Two quantum analogues of Fisher information from a large deviation viewpoint of quantum estimation

2002 J. Phys. A: Math. Gen. 35 7689

(http://iopscience.iop.org/0305-4470/35/36/302)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 7689–7727 PII: S0305-4470(02)33676-X

Two quantum analogues of Fisher information from a large deviation viewpoint of quantum estimation

Masahito Hayashi

Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN 2-1, Hirosawa, Wako, Saitama 351-0198, Japan

E-mail: masahito@brain.riken.go.jp

Received 12 February 2002, in final form 14 June 2002
Published 28 August 2002
Online at stacks.iop.org/JPhysA/35/7689

Abstract

We discuss two quantum analogues of the Fisher information, the symmetric logarithmic derivative Fisher information and Kubo–Mori–Bogoljubov Fisher information from a large deviation viewpoint of quantum estimation and prove that the former gives the true bound and the latter gives the bound of consistent superefficient estimators. As another comparison, it is shown that the difference between them is characterized by the change of the order of limits.

PACS numbers: 03.67.−a, 02.50.Tt

1. Introduction

Fisher information plays a central role in statistical inference, but also coincides with a natural inner product in a distribution family. It is defined as

$$J_\theta := \int_\Omega l_\theta(\omega)^2 p_\theta(\omega)\,\mathrm{d}\omega \qquad l_\theta(\omega)\,p_\theta(\omega) = \frac{\mathrm{d}\,p_\theta(\omega)}{\mathrm{d}\theta} \tag{1}$$

for a probability distribution family $\{p_\theta|\theta \in \Theta \subset \mathbb{R}\}$ with a probability space Ω. However, the quantum version of Fisher information cannot be uniquely determined. In general, there is a serious arbitrariness concerning the order among non-commutative observables in the quantization of products of several variables. The problem of the arbitrarity of the quantum version of Fisher information is caused by the same reason. The geometrical properties of its quantum analogues have been discussed by many authors [1–4].

One quantum analogue is the Kubo–Mori–Bogoljubov (KMB) Fisher information $\tilde{J}_\rho$ defined by

$$\tilde{J}_\theta := \int_0^1 \mathrm{Tr}\,\rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t} \tilde{L}_\theta\,\mathrm{d}t \qquad \int_0^1 \rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t}\,\mathrm{d}t = \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta} \tag{2}$$

for a quantum state family $\{\rho_\theta \in \mathcal{S}(\mathcal{H})|\theta \in \Theta\}$, where $\mathcal{S}(\mathcal{H})$ is the set of density matrices on $\mathcal{H}$ and the Hilbert space $\mathcal{H}$ corresponds to the physical system of interest [1–4]. As proved

in appendix B, it can be characterized as the limit of quantum relative entropy, which plays an important role in several topics of quantum information theory, for example, quantum channel coding [5, 6], quantum source coding [7–9] and quantum hypothesis testing [10, 11]. Moreover, as mentioned in section 3, this inner product is closely related to the canonical correlation of the linear response theory in statistical mechanics [12]. As mentioned in appendix A, it appears to be the most natural quantum extension from an information-geometrical viewpoint. Thus, one might expect that it is significant in quantum estimation, but its estimation-theoretical characterization has not been sufficiently clarified.

Another quantum analogue is the symmetric logarithmic derivative (SLD) Fisher information

$$J_\theta := \mathrm{Tr}\, L_\theta^2 \rho_\theta \qquad \frac{1}{2}(L_\theta \rho_\theta + \rho_\theta L_\theta) = \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta} \tag{3}$$

where L_θ is called the symmetric logarithmic derivative [13]. It is closely related to the achievable lower bound of mean square error (MSE) not only for the one-parameter case [13–15], but also for the multi-parameter case [16–18] in quantum estimation. The difference between the two can be regarded as the difference in the order of the operators, and reflects the two ways of defining Fisher information for a probability distribution family.

Currently, the former is closely related to the quantum information theory while the latter is related to the quantum estimation theory. These two inner products have been discussed from separated contexts only. In this paper, to clarify the difference, we introduce a large deviation viewpoint of quantum estimation as a unified viewpoint, whose classical version was initiated by Bahadur [19–21]. This method may not be standard in mathematical statistics, but seems a suitable setting for a comparison between two quantum analogues from an estimation viewpoint. This type of comparison was initiated by Nagaoka [22, 23], and is discussed in further depth in this paper. Such a large deviation evaluation of quantum estimation is closely related to the exponent of the overflow probability of quantum universal variable-length coding [24].

This paper is structured as follows: before we state the main results, we summarize the classical estimation theory including Bahadur's large deviation theory, which has been done in section 2. After this summary, we briefly outline the main results in section 3, i.e. the difference is characterized from three contexts. To simplify the notation, even if we need the Gauss notation [ ], we omit it when this does not cause confusion. Some proofs are very complicated and are presented in the appendices.

2. Summary of classical estimation theory

We summarize the relationship between the parameter estimation for the probability distribution family $\{p_\theta | \theta \in \Theta \subset \mathbb{R}\}$ with a probability space Ω and its Fisher information. The definitions of Fisher information are given not only by (1), but also by the limit of the relative entropy (Kullback–Leibler divergence) $D(p\|q) := \int_\Omega (\log p(\omega) - \log q(\omega)) p(\omega)\, \mathrm{d}\omega$ as

$$J_\theta := \lim_{\epsilon \to 0} \frac{2}{\epsilon^2} D(p_{\theta+\epsilon} \| p_\theta). \tag{4}$$

These two definitions (1) and (4) coincide under some regularity conditions for a family.

Next, we consider a map f from Ω to Ω'. Similar to other information quantities (for example, Kullback divergence, etc) the inequality

$$J_\theta \geqslant J'_\theta \tag{5}$$

holds, where J'_θ is the Fisher information of the family $\{p_\theta \circ f^{-1} | \theta \in \Theta\}$. Inequality (5) is called monotonicity. According to Čencov [25], any information quantity satisfying (5) coincides with a constant times Fisher information J_θ.

For an estimator that is defined as a map from the dataset Ω to the parameter set Θ, we sometimes consider the *unbiasedness* condition:

$$\int_\Omega T(\omega) p_\theta(\omega)\, \mathrm{d}\omega = \theta \qquad \forall \theta \in \Theta. \tag{6}$$

The MSE of any unbiased estimator T is evaluated by the following inequality (Cramér–Rao inequality):

$$\int_\Omega (T(\omega) - \theta)^2 p_\theta(\omega)\, \mathrm{d}\omega \geqslant \frac{1}{J_\theta} \tag{7}$$

which follows from the Schwartz inequality w.r.t. the inner product $\langle X, Y \rangle := \int_\Omega X(\omega) Y(\omega) p_\theta(\omega)\, \mathrm{d}\omega$ for variables X, Y. When the number of data $\vec{\omega}_n := (\omega_1, \ldots, \omega_n)$, which obey the unknown probability p_θ, is sufficiently large, we discuss a sequence $\{T_n\}$ of estimators $T_n(\vec{\omega}_n)$. If $\{T_n\}$ is suitable as a sequence of estimators, we can expect that it converges to the true parameter θ in probability, i.e. it satisfies the *weak consistency* condition:

$$\lim_{n\to\infty} p_\theta^n\{|T_n - \theta| > \epsilon\} = 0 \qquad \forall \epsilon > 0 \quad \forall \theta \in \Theta. \tag{8}$$

Usually, the performance of a sequence $\{T_n\}$ of estimators is measured by the speed of its convergence. As one criterion, we focus on the speed of the convergence in MSE. If a sequence $\{T_n\}$ of estimators satisfies the weak consistency condition (WC) and some regularity conditions, the asymptotic version of the Cramér–Rao inequality

$$\lim_{n\to\infty} n \int_\Omega (T_n(\vec{\omega}_n) - \theta)^2 p_\theta^n(\omega)\, \mathrm{d}\omega \geqslant \frac{1}{J_\theta} \tag{9}$$

holds. If it satisfies only the weak consistency condition, it is possible that it surpasses the bound of (9) at a specific subset. Such a sequence of estimators is called superefficient. We can reduce its error to any amount at a specific subset with the measure 0 under the weak consistency condition (8).

