alpha^\dagger$, $\mathrm{Tr}[\hat{\sigma}_\alpha] = 0$, and $\mathrm{Tr}[\hat{\sigma}_\alpha\hat{\sigma}_\beta] = 2\delta_{\alpha,\beta}$ ($\alpha,\beta = 1,\ldots,d^2-1$). To determine the representation uniquely, we need more additional conditions on $\hat{\boldsymbol{\sigma}}$, but the additional conditions are not used in the following discussion. Each element of the tester POVM can be represented as

$$\hat{\Pi}_x = v(x)\hat{I} + \boldsymbol{w}(x)\cdot\hat{\boldsymbol{\sigma}},\ x\in\Omega, \tag{A.46}$$

where $v(x)$ and $\boldsymbol{w}(x)$ should satisfy $\sum_{x\in\Omega} v(x) = 1$, $\sum_{x\in\Omega}\boldsymbol{w}(x) = \boldsymbol{0}$, and $\hat{\Pi}_x \geq 0$ for any $x\in\Omega$. Then, the probability distribution describing the tomographic experiment is represented as Eq.(A.39), and the Fisher matrix is

$$F_{\boldsymbol{s}} = \sum_{x\in\Omega} p_{\boldsymbol{s}}(x)\nabla_{\boldsymbol{s}}\log p_{\boldsymbol{s}}(x)\nabla_{\boldsymbol{s}}\log p_{\boldsymbol{s}}(x)^T \tag{A.47}$$

$$= \sum_{x\in\Omega}\frac{\boldsymbol{w}(x)\boldsymbol{w}(x)^T}{p_{\boldsymbol{s}}(x)}. \tag{A.48}$$

Therefore the full rankness of the Fisher matrix is equivalent to the condition that $\{\boldsymbol{w}(x)\}_{x\in\Omega}$ spans $\mathbf{R}^{d^2-1}$, and this implies that the tester POVM $\boldsymbol{\Pi}$ is informationally complete. □

2. A more general theorem

From Eq.(A.13) and Eq.(A.31), we obtain the following theorem.

Theorem 2 *Suppose that Δ on Θ is a sufficiently smooth pseudo-distance with a non-zero same point Hesse matrix H_θ. If $\theta\in\Theta^o$, for an arbitrary consistent estimator θ^{est}, the following inequality holds:*

$$\varliminf_{\epsilon\to 0}\varliminf_{N\to\infty}\frac{1}{\epsilon N}\log P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{est}},\theta) > \epsilon) \geq -1/\sigma_1(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta}). \tag{A.49}$$

Furthermore, when the identifiability condition is satisfied, a maximum likelihood estimator θ^{ml} is consistent and achieves the equality in Eq.(A.49), i.e.,

$$\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{1}{\epsilon N}\log P_\theta^{(N)}(\Delta(\theta_N^{\mathrm{ml}},\theta) > \epsilon) = -1/\sigma_1(\sqrt{H_\theta}F_\theta^{-1}\sqrt{H_\theta}) \tag{A.50}$$

holds.

Theorem 2 is in fact more general than Theorem 1, since identifiability is more general than informational completeness. Hence, the properties that the error probabilities of consistent estimators can decrease at most exponentially, the rate of decrease is bounded by the maximal eigenvalue of a matrix, and the bound is achievable by a maximum-likelihood estimator are common to a larger class of probability distributions than those of quantum mechanics.

By applying Theorem 2 to quantum state tomography and using Lemma 6, we can obtain Theorem 1. Theorem 2 is applicable to the other types of quantum tomography. The conditions corresponding to the identifiability condition are different, and can be derived in the same way as in the proof of Lemma 6. For example, let us consider ancilla-unassisted quantum process tomography. To identify an unknown quantum process described by a linear, completely-positive, and trace-preserving map κ on $\mathcal{S}(\mathcal{H})$, we prepare a set of input states $\boldsymbol{\rho} = \{\hat{\rho}_n\}_{n=1}^{N_s}$ where $\hat{\rho}_n\in\mathcal{S}(\mathcal{H})$ and a measurement described by a POVM $\boldsymbol{\Pi} = \{\hat{\Pi}_x\}_{x\in\Omega}$ on $\mathcal{H}$. The set $\{\boldsymbol{\rho},\boldsymbol{\Pi}\}$ is the tester for ancilla-unassisted process tomography. When $\boldsymbol{\rho}$ spans $\mathcal{S}(\mathcal{H})$, it is called tomographically complete. In ancilla-unassisted process tomography, the informational completeness of $\boldsymbol{\Pi}$ and the tomographical completeness of $\boldsymbol{\rho}$ both are required. We can prove that these conditions are equivalent to the identifiability condition in the same way as in lemma 6.

For the case where the same point Hesse matrix of the loss function is positive semidefinite, as mentioned in subsection IV D, the identifiability condition is modified as follows: for any $\theta\in\Theta^o$ and $\theta'\in\Theta$, if $g(\theta)\neq g(\theta')$, then there exists at least one single outcome $x\in\Omega$ satisfying $p_\theta(x)\neq p_{\theta'}(x)$. Theorem 2 holds for this modification.