As another criterion, we evaluate the decreasing rate of the tail probability:

$$\beta(\{T_n\}, \theta, \epsilon) := \lim_{n\to\infty} \frac{-1}{n} \log p_\theta^n\{|T_n - \theta| > \epsilon\}. \tag{10}$$

This method was initiated by Bahadur [19–21], and was a much discussed topic among mathematical statisticians in the 1970s. From the monotonicity of the divergence, we can prove the inequality

$$\beta(\{T_n\}, \theta, \epsilon) \leqslant \min\{D(p_{\theta+\epsilon} \| p_\theta), D(p_{\theta-\epsilon} \| p_\theta)\} \tag{11}$$

for any weakly consistent sequence $\{T_n\}$ of estimators. Its proof is essentially given in our proof of theorem 2. Since it is difficult to analyse $\beta(\{T_n\}, \theta, \epsilon)$ except in the case of an exponential family, we focus on another quantity $\alpha(\{T_n\}, \theta) := \lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \beta(\{T_n\}, \theta, \epsilon)$. For an exponential family, see appendix K. Taking the limit $\epsilon \to +0$, we obtain the inequality

$$\alpha(\{T_n\}, \theta) \leqslant \frac{J_\theta}{2}. \tag{12}$$

If T_n is the maximum likelihood estimator (MLE), the equality of (12) holds under some regularity conditions for the family [21, 26]. This type of discussion is different from the MSE type of discussion in deriving (12) from only the weak consistency condition. Therefore, there is no consistent superefficient estimator w.r.t. the large deviation evaluation.

Indeed, we can relate the above large deviation type of discussion in the estimation to Stein's lemma in simple hypothesis testing as follows. In simple hypothesis testing, we decide whether the null hypothesis should be accepted or rejected from the data $\vec{\omega}_n := (\omega_1, \ldots, \omega_n)$ which obey an unknown probability. For the decision, we must define an *accept region* A_n as a subset of Ω^n. If the null hypothesis is p and the alternative is q, the first error (though the true distribution is p, we reject the null hypothesis) probability $\beta_{1,n}(A_n)$ and the second error (though the true distribution is q, we accept the null hypothesis) probability $\beta_{2,n}(A_n)$ are given by

$$\beta_{1,n}(A_n) := 1 - p^n(A_n) \qquad \beta_{2,n}(A_n) := q^n(A_n).$$

Regarding the decreasing rate of the second error probability under the constant constraint of the first error probability, the equation

$$\lim_{n\to\infty} \frac{-1}{n} \log \min\{\beta_{2,n}(A_n) | \beta_{1,n}(A_n) \leqslant \epsilon\} = D(p\|q) \qquad \epsilon > 0 \tag{13}$$

holds (Stein's lemma). Inequality (11) can be derived from this lemma. We can regard the large deviation type of evaluation in the estimation to be the limit of Stein's lemma in the case where the null hypothesis is close to the alternative one.

3. Outline of main results

Let us return to the quantum case. In a quantum setting, we focus two quantum analogues of the Fisher information, the KMB Fisher information $\tilde{J}_\theta$ and the SLD Fisher information J_θ. Indeed, if the state ρ_θ is non-degenerate, SLD L_θ is not uniquely determined. However, as is proved in appendix C, the SLD Fisher information J_θ is uniquely determined, i.e. it is independent of the choice of the SLD L_θ.

On the other hand, according to chapter 7 in Amari and Nagaoka [1], $\tilde{L}_\theta$ has another form

$$\tilde{L}_\theta = \frac{\mathrm{d}\log\rho_\theta}{\mathrm{d}\theta}. \tag{14}$$

As is proved by using formula (14) in appendix B, the KMB Fisher information $\tilde{J}_\theta$ can be characterized as the limit of the quantum relative entropy $D(\rho\|\sigma) := \mathrm{Tr}\,\rho(\log\rho - \log\sigma)$ in the following way:

$$\tilde{J}_\theta = \lim_{\epsilon\to 0} \frac{2}{\epsilon^2} D(\rho_{\theta+\epsilon}\|\rho_\theta). \tag{15}$$

Moreover, in the linear response theory of statistical physics, given an equilibrium state ρ, when a variable A fluctuates with a small value δ, another variable B also is thought to fluctuate with a constant times δ [12]. Its coefficient is called the canonical correlation and is given by

$$\int_0^1 \mathrm{Tr}\,\rho_\theta^t (A - \mathrm{Tr}\,\rho A)\rho_\theta^{1-t}(B - \mathrm{Tr}\,\rho B)\,\mathrm{d}t. \tag{16}$$

Thus, the KMB Fisher information $\tilde{J}_\theta$ is thought to be more natural from the viewpoint of statistical physics.

As another quantum analogue, the right logarithmic derivative (RLD) Fisher information $\check{J}_\theta$

$$\check{J}_\theta := \mathrm{Tr}\,\rho_\theta \check{L}_\theta \check{L}_\theta^* \qquad \rho_\theta \check{L}_\theta = \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}$$

is known. When ρ_θ does not commute $\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}$ and $\rho_\theta > 0$, the RLD $\check{L}_\theta$ is not self-adjoint. Since it is not useful in the one-parameter case, we do not discuss it in this paper. Since

the difference in definitions can be regarded as the difference in the order of operators, these quantum analogues coincide when all states of the family are commutative with each other. However, in the general case, they do not coincide and the inequality $\tilde{J}_\theta \geqslant J_\theta$ holds, as exemplified in section 4. Concerning some information-geometrical properties, see appendix A.

In the following, we consider what roles these quantum analogues of Fisher information play in the parameter estimation for the state family. As is discussed in detail in section 4, the estimator is described by the pair of positive operator valued measure (POVM) M (which corresponds to the measurement and is defined in section 4) and the map from the dataset to the parameter space Θ. Similar to the classical case, we can define an unbiased estimator. For any unbiased estimator E, the SLD Cramér–Rao inequality

$$V(E) \geqslant \frac{1}{J_\theta} \tag{17}$$

holds, where $V(E)$ is the MSE of the estimator E.

In an asymptotic setting, as a quantum analogue of the n-i.i.d. condition, we treat the quantum n-i.i.d. condition, i.e. we consider the case where the number of systems that are independently prepared in the same unknown state is sufficiently large, in section 5. In this case, the measurement is denoted by a POVM M^n on the composite system $\mathcal{H}^{\otimes n}$ and the state is described by the density $\rho^{\otimes n}$. Of course, such POVMs include a POVM that requires quantum correlations between the respective quantum systems in the measurement apparatus. Similar to the classical case, for a sequence $\vec{E} = \{E^n\}$ of estimators, we can define the *weak consistency* condition given in (31). In mathematical statistics, the square root n consistency, local asymptotic minimax theorems and Bayesian theorem are important topics as the asymptotic theory, but it seems too difficult to link these quantum settings and the KMB Fisher information $\tilde{J}_\theta$. Thus, in this paper, in order to compare two quantum analogues from a unified framework, we adopt Bahadur's large deviation theory as follows. As is discussed in section 5, we can similarly define the quantities $\beta(\vec{E}, \theta, \epsilon)$ and $\alpha(\vec{E}, \theta)$. Similar to (11) and (12), under the weak consistency (WC) condition, the inequalities

$$\beta(\vec{E}, \theta, \epsilon) \leqslant \min\{D(\rho_{\theta+\epsilon} \| \rho_\theta), D(\rho_{\theta-\epsilon} \| \rho_\theta)\} \qquad \alpha(\vec{E}, \theta) \leqslant \tfrac{1}{2}\tilde{J}_\theta \tag{18}$$

hold. From these discussions, the bound in the large deviation type of evaluation seems different from that in the MSE case. However, as mentioned in section 6, the inequality

$$\alpha(\vec{E}, \theta) \leqslant \tfrac{1}{2} J_\theta \tag{19}$$

holds if the sequence $\vec{E}$ satisfies the strong consistency (SC) condition introduced in section 6 as a stronger condition. As is mentioned in section 7, these bounds can be attained in their respective senses. Therefore, roughly speaking, the difference between the two quantum analogues can be regarded as the difference in consistency conditions and can be characterized as

$$\sup_{\vec{E}:\mathrm{SC}} \lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\beta(\vec{E}, \theta, \epsilon) = \frac{1}{2} J_\theta \qquad \sup_{\vec{E}:\mathrm{WC}} \lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\beta(\vec{E}, \theta, \epsilon) = \frac{1}{2}\tilde{J}_\theta.$$

Even if we restrict our estimators to strongly consistent ones, the difference between the two appears as

$$\sup_{\vec{M}:\mathrm{SC}} \liminf_{\epsilon\to 0} \frac{1}{\epsilon^2}\beta(\vec{M}, \theta, \epsilon) = \frac{J_\theta}{2} \tag{20}$$

$$\liminf_{\epsilon\to 0} \frac{1}{\epsilon^2} \sup_{\vec{M}:\mathrm{SC}} \beta(\vec{M}, \theta, \epsilon) = \frac{\tilde{J}_\theta}{2} \tag{21}$$

where, for a precise statement, as expressed in section 9, we need more complicated definitions.

However, we should consider that the bound $\frac{J_\theta}{2}$ is more meaningful for the following two reasons. The first reason is the fact that we can construct the sequence of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ at all points, which is proved in section 7. On the other hand, there is a sequence of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ at one point θ, but it cannot attain the bound at all points. The other reason is the naturalness of the conditions for deriving the bound $\frac{J_\theta}{2}$. In other words, an estimator attaining $\frac{J_\theta}{2}$ is natural, but an estimator attaining $\frac{\tilde{J}_\theta}{2}$ is very irregular. Such a sequence of estimators can be regarded as a consistent superefficient estimator and does not satisfy regularity conditions other than the weak consistency condition. This type of discussion of the superefficiency is different from the MSE type of discussion in that any consistent superefficient estimator is bounded by inequality (18).

To consider the difference between the two quantum analogues of the Fisher information in more details, we must analyse how we can achieve the bound $\frac{\tilde{J}_\theta}{2}$. It is important for this analysis to consider the relationship between the above discussion and the quantum version of Stein's lemma in simple hypothesis testing. Similar to the classical case, when the null hypothesis is the state ρ and the alternative is the state σ, we evaluate the decreasing rate of the second error probability under the constant constraint of the first error probability. As was proved in quantum Stein's lemma, its exponential component is given by the quantum relative entropy $D(\rho\|\sigma)$ for any $\epsilon > 0$. Hiai and Petz [10] constructed a sequence of tests to attain the optimal rate $D(\rho\|\sigma)$, by constructing the sequence $\{M^n\}$ of POVMs such that

$$\lim_{n\to\infty} \frac{1}{n} D\left(\mathrm{P}_\rho^{M^n} \middle\| \mathrm{P}_\sigma^{M^n}\right) = D(\rho\|\sigma). \tag{22}$$

Ogawa and Nagaoka [11] proved that there is no test exceeding the bound $D(\rho\|\sigma)$. It was proven by Hayashi that by using the group representation theory, we can construct the POVM satisfying (22) independently of ρ. For the reader's convenience, we give a summary of this in appendix J. As discussed in section 7.2, this type of construction is useful for the construction of an estimator attaining the bound $\frac{\tilde{J}_\theta}{2}$ at one point. Since the proper bound of the large deviation is $\frac{J_\theta}{2}$, we cannot regard the quantum estimation as the limit of the quantum Stein's lemma.

In order to consider the properties of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ at one point from another viewpoint, we consider the restriction that makes such a construction impossible. We introduce a class of estimators whose POVMs do not need a quantum correlation in the quantum apparatus in section 8. In this class, we assume that the POVM on the lth system is chosen from $l-1$ data. We call such an estimator an adaptive estimator. When an adaptive estimator $\vec{E}$ satisfies the weak consistency condition, the inequality

$$\alpha(\vec{E}, \theta) \leqslant \tfrac{1}{2} J_\theta \tag{23}$$

holds (see section 6). Similarly, we can define a class of estimators that use quantum correlations up to m systems. We call such an estimator an m-adaptive estimator. For any m-adaptive weakly consistent estimator $\vec{E}$, inequality (23) holds. Therefore, it is impossible to construct a sequence of estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ if we fix the number of systems in which we use quantum correlations. As mentioned in section 8, taking limit $m \to \infty$, we have

$$\lim_{m\to\infty} \lim_{\epsilon\to 0} \sup_{\vec{M}:m\text{-AWC}} \frac{1}{\epsilon^2} \beta(\vec{M}, \theta, \epsilon) = \frac{J_\theta}{2} \tag{24}$$

where m-AWC denotes an m-adaptive weakly consistent estimator. However, as the third characterization of the difference between the two quantum analogues, as precisely mentioned in section 9, we have

$$\lim_{\epsilon\to 0}\lim_{m\to\infty}\sup_{\vec{M}:m\text{-ASC}}\frac{1}{\epsilon^2}\beta(\vec{M},\theta,\epsilon)=\frac{\tilde{J}_\theta}{2} \tag{25}$$

where m-ASC denotes an m-adaptive strongly consistent estimator. A more narrow class of estimators is treated in equation (25) than in equation (21). Equations (24) and (25) indicate that the order of limits $\lim_{m\to\infty}$ and $\lim_{\epsilon\to 0}$ is more crucial than the difference between two types of consistencies.

Remark 1. In the estimation of only the spectrum of a density operator in a unitary-invariant family, the natural inner product in the parameter space is unique and equals the Fisher inner product in the distribution family whose element is the probability distribution corresponding to the eigenvalues of a density matrix. In addition, the achievable bound was as was derived by Keyl and Werner [28], and coincides with the bound uniquely given by the above inner product.

4. Summary of non-asymptotic setting in quantum estimation

In a quantum system, in order to discuss the probability distribution which the data obey, we must define a POVM.

A POVM M is defined as a map from the Borel sets of the dataset Ω to the set of bounded, self-adjoint and positive semi-definite operators, which satisfies

$$M(\emptyset)=0 \qquad M(\Omega)=\mathrm{I} \qquad \sum_i M(B_i)=M(\cup B_i) \quad \text{for disjoint sets.}$$

If the state on the quantum system $\mathcal{H}$ is a density operator ρ and we perform a measurement corresponding to a POVM M on the system, the data obey the probability distribution $\mathrm{P}_\rho^M(B):=\mathrm{Tr}\,\rho M(B)$. If a POVM M satisfies $M(B)^2=M(B)$ for any Borel set B, M is called a projection-valued measure (PVM). The spectral measure of a self-adjoint operator X is a PVM, and is denoted by $E(X)$. For $1>\lambda>0$ and any POVMs M_1 and M_2 taking values in Ω, the POVM $B\mapsto\lambda M_1(B)+(1-\lambda)M_2(B)$ is called the *random combination* of M_1 and M_2 in the ratio $\lambda:1-\lambda$. Even if M_1's dataset Ω_1 is different from M_2's dataset Ω_2, M_1 and M_2 can be regarded as POVMs taking values in the disjoint union set $\Omega_1\coprod\Omega_2:=(\Omega_1\times\{1\})\cup(\Omega_2\times\{2\})$. In this case, we can define a random combination of M_1 and M_2 as a POVM taking values in $\Omega_1\coprod\Omega_2$ and call it the *disjoint random combination*. In this paper, we simplify the probability $\mathrm{P}_{\rho_\theta}^M$ and the relative entropies $D\big(\rho_{\theta_0}\big\|\rho_{\theta_1}\big)$ and $D\big(\mathrm{P}_{\rho_{\theta_0}}^M\big\|\mathrm{P}_{\rho_{\theta_1}}^M\big)$ to P_θ^M, $D(\theta_0\|\theta_1)$ and $D^M(\theta_0\|\theta_1)$, respectively.

In the one-parameter quantum estimation, the estimator is described by a pair comprised of a POVM and a map from its dataset to the real number set $\mathbb{R}$. Since the POVM $M\circ T^{-1}$ takes values in the real number set $\mathbb{R}$, we can regard any estimator as a POVM taking values in the real number set $\mathbb{R}$. In order to evaluate MSE, Helstrom [13, 14] derived the SLD Cramér–Rao inequality as a quantum counterpart of the Cramér–Rao inequality (29). If an estimator M satisfies

$$\int_{\mathbb{R}} x\,\mathrm{Tr}\,\rho_\theta M(\mathrm{d}x)=\theta \qquad \forall\theta\in\Theta \tag{26}$$

it is called unbiased. If $\theta - \theta_0$ is sufficiently small, we can obtain the following approximation in the neighbourhood of θ_0:

$$\int_{\mathbb{R}} x \operatorname{Tr} \rho_{\theta_0} M(\mathrm{d}x) + \left(\int_{\mathbb{R}} x \operatorname{Tr} \left. \frac{\partial \rho_\theta}{\partial \theta} \right|_{\theta=\theta_0} M(\mathrm{d}x) \right) (\theta - \theta_0) \cong \theta_0 + (\theta - \theta_0).$$

It implies the following two conditions:

$$\int_{\mathbb{R}} x \operatorname{Tr} \left. \frac{\partial \rho_\theta}{\partial \theta} \right|_{\theta=\theta_0} M(\mathrm{d}x) = 1 \tag{27}$$

$$\int_{\mathbb{R}} x \operatorname{Tr} \rho_{\theta_0} M(\mathrm{d}x) = \theta_0. \tag{28}$$

If an estimator M satisfies (27) and (28), it is called locally unbiased at θ_0. For any locally unbiased estimator M (at θ), the inequality, which is called the SLD Cramér–Rao inequality

$$\int_{\mathbb{R}} (x - \theta)^2 \operatorname{Tr} \rho_\theta M(\mathrm{d}x) \geqslant \frac{1}{J_\theta} \tag{29}$$

holds. Similar to the classical case, this inequality is derived from the Schwartz inequality with respect to the SLD Fisher information $\langle X|Y\rangle := \operatorname{Tr} \rho_\theta \frac{XY+YX}{2}$ [13–15].