ACKNOWLEDGMENTS

T.S. thanks Yu Watanabe for helpful discussions. P.S.T. acknowledges useful discussion with P. Kim. This work was supported by JSPS Rearch Fellowships for Young Scientists, JSPS KAKENHI (20549002) for Scientific Research (C), and Special Coordination Funds for Promoting Science and Technology.

[1] *Elements of Quantum Information*, edited by W. P. Schleich and H. Walther (Wiley-VCH, Weinheim, 2007).
[2] *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics (Springer, Berlin, 2004).
[3] C. R. Rao, *Linear statistical inference and its applications* (John Wiley, New York, 1973).
[4] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[5] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Phys. Rev. A **70**, 052321 (2004).
[6] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz-Tapia, Phys. Rev. A **73**, 032301 (2006).
[7] J. Nunn, B. J. Smith, G. Puentes, and I. A. Walmsley, arXiv: quant-ph/0911.4310.
[8] R. R. Bahadur, J. C. Gupta, and S. L. Zabell, in *Asymptotic Theory of Statistical Tests and Estimates* (Academic Press, New York, 1980), pp. 33–64.
[9] X. Shen, Statistica Sinica **11**, 479 (2001).
[10] R. R. Bahadur, Sankhyā, **22**, 229 (1960).
[11] R. R. Bahadur, Ann. Math. Statist. **38**, 303 (1967).
[12] M. Hayashi, J. Phys. A: Math. and Gen. **35**, 7689 (2002).
[13] M. Hayashi and K. Matsumoto., IEICE Trans. A **83**, 629 (2000) (in Japanese).
[14] I. N. Sanov, Mat. Sbornik. **42**, 11 (1957). (English translation in *Sel. Transl. Math. Stat. and Prob.*, Vol.1, pp.213-244, 1961).
[15] A. Dembo and O. Zeitouni, *Large deviations Techniques and Applications*, 2nd ed. (Springer, New York, 1998).
[16] G. M. D'Ariano and P. L. Presti, Phys. Rev. Lett. **86**, 4195 (2001).
[17] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[18] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[19] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Phys. Rev. A **61**, 010304(R) (1999).
[20] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).
[21] I. L. Chuang and M. A. Nielsen, J. Mod. Phys. **44**, 2455 (1997).
[22] V. Buzek, Phys. Rev. A **58**, 1723 (1998).
[23] J. Fiurasek and Z. Hradil, Phys. Rev. A **63**, 020101(R) (2001).
[24] M. F. Sacchi, Phys. Rev. A **63**, 054104 (2001).
[25] A. Luis and L. L. Sanchez-Sato, Phys. Rev. Lett. **83**, 3573 (1999).
[26] J. Fiurasek, Phys. Rev. A **64**, 024102 (2001).
[27] G. Kimura, Phys. Lett. A **314**, 339 (2003).
[28] M. S. Byrd and N. Khaneja, Phys. Rev. A **68**, 062322 (2003).
[29] In quantum tomography, it is possible to change the tester used in the next trial depending on the previous observation results. Such an experimental scheme is called *adaptive*. An experimental scheme which allows a global measurement on more than one system at once is called *collective*. Both of these generalizations constitute significantly more complicated experiments that are not currently the norm, and are not treated in this paper.
[30] E. Prugovečki, Int. J. Theor. Phys. **16**, 321 (1977).
[31] G. A. V. Buzek, R. Derka and P. L. Knight, Ann. Phys., **266**, 454 (1998).
[32] R. Schack, T. A. Brun, and C. M. Caves, Phys. Rev. A **64**, 014305 (2001).
[33] C. A. Fuchs, R. Schack, and P. F. Scudo, Phys. Rev. A **69**, 062305 (2004).
[34] V. Buzek and G. Drobny, J. Mod. Opt. **47**, 2823 (2000).
[35] R. L. Kosut, arXiv: quant-ph/0812.4323.
[36] The terminology differs by textbook.
[37] A. Wald, Ann. Math. Statist. **20**, 595 (1949).
[38] A. L. Rukhin, Ann. Statist. **11**, 202 (1983).
[39] M. Akahara and K. Takeuchi, *Non-regular Statistical estimation* (Springer, New York, 1995).
[40] *Asymptotic Theory of Quantum Statistical Inference*, edited by M. Hayashi (World Scientific, Singapore, 2005).
[41] The upper bound of ϵ_0 depends on the true parameter θ, and if ϵ is in $(0, \epsilon_0)$, a maximum likelihood estimator achieves the equality in Eq.(11) as pointed out in [42]. This is because the sample space is finite. If the sample space is infinite, the equality for arbitrary finite ϵ is not achievable by any estimate [38].
[42] A. D. Kester and W. C. M. Kallenberg, Ann. Statist. **14**, 648 (1986).