The equality of (29) holds when the estimator is given by the spectral decomposition $E\left(\frac{L_\theta}{J_\theta} + \theta\right)$ of $\frac{L_\theta}{J_\theta} + \theta$, where L_θ is the SLD at θ and is defined by (3). This implies that the SLD Fisher information J_{θ_0} coincides with the Fisher information at θ_0 of the probability family $\left\{ \mathrm{P}_\theta^{E\left(\frac{L_{\theta_0}}{J_{\theta_0}} + \theta_0\right)} \middle| \theta \in \Theta \right\}$. The monotonicity of quantum relative entropy [29, 30] gives the following evaluation of the probability family $\left\{ \mathrm{P}_\theta^{E\left(\frac{L_{\theta_0}}{J_{\theta_0}} + \theta_0\right)} \middle| \theta \in \Theta \right\}$:

$$D^{E\left(\frac{L_{\theta_0}}{J_{\theta_0}} + \theta_0\right)}(\theta \| \theta_0) \leqslant D(\theta \| \theta_0).$$

Taking the limit $\theta \to \theta_0$, we have

$$J_\theta \leqslant \tilde{J}_\theta. \tag{30}$$

In this paper, we discuss inequality (30) from the viewpoint of the large deviation type of evaluation of the quantum estimation. The following families are treated as simple examples of the one-parameter quantum state family, in the latter.

Example 1 (One-parameter equatorial spin 1/2 system state family)**.**

$$\mathcal{S}_r := \left\{ \rho_\theta := \frac{1}{2} \begin{pmatrix} 1 + r\cos\theta & r\sin\theta \\ r\sin\theta & 1 - r\cos\theta \end{pmatrix} \middle| 0 \leqslant \theta < 2\pi \right\}.$$

In this family, we calculate

$$D(\rho_\theta \| \rho_0) = \frac{r}{2}(1 - \cos\theta) \log \frac{1+r}{1-r} \qquad \tilde{J}_\theta = \frac{r}{2} \log \frac{1+r}{1-r} \qquad J_\theta = r^2.$$

Since the relations $\tilde{J}_\theta = \infty$ and $J_\theta = 1$ hold in the case of $r = 1$, the two quantum analogues are completely different.

Example 2 (One-parameter quantum Gaussian state family and half-line quantum Gaussian state family)**.** We define the boson coherent vector $|\alpha\rangle := \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$, where $|n\rangle$ is the number vector on $L^2(\mathbb{R})$. The quantum Gaussian state is defined as

$$\rho_\theta := \frac{1}{\pi \overline{N}} \int_{\mathbb{C}} |\alpha\rangle\langle\alpha| \, \mathrm{e}^{-\frac{|\alpha-\theta|^2}{\overline{N}}} \mathrm{d}^2\alpha \qquad \forall \theta \in \mathbb{C}.$$

We call $\{\rho_\theta|\theta \in \mathbb{R}\}$ the *one-parameter quantum Gaussian state family* and call $\{\rho_\theta|\theta \geqslant 0\,(\theta \in \mathbb{R}^+ = [0,\infty))\}$ the *half-line quantum Gaussian state family*. In this family, we can calculate

$$D\left(\rho_\theta \middle\| \rho_{\theta_0}\right) = \log\left(1+\frac{1}{\overline{N}}\right)|\theta-\theta_0|^2 \qquad \tilde{J}_\theta = 2\log\left(1+\frac{1}{\overline{N}}\right) \qquad J_\theta = \frac{2}{\overline{N}+\frac{1}{2}}.$$

5. The bound under the weak consistency condition

We introduce the quantum independent-identical density (i.i.d.) condition in order to treat an asymptotic setting. Suppose that n-independent physical systems are prepared in the same state ρ. Then, the quantum state of the composite system is described by

$$\rho^{\otimes n} := \underbrace{\rho \otimes \cdots \otimes \rho}_{n} \text{ on } \mathcal{H}^{\otimes n}$$

where the tensored space $\mathcal{H}^{\otimes n}$ is defined by

$$\mathcal{H}^{\otimes n} := \underbrace{\mathcal{H} \otimes \cdots \otimes \mathcal{H}}_{n}.$$

We call this condition the quantum i.i.d. condition, which is a quantum analogue of the independent-identical distribution condition. In this setting, any estimator is described by a POVM M^n on $\mathcal{H}^{\otimes n}$, whose dataset is $\mathbb{R}$. In this paper, we simplify $\mathrm{P}_{\rho_\theta^{\otimes n}}^{M^n}$ and $D\left(\mathrm{P}_{\rho_{\theta_0}^{\otimes n}}^{M^n} \middle\| \mathrm{P}_{\rho_{\theta_1}^{\otimes n}}^{M^n}\right)$ to $\mathrm{P}_\theta^{M^n}$ and $D^{M^n}(\theta_0\|\theta_1)$. The notation $M \times n$ denotes the POVM in which we perform the POVM M for the respective n systems.

Definition 1 (Weak consistency condition)**.** *A sequence of estimators $\vec{M} := \{M^n\}_{n=1}^\infty$ is called* weakly consistent *if*

$$\lim_{n\to\infty} \mathrm{P}_\theta^{M^n}\{|\hat{\theta}-\theta| > \epsilon\} = 0 \qquad \forall\theta \in \Theta \quad \forall\epsilon > 0 \tag{31}$$

where $\hat{\theta}$ is the estimated value.

This definition means that the estimated value $\hat{\theta}$ converges to the true value θ in probability, and can be regarded as the quantum extension of (8).

Now, we focus on the exponential component of the tail probability as follows:

$$\beta(\vec{M},\theta,\epsilon) := \limsup_{n\to\infty} \frac{-1}{n}\log \mathrm{P}_\theta^{M^n}\{|\hat{\theta}-\theta| > \epsilon\}.$$

We usually discuss the following value instead of $\beta(\vec{M},\theta,\epsilon)$:

$$\alpha(\vec{M},\theta) := \limsup_{\epsilon\to 0} \frac{1}{\epsilon^2}\beta(\vec{M},\theta,\epsilon) \tag{32}$$

because it is too difficult to discuss $\beta(\vec{M},\theta,\epsilon)$. The following theorem can be proved from the monotonicity of the quantum relative entropy.

Theorem 2. (Nagaoka [22, 23]) *If a POVM M^n on $\mathcal{H}^{\otimes n}$ satisfies the weakly consistent condition (31), the inequalities*

$$\beta(\vec{M},\theta,\epsilon) \leqslant \inf\{D(\rho_{\theta'}\|\rho_\theta)||\theta-\theta'| < \epsilon\} \tag{33}$$

$$\alpha(\vec{M},\theta) \leqslant \frac{\tilde{J}_\theta}{2} \tag{34}$$

hold.

Even if the parameter set Θ is not open (e.g., the closed half-line $\mathbb{R}^+ := [0, \infty)$), this theorem holds.

Proof. The monotonicity of the quantum relative entropy yields the inequality

$$D\left(\rho_{\theta'}^{\otimes n} \middle\| \rho_{\theta}^{\otimes n}\right) \geqslant p_{n,\theta'} \log \frac{p_{n,\theta'}}{p_{n,\theta}} + (1 - p_{n,\theta'}) \log \frac{1 - p_{n,\theta'}}{1 - p_{n,\theta}}$$

for any θ' satisfying $|\theta' - \theta| > \epsilon$, where we denote the probability $\mathrm{P}_{\theta''}^{M^n}\{|\hat{\theta} - \theta| > \epsilon\}$ by $p_{n,\theta''}$. Using the inequality $-(1 - p_{n,\theta'}) \log(1 - p_{n,\theta}) \geqslant 0$, we have

$$-\frac{\log \mathrm{P}_{\theta}^{M^n}\{|\hat{\theta} - \theta| > \epsilon\}}{n} = -\frac{\log p_{n,\theta}}{n} \leqslant \frac{D\left(\rho_{\theta'}^{\otimes n} \middle\| \rho_{\theta}^{\otimes n}\right) + h(p_{n,\theta'})}{n p_{n,\theta'}} \tag{35}$$

where h is the binary entropy defined by $h(x) := -x \log x - (1 - x) \log(1 - x)$. Since the assumption guarantees that $p_{n,\theta'} \to 1$, the inequality

$$\beta(\vec{M}, \theta, \epsilon) \leqslant D(\rho_{\theta'} \| \rho_{\theta}) \tag{36}$$

holds, where we use the additivity of quantum relative entropy:

$$D\left(\rho_{\theta'}^{\otimes n} \middle\| \rho_{\theta}^{\otimes n}\right) = n D(\rho_{\theta'} \| \rho_{\theta}).$$

Thus, we obtain (33). Taking the limit $\epsilon \to 0$ in inequality (36), we obtain (34). □

As another proof, we can prove this inequality as a corollary of the quantum Stein's lemma [10, 11].

6. The bound under the strong consistency condition

As discussed in section 4, the SLD Cramér–Rao inequality guarantees that the lower bound of MSE is given by the SLD Fisher information. Therefore, it is expected that the bound is connected with the SLD Fisher information for large deviation. In order to discuss the relationship between the SLD Fisher information and the bound for large deviation, we need another characterization with respect to the limit of the tail probability. We thus define

$$\underline{\beta}(\vec{M}, \theta, \epsilon) := \liminf_{n \to \infty} \frac{-1}{n} \log \mathrm{P}_{\theta}^{M^n}\{|\hat{\theta} - \theta| > \epsilon\} \qquad \underline{\alpha}(\vec{M}, \theta) := \liminf_{\epsilon \to 0} \frac{1}{\epsilon^2} \underline{\beta}(\vec{M}, \theta, \epsilon). \tag{37}$$

In the following, we attempt to link the quantity $\underline{\alpha}(\vec{M}, \theta)$ with the SLD Fisher information. For this purpose, it is suitable to focus on an information quantity that satisfies the additivity and the monotonicity, as in the proof of theorem 1. Its limit should be the SLD Fisher information. The Bures distance $b(\rho, \sigma) := \sqrt{2(1 - \mathrm{Tr}\,|\sqrt{\rho}\sqrt{\sigma}|)} = \sqrt{\min_{U:\text{unitary}} \mathrm{Tr}(\sqrt{\rho} - \sqrt{\sigma}U)(\sqrt{\rho} - \sqrt{\sigma}U)^*}$ is known to be an information quantity whose limit is the SLD Fisher information, as mentioned in lemma 3. Of course, it can be regarded as a quantum analogue of the Hellinger distance, and satisfies the monotonicity.

Lemma 3 (Uhlmann [31], Matsumoto [32])**.** *If there exists an SLD L_θ satisfying (3), then the equation*

$$\frac{1}{4} J_\theta = \lim_{\epsilon \to 0} \frac{b^2(\rho_\theta, \rho_{\theta+\epsilon})}{\epsilon^2} \tag{38}$$

holds.

A proof of lemma 3 is given in appendix C. As discussed in the latter, the Bures distance satisfies the monotonicity. Unfortunately, the Bures distance does not satisfy the additivity.

However, the quantum affinity $I(\rho\|\sigma) := -8\log \mathrm{Tr}\,|\sqrt{\rho}\sqrt{\sigma}| = -8\log\left(1-\frac{1}{2}b(\rho,\sigma)^2\right)$ satisfies the additivity:

$$I(\rho^{\otimes n}\|\sigma^{\otimes n}) = nI(\rho\|\sigma). \tag{39}$$

Its classical version is called affinity in the following form [33]:

$$I(p\|q) = -8\log\left(\sum_i \sqrt{p_i}\sqrt{q_i}\right). \tag{40}$$

As a trivial deformation of (38), the equation

$$\lim_{\epsilon\to 0}\frac{I(\rho_\theta\|\rho_{\theta+\epsilon})}{\epsilon^2} = J_\theta \tag{41}$$

holds. The quantum affinity satisfies the monotonicity w.r.t. any measurement M (Jozsa [34], Fuchs [35]):

$$I(\rho\|\sigma) \geqslant I\left(\mathrm{P}_\rho^M \,\|\, \mathrm{P}_\sigma^M\right) = -8\log\sum_\omega\left(\sqrt{\mathrm{P}_\rho^M(\omega)}\sqrt{\mathrm{P}_\sigma^M(\omega)}\right). \tag{42}$$

The most simple proof of (42) is given by Fuchs [35] who directly proved that

$$\mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \leqslant \sum_\omega\left(\sqrt{\mathrm{P}_\rho^M(\omega)}\sqrt{\mathrm{P}_\sigma^M(\omega)}\right). \tag{43}$$

For reader's convenience, a proof of (43) is given in appendix D. From (39), (41) and (42), we can expect that the SLD Fisher information is, in a sense, closely related to a large deviation type of bound. From the additivity and the monotonicity of the quantum affinity, we can show the following lemma.

Lemma 4. *The inequality*

$$4\inf_{\{s|1\geqslant s\geqslant 0\}}(\underline{\beta}'(\vec{M},\theta,s\delta)+\underline{\beta}'(\vec{M},\theta+\delta,(1-s)\delta)) \leqslant I(\rho_\theta\|\rho_{\theta+\delta}) \tag{44}$$

holds, where we define $\underline{\beta}'(\vec{M},\theta,\delta) := \lim_{\epsilon\to+0}\underline{\beta}(\vec{M},\theta,\delta-\epsilon)$.

A proof of lemma 4 is given in appendix E. However, lemma 4 cannot yield an inequality w.r.t. $\alpha(\vec{M},\theta)$ under the weak consistency condition, unlike inequality (36). Therefore, we consider a stronger condition, which is given in the following.

Definition 5 (Strong consistency condition)**.** *A sequence of estimators* $\vec{M} = \{M^n\}_{n=1}^\infty$ *is called* strongly consistent *if the convergence of (37) is uniform for the parameter* θ *and if* $\underline{\alpha}(\vec{M},\theta)$ *is continuous for* θ*. A sequence of estimators is called* strongly consistent at θ *if there exists a neighbourhood* U *of* θ *such that it is strongly consistent in* U*.*

The square root n consistency is familiar in the field of mathematical statistics. However, in the large deviation setting, this strong consistency seems more suitable than the square root n consistency.

As a corollary of lemma 4, we have the following theorem.

Theorem 6. *Assume that there exists the SLD* L_θ *satisfying (3). If a sequence of estimators* $\vec{M} = \{M^n\}_{n=1}^\infty$ *is strongly consistent at* θ*, then the inequality*

$$\underline{\alpha}(\vec{M},\theta) \leqslant \frac{J_\theta}{2} \tag{45}$$

holds.

Proof. From the above assumption, for any real $\epsilon > 0$ and any element $\theta \in \Theta$, there exists a sufficiently small real $\delta > 0$ such that $(\underline{\alpha}(\vec{M}, \theta) - \epsilon)\epsilon'^2 \leqslant \underline{\beta}'(\vec{M}, \theta, \epsilon'), \underline{\beta}'(\vec{M}, \theta + \delta, \epsilon')$ for $\forall \epsilon' < \delta$. Therefore, inequality (44) yields the relations

$$2(\underline{\alpha}(\vec{M}, \theta) - \epsilon)\delta^2 = 4(\underline{\alpha}(\vec{M}, \theta) - \epsilon) \inf_{\{s|1 \geqslant s \geqslant 0\}} (s^2\delta^2 + (1-s)^2\delta^2)$$
$$\leqslant 4 \inf_{\{s|1 \geqslant s \geqslant 0\}} (\underline{\beta}'(\vec{M}, \theta, s\delta) + \underline{\beta}'(\vec{M}, \theta + \delta, (1-s)\delta)) \leqslant I(\rho_\theta \| \rho_{\theta+\delta}). \tag{46}$$

Lemma 3 and (46) guarantee (45) for $\forall \theta \in \Theta$. □

Remark 2. Inequality (43) can be regarded as a special case of the monotonicity w.r.t. any trace-preserving completely positive (CP) map $C : \mathcal{S}(\mathcal{H}_1) \to \mathcal{S}(\mathcal{H}_2)$:

$$(\mathrm{Tr}\,|\sqrt{\rho}\sqrt{\sigma}|)^2 \leqslant (\mathrm{Tr}\,|\sqrt{C(\rho)}\sqrt{C(\sigma)}|)^2 \tag{47}$$

which is proved by Jozsa [34] because the map $\rho \mapsto \mathrm{P}_\rho^M$ can be regarded as a trace-preserving CP map from the C^* algebra of bounded operators on $\mathcal{H}$ to the commutative C^* algebra $C(\Omega)$, where Ω is the dataset.

7. Achievabilities of the bounds

Next, we discuss the achievabilities of the two bounds $\tilde{J}_\theta$ and J_θ in their respective senses. In this section, we discuss the achievabilities in two cases: the first case is the one-parameter quantum Gaussian state family and the second case is an arbitrary one-parameter finite-dimensional quantum state family that satisfies some assumptions.

7.1. One-parameter quantum Gaussian state family

In this subsection, we discuss the achievabilities in the one-parameter quantum Gaussian state family.

Theorem 7. *In the one-parameter quantum Gaussian state family, the sequence of estimators $\vec{M}^s = \{M^{s,n}\}_{n=1}^\infty$ (defined in the following) satisfies the strong consistency condition and the relations*

$$\alpha(\vec{M}^s, \theta) = \underline{\alpha}(\vec{M}^s, \theta) = \frac{J_\theta}{2} = \frac{1}{\overline{N} + \frac{1}{2}}. \tag{48}$$

Construction of $\vec{M}^s$. We perform the POVM $E(Q)$ for all systems, where Q is the position operator on $L^2(\mathbb{R})$. The estimated value ξ_n is determined to be the mean value of n data.

Proof. Since the equation

$$\mathrm{P}_{|\alpha\rangle\langle\alpha|}^{E(Q)}(\mathrm{d}x) = \sqrt{\frac{2}{\pi}}\,\mathrm{e}^{-2(x-\alpha_x)^2}\mathrm{d}x$$

holds, we have the equation

$$\mathrm{P}_\theta^{E(Q)}(\mathrm{d}x) = \left(\mathrm{P}_{\rho_\theta}^{E(Q)}(\mathrm{d}x)\right) = \frac{1}{\pi N}\int_{\mathbb{C}} \mathrm{P}_{|\alpha\rangle\langle\alpha|}^{E(Q)}(\mathrm{d}x)\,\mathrm{e}^{-\frac{|\alpha-\theta|^2}{N}}\,\mathrm{d}^2\alpha = \sqrt{\frac{2}{\pi(2\overline{N}+1)}}\,\mathrm{e}^{-\frac{2(x-\theta)^2}{2\overline{N}+1}}\mathrm{d}x.$$

Thus, we obtain the equation

$$\mathrm{P}_\theta^{M^{s,n}}(\mathrm{d}\xi_n) = \sqrt{\frac{2}{\pi(2\overline{N}+1)n}}\,\mathrm{e}^{-\frac{2(\xi_n-\theta)^2}{(2\overline{N}+1)n}}\mathrm{d}\xi_n$$

which implies that

$$\beta(\vec{M}^s, \theta, \epsilon) = \lim \frac{-1}{n} \log \mathrm{P}_\theta^{M^{s,n}}\{|\xi_n - \theta| > \epsilon\} = \frac{\epsilon^2}{\overline{N} + \frac{1}{2}}. \tag{49}$$

Therefore, the sequence of estimators $\vec{M}^s = \{M^{s,n}\}_{n=1}^\infty$ attains the bound $\frac{J_\theta}{2}$ and satisfies the strong consistency condition. □

Proposition 8. *In the half-line quantum Gaussian state family, the sequence of estimators $\vec{M}^w = \{M^{w,n}\}_{n=0}^\infty$ (defined in the following) satisfies the weak consistency condition and the strong consistency condition at $\mathbb{R}^+\backslash\{0\}$ and the relations*

$$\underline{\alpha}(\vec{M}^w, 0) = \alpha(\vec{M}^w, 0) = \frac{\tilde{J}_0}{2} = \log\left(1 + \frac{1}{\overline{N}}\right) \tag{50}$$

$$\underline{\alpha}(\vec{M}^w, \theta) = \alpha(\vec{M}^w, \theta) = \frac{J_\theta}{2} = \frac{1}{\overline{N} + \frac{1}{2}} \qquad \forall \theta \in \mathbb{R}^+\backslash\{0\}. \tag{51}$$

This proposition indicates the significance of the uniformity of the convergence of (37). This proposition is proved in appendix G.

Construction of $\vec{M}^w$. We perform the following unitary evolution:

$$\rho_\theta^{\otimes n} \mapsto \rho_{\sqrt{n}\theta} \otimes \rho_0^{\otimes(n-1)}.$$

For detail, see appendix F. We perform the number measurement $E(N)$ of the first system whose state is $\rho_{\sqrt{n}\theta}$, and let k be its data, where the number operator N is defined as $N := \sum_n n|n\rangle\langle n|$. The estimated value T_n is determined by $T_n := \sqrt{\frac{k}{n}}$.

Theorem 9. *In the one-parameter quantum Gaussian state family, for any $\theta \in \mathbb{R}$, the sequence of estimators $\vec{M}^w_{\theta_1} = \left\{M^{w,n}_{\theta_1}\right\}_{n=1}^\infty$ (defined in the following) satisfies the weak consistency condition and the relations*

$$\underline{\alpha}\left(\vec{M}^w_{\theta_1}, \theta_1\right) = \alpha\left(\vec{M}^w_{\theta_1}, \theta_1\right) = \frac{\tilde{J}_\theta}{2} = \log\left(1 + \frac{1}{\overline{N}}\right). \tag{52}$$

Construction of $\vec{M}^w_{\theta_1}$. We divide n systems into two groups. One consists of $\sqrt{n}$ systems and the other of $n - \sqrt{n}$ systems. We perform the PVM $E(Q)$ for every system in the first group. Let $\xi_{\sqrt{n}}$ be the mean value in the first group, i.e. we perform the PVM $M^{s,\sqrt{n}}$ for the first system. At the second step, we perform the following unitary evolution for the second group:

$$\rho_\theta^{\otimes(n-\sqrt{n})} \mapsto \rho_{\theta-\theta_1}^{\otimes(n-\sqrt{n})}.$$

For details, see appendix F. We perform the POVM $M^{w,n-\sqrt{n}}$ for the system whose state is $\rho_{\theta-\theta_1}^{\otimes n-\sqrt{n}}$; the data are written as $T_{n-\sqrt{n}}$. Then, we decide the final estimated value $\hat{\theta}$ as

$$\hat{\theta} := \theta_1 + \mathrm{sgn}(\xi_{\sqrt{n}} - \theta_1) T_{n-\sqrt{n}}.$$

Proof. Since

$$\mathrm{P}_{\theta_1}^{M^{w,n}_{\theta_1}}\{|\hat{\theta} - \theta_1| > \epsilon\} = \mathrm{P}_0^{M^{w,n-\sqrt{n}}}\{|T_{n-\sqrt{n}}| > \epsilon\}$$

we have

$$\beta\left(\vec{M}^{w}_{\theta_1},\theta_1\right)=\lim\frac{-1}{n}\log \mathrm{P}_{\theta_1}^{M^{w,n}_{\theta_1}}\{|\hat{\theta}-\theta_1|>\epsilon\}$$
$$=\lim\frac{n-\sqrt{n}}{n}\frac{-1}{n-\sqrt{n}}\log \mathrm{P}_0^{M^{w,n-\sqrt{n}}}\{|T_{n-\sqrt{n}}|>\epsilon\}=\beta(\vec{M}^{w},0).$$

As shown in appendix G, we have

$$\beta(\vec{M}^{w},0)=\epsilon^2\log\left(1+\frac{1}{\overline{N}}\right)$$

which implies (52). Next, we prove the consistency in the case where $\theta>\theta_1$. In this case, it is sufficient to discuss the case where $\theta-\theta_1>\epsilon>0$. Since the first measurement $M^{s,\sqrt{n}}$ and the second one $M^{w,n-\sqrt{n}}$ are performed independently, we obtain

$$\mathrm{P}_{\theta}^{M^{w,n}_{\theta_1}}\{|\hat{\theta}-\theta_1|>\epsilon\}\leqslant \mathrm{P}_{\theta}^{M^{w,n-\sqrt{n}}}\{|T_{n-\sqrt{n}}-(\theta-\theta_1)|>\epsilon\}+\mathrm{P}_{\theta}^{M^{s,\sqrt{n}}}\{\xi_{\sqrt{n}}-\theta_1\leqslant 0\}.$$

Proposition 8 guarantees that the first term goes to 0 and theorem 7 guarantees that the second term goes to 0. Thus, we obtain the consistency of $\vec{M}^{w}_{\theta_1}$. Similarly, we can prove the weak consistency in the case where $\theta<\theta_1$. □

7.2. Finite-dimensional family

In this subsection, we treat the case where the dimension of the Hilbert space $\mathcal{H}$ is k (finite). As for the achievability of inequality (45), we have the following lemma.

Lemma 10. *Let θ_0 be fixed in Θ. Under assumptions 1 and 2, the sequence of estimators $\vec{M}^{s}_{\theta_0}$ (defined in the following) satisfies the strong consistency condition at θ_0 (defined in definition 5) and the relation*

$$\alpha\left(\vec{M}^{s}_{\theta_0},\theta_0\right)=\underline{\alpha}\left(\vec{M}^{s}_{\theta_0},\theta_0\right)=\frac{J_{\theta_0}}{2}. \tag{53}$$

Assumption 1. The map $\theta\mapsto\rho_\theta$ is C^1 and $\rho_\theta>0$.

Assumption 2. The map $\theta\mapsto \mathrm{Tr}\,\rho_\theta\frac{L_{\theta_0}}{J_{\theta_0}}$ is injective, i.e. one-to-one.

Construction of $\vec{M}^{s}_{\theta_0}$. We perform the POVM $E\left(\frac{L_{\theta_0}}{J_{\theta_0}}\right)$ for all systems. The estimated value is determined to be the mean value plus θ_0.

Proof of lemma 10. From assumption 2, the weak consistency is satisfied. Let $\delta>0$ be a sufficiently small number. Define the function

$$\phi_{\theta,\theta_0}(s):=\mathrm{Tr}\,\rho_\theta\exp\left(s\left(\frac{L_{\theta_0}}{J_{\theta_0}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\theta_0}}{J_{\theta_0}}\right)\right). \tag{54}$$

Since $\left\|\frac{L_{\theta_0}}{J_{\theta_0}}\right\|<\infty$ and $\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\theta_0}}{J_{\theta_0}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\theta_0}}{J_{\theta_0}}\right)=0$, we have

$$\lim_{s\to 0}\frac{\phi_{\theta,\theta_0}(s)-1}{s^2}=\frac{1}{2}\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\theta_0}}{J_{\theta_0}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\theta_0}}{J_{\theta_0}}\right)^2.$$

When $\|\theta-\theta_0\|$ is sufficiently small, the function $x\to\sup_s(xs-\log\phi_{\theta,\theta_0}(s))$ is continuous in $(-\delta,\delta)$. Using Cramér's theorem [36], we have

$$\lim_{n\to\infty}\frac{-1}{n}\log \mathrm{P}_{\theta}^{M^{s,n}_{\theta_0}}\{|\hat{\theta}-\theta_0|>\epsilon\}=\min\left\{\sup_s(\epsilon s-\log\phi_{\theta,\theta_0}(s)),\sup_{s'}(-\epsilon s'-\log\phi_{\theta,\theta_0}(s'))\right\}$$

for $\epsilon < \delta$. Taking the limit $\epsilon \to 0$, we have

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{-1}{\epsilon^2 n}\mathrm{P}_{\theta_0}^{M^{s,n}_{\theta_0}}\{|\hat{\theta}-\theta_0|>\epsilon\}$$
$$= \min\left\{\lim_{\epsilon\to 0}\frac{\sup_s(\epsilon s-\log\phi_{\theta,\theta_0}(s))}{\epsilon^2}, \lim_{\epsilon\to 0}\frac{\sup_s(-\epsilon s-\log\phi_{\theta,\theta_0}(s))}{\epsilon^2}\right\} = \frac{1}{2}c^{-1}_{\theta,\theta_0}$$

where

$$c_{\theta,\theta_0} := \mathrm{Tr}\,\rho_\theta\left(\frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\mathrm{Tr}\,\rho_\theta L_{\theta_0}}{J_{\theta_0}}\right)^2$$

because

$$\epsilon s-\log\phi_{\theta,\theta_0}(s)\cong\epsilon s-\log\left(1+\frac{1}{2}c_{\theta,\theta_0}s^2\right)\cong\epsilon s-\frac{1}{2}c_{\theta,\theta_0}s^2=-\frac{c_{\theta,\theta_0}}{2}\left(s-\frac{\epsilon}{c_{\theta,\theta_0}}\right)+\frac{\epsilon^2}{2c_{\theta,\theta_0}}.$$

The above convergence is uniform for the neighbourhood of θ_0. Taking the limit $\theta\to\theta_0$, we have

$$\lim_{\theta\to\theta_0}\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\theta_0}}{J_{\theta_0}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\theta_0}}{J_{\theta_0}}\right)^2 = J^{-1}_{\theta_0} = \mathrm{Tr}\,\rho_{\theta_0}\left(\frac{L_{\theta_0}}{J_{\theta_0}}-\frac{\mathrm{Tr}\,\rho_{\theta_0}L_{\theta_0}}{J_{\theta_0}}\right)^2.$$

Thus, we can check (53) and the strong consistency in the neighbourhood of θ_0. □

However, this sequence of estimators $\vec{M}^s_\delta$ depends on the true parameter θ_0. We should construct a sequence of estimators that satisfies the strong consistency condition and attains the bound $\frac{J_{\theta_0}}{2}$ at all points θ_0. Since such a construction is too difficult, we introduce another strong consistency condition that is weaker than the above and under which inequality (45) holds. We construct a sequence of estimators that satisfies this strong consistency condition and attains the bound given in (45) for all θ in a weak sense.

Second strong consistency condition. A sequence of estimators $\vec{M} = \{M^n\}$ is called second strongly consistent if there exists a sequence of functions $\{\underline{\beta}_m(\vec{M},\theta,\epsilon)\}_{m=1}^\infty$ such that

- $\lim_{m\to\infty}\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}_m(\vec{M},\theta,\epsilon) = \underline{\alpha}(\vec{M},\theta)$.
- $\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}_m(\vec{M},\theta,\epsilon) \leqslant \underline{\alpha}(\vec{M},\theta)$ holds. Its lhs converges locally uniformly to θ.
- $\forall m, \exists\delta>0$ s.t. $\underline{\beta}(\vec{M},\theta,\epsilon) \geqslant \underline{\beta}_m(\vec{M},\theta,\epsilon)$, for $\delta > \forall\epsilon > 0$.

Similar to theorem 2, we can prove inequality (45) under the second strong consistency condition.

Under these preparations, we state a theorem with respect to the attainability of the bound J_θ. The following theorem can be regarded as a special case of theorem 8 of [37].

Theorem 11. *Under assumptions 1 and 3, the sequence of estimators $\vec{M}^s_\delta = \left\{M^{s,n}_\delta\right\}_{n=1}^\infty$ (defined in the following) satisfies the second strong consistency condition and the relations*

$$\alpha\left(\vec{M}^s_\delta,\theta\right) = \underline{\alpha}\left(\vec{M}^s_\delta,\theta\right) = (1-\delta)\frac{J_\theta}{2}. \tag{55}$$

The sequence of estimators $\vec{M}^s_\delta$ is independent of the unknown parameter θ. Every $M^{s,n}_\delta$ is an adaptive estimator and will be defined in section 8.

Its proof is given in appendix H.

Assumption 3. The following set is compact:

$$\left\{ \left(\mathrm{Tr}\, \rho_\theta \left(\frac{L_{\check\theta}}{J_{\check\theta}} - \frac{\mathrm{Tr}\, \rho_\theta L_{\check\theta}}{J_{\check\theta}} \right)^2 \right)^{-1}, \mathrm{Tr}\, \rho_\theta \left(\frac{L_{\check\theta}}{J_{\check\theta}} - \frac{\mathrm{Tr}\, \rho_\theta L_{\check\theta}}{J_{\check\theta}} \right)^2 \middle| \forall \theta, \check\theta \in \Theta \right\}.$$

If the state family is included by a bounded closed set consisting of positive definite operators, assumption 3 is satisfied.

Construction of $\vec{M}^s_\delta$. We perform a faithful POVM M_f (defined in the following) for the first δn systems. Then, the data $(\omega_1, \ldots, \omega_{\delta n})$ obey the probability family $\{\mathrm{P}^{M_f}_\theta | \theta \in \Theta\}$. We denote the maximum likelihood estimator (MLE) w.r.t. the data $(\omega_1, \ldots, \omega_{\delta n})$ by $\check\theta$. Next, we perform the measurement $E(L_{\check\theta})$ defined by the spectral measure of $L_{\check\theta}$ for other $(1-\delta)n$ systems. Then, we have the data $(\omega_{\delta n+1}, \ldots, \omega_n)$. We decide the final estimated value $T^n_{\check\theta}$ as

$$\mathrm{Tr}\, \rho_{T^n_{\check\theta}} L_{\check\theta} = \frac{1}{(1-\delta)n} \sum_{i=\delta n+1}^{n} \omega_i.$$

Definition 12. *A POVM M is called* faithful, *if the map $\rho \in \mathcal{S}(\mathcal{H}) \mapsto \mathrm{P}^M_\rho$ is one-to-one.*

An example of faithful POVM, which is a POVM taking values in the set of pure states on $\mathcal{H}$, is given by $M_h(\mathrm{d}\rho) := k\rho\nu(\mathrm{d}\rho)$, where ν is the invariant (w.r.t. the action of $SU(\mathcal{H})$) probability measure on the set of pure states on $\mathcal{H}$. As another example, if $L_1, \ldots, L_{k^2-1}$ is a basis of the space of self-adjoint traceless operators, a disjoint random combination of PVMs $E(L_1), \ldots, E(L_{k^2-1})$ is faithful. Note that a disjoint random combination is defined in section 4.

Remark 3. By dividing n systems into $\sqrt{n}$ and $n - \sqrt{n}$ systems, Gill and Massar [16] constructed an estimator which asymptotically attains the optimal bound w.r.t. MSE, and Hayashi and Matsumoto [38] constructed a similar estimator by dividing them into b_n and $n - b_n$ systems, where $\lim \frac{b_n}{n} = 0$. However, in our proof, it is difficult to show the attainability of the bound (45) in such a division. Perhaps, there may exist a family in which such an estimator does not attain the bound (45). At least, it is essential in our proof that the number of the first group b_n satisfies $\lim \frac{b_n}{n} > 0$.

Conversely, as is mentioned in theorems 9 and 13, by dividing n systems into $\sqrt{n}$ and $n - \sqrt{n}$ systems, we can construct an estimator attaining the bound (34) at one point.

We must use quantum correlations in the quantum apparatus to achieve the bound $\frac{\tilde J_\theta}{2}$. The following theorem can easily be extended to the multi-parameter case.

Theorem 13. *We assume assumption 1 and that $D\left(\rho_{\theta'} \middle\| \rho_{\theta_1}\right) < \infty$ for $\forall \theta_1, \forall \theta' \in \Theta$. Then, for any $\theta_1 \in \Theta$, the sequence of estimators $\vec{M}^w_{\theta_1} = \left\{M^{w,n}_{\theta_1}\right\}_{n=1}^\infty$ satisfies the weak consistency condition (31), and the equations*

$$\underline{\beta}\left(\vec{M}^w_{\theta_1}, \theta_1, \epsilon\right) = \beta\left(\vec{M}^w_{\theta_1}, \theta_1, \epsilon\right) = \inf_{\theta' \in \Theta} \left\{ D\left(\rho_{\theta'} \middle\| \rho_{\theta_1}\right) \middle| |\theta_1 - \theta'| > \epsilon \right\} \tag{56}$$

$$\underline{\alpha}\left(\vec{M}^w_{\theta_1}, \theta_1\right) = \alpha\left(\vec{M}^w_{\theta_1}, \theta_1\right) = \frac{\tilde J_{\theta_1}}{2}. \tag{57}$$

The sequence of estimators $\vec{M}^w_{\theta_1}$ depends on the unknown parameter θ_1 but not on $\epsilon > 0$.

Its proof is given in appendix I. In the following construction, $M^{w,n}_{\theta_1}$ is constructed from the PVM $E^n_{\theta_1}$, which is defined from a group-theoretical viewpoint in definition 29 in appendix J.3.

Construction of $M^{w,n}_{\theta_1}$. We divide the n systems into two groups. We perform a faithful POVM M_f for the first group of $\sqrt{n}$ systems. Then, the data $(\omega_1, \ldots, \omega_{\sqrt{n}})$ obey the probability $\mathrm{P}_\theta^{M_f}$. We let $\check{\theta}$ be the MLE of the data $(\omega_1, \ldots, \omega_{\sqrt{n}})$ under the probability family $\{\mathrm{P}_\theta^{M_f} | \theta \in \Theta\}$. Next, we perform the correlational PVM $E_{\theta_1}^{n-\sqrt{n}}$ for the composite system which consists of the other group of $n - \sqrt{n}$ systems. Then, the data ω obey the probability $\mathrm{P}_\theta^{E_{\theta_1}^{n-\sqrt{n}}}$. If $\mathrm{e}^{n(1-\delta_{n-\sqrt{n}})D(\rho_{\check{\theta}}\|\rho_{\theta_1})}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega) \geqslant \mathrm{P}_{\check{\theta}}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)$, the estimated value T_n is decided to be θ_1, where $\delta_n := \frac{1}{n^{\frac{1}{5}}}$. If not, T_n is decided to be $\check{\theta}$.

The following lemma proved in appendix J plays an important role in the proof of theorem 13.

Lemma 14. *For three parameters θ_0, θ_1 and θ_2 and $\delta > 0$, the inequalities*

$$\mathrm{P}_{\theta_0}^{E_{\theta_1}^n}\left\{-\frac{1}{n}\log \mathrm{P}_{\theta_2}^{E_{\theta_1}^n}(\omega) + \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_2} \geqslant \delta\right\} \leqslant \exp -n\left(\sup_{0\leqslant t\leqslant 1}\big(\delta - \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_2}\big)t - t\frac{(k+1)\log(n+1)}{n} - \log \mathrm{Tr}\,\rho_{\theta_0}{\rho_{\theta_2}}^{-t}\right) \tag{58}$$

$$\mathrm{P}_{\theta_0}^{E_{\theta_1}^n}\left\{\frac{1}{n}\log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega) - \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_1} \geqslant \delta\right\} \leqslant \exp -n\left(\sup_{0\leqslant t}\big(\delta + \mathrm{Tr}\,\rho_{\theta_0}\log\rho_{\theta_1}\big)t - \log \mathrm{Tr}\,\rho_{\theta_0}\rho_{\theta_1}^t\right) \tag{59}$$

hold.

We obtain the following theorem as a summary of the above discussion.

Theorem 15. *From theorems 2, 6 and 11 and lemma 10, we have the equations*

$$\sup_{\vec{M}:\mathrm{WC}} \limsup_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \sup_{\vec{M}:\mathrm{WC}} \liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \frac{\tilde{J}_\theta}{2} \tag{60}$$

$$\sup_{\vec{M}:\mathrm{SC\,at}\,\theta} \liminf_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}(\vec{M},\theta,\epsilon) = \frac{J_\theta}{2} \tag{61}$$

as an operational comparison of $\tilde{J}_\theta$ and J_θ under assumptions 1–3. We can replace $\beta(\vec{M},\theta,\epsilon)$ with $\underline{\beta}(\vec{M},\theta,\epsilon)$ in equations (60).

We can also prove (30) as a consequence of equations (60) and (61).

8. Adaptive estimators

In this section, we assume that the dimension of the Hilbert space $\mathcal{H}$ is finite. We consider estimators whose POVM is adaptively chosen from the data. We choose the lth POVM $M_l(\vec{\omega}_{l-1})$ on $\mathcal{H}$ from $l-1$ data $\vec{\omega}_{l-1} := (\omega_1, \ldots, \omega_{l-1})$. Its POVM M^n is described by

$$M^n(\vec{\omega}_n) := M_1(\omega_1) \otimes M_2(\vec{\omega}_1;\omega_2) \otimes \cdots \otimes M_n(\vec{\omega}_{n-1};\omega_n). \tag{62}$$

In this setting, the estimator is written as the pair $\mathcal{E}_n = (M^n, T_n)$ of the POVM M^n satisfying (62) and the function $T_n : \Omega^n \mapsto \Theta$. Such an estimator $\mathcal{E}_n$ is called an adaptive estimator.

As a larger class of POVMs, the separable POVM is well known. A POVM M^n on $\mathcal{H}^{\otimes n}$ is called separable if it is written as

$$M^n = \{M_1(\omega) \otimes \cdots \otimes M_n(\omega)\}_{\omega \in \Omega}$$

on $\mathcal{H}^{\otimes n}$, where $M_i(\omega)$ is a positive semi-definite operator on $\mathcal{H}$. For any separable estimator (M^n, T_n), the relations

$$\begin{aligned} D^{M^n}(\theta\|\theta') &= \sum_{\omega\in\Omega} \prod_{l'=1}^{n} \operatorname{Tr} \rho_\theta M_{l'}(\omega) \log \frac{\prod_{l=1}^n \operatorname{Tr} \rho_\theta M_l(\omega)}{\prod_{l=1}^n \operatorname{Tr} \rho_{\theta'} M_l(\omega)} \\ &= \sum_{\omega\in\Omega} \prod_{l'=1}^{n} \operatorname{Tr} \rho_\theta M_{l'}(\omega) \sum_{l=1}^{n} \log \frac{\operatorname{Tr} \rho_\theta M_l(\omega)}{\operatorname{Tr} \rho_{\theta'} M_l(\omega)} \\ &= \sum_{l=1}^{n} \sum_{\omega\in\Omega} a_{\theta,l}(\omega) \operatorname{Tr} \rho_\theta M_l(\omega) \log \frac{a_{\theta,l}(\omega) \operatorname{Tr} \rho_\theta M_l(\omega)}{a_{\theta,l}(\omega) \operatorname{Tr} \rho_{\theta'} M_l(\omega)} \\ &= \sum_{l=1}^{n} D^{M_{\theta,l}}(\theta\|\theta') \leqslant n \sup_{M:\text{POVM on}\,\mathcal{H}} D^M(\theta\|\theta') \end{aligned} \tag{63}$$

hold, where the POVM $M_{\theta,l}$ on $\mathcal{H}$ is defined by

$$M_{\theta,l}(\omega) := a_{\theta,l}(\omega) M_l(\omega) \qquad a_{\theta,l}(\omega) := \left(\prod_{l' \neq l} \operatorname{Tr} \rho_\theta M_{l'}(\omega) \right).$$

Theorem 16. *If a sequence of separable estimators $\vec{M} = \{\mathcal{E}_n\} = \{(M^n, T_n)\}$ satisfies the weak consistency condition, the inequalities*

$$\beta(\vec{M}, \theta_1, \epsilon) \leqslant \inf_{|\theta-\theta_1|>\epsilon} \sup_{M:\text{POVM on}\,\mathcal{H}} D^M(\theta\|\theta_1) \tag{64}$$

$$\alpha(\vec{M}, \theta_1) \leqslant \frac{J_{\theta_1}}{2} \tag{65}$$

hold.

Proof. Similar to (35), the monotonicity of quantum relative entropy yields

$$-\frac{\log \mathrm{P}_{\theta_1}^{M^n}\{|T_n(\vec{\omega}_n) - \theta_1| > \epsilon\}}{n} \leqslant \frac{D^{M^n}(\theta\|\theta_1) + h(P_n)}{n P_n}$$

where $P_n := \mathrm{P}_\theta^{M^n}\{|T_n(\vec{\omega}_n) - \theta_1| > \epsilon\}$. From the weak consistency, we have $P_n \to 1$. Thus, we obtain (64) from (63). Since $\mathcal{H}$ is finite dimensional, the set of extremal points of POVMs is compact. Therefore, the convergence $\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} D^M(\theta_1 + \epsilon\|\theta_1)$ is uniform w.r.t. M. This implies that

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \sup_{M:\text{POVM on}\,\mathcal{H}} D^M(\theta_1 + \epsilon\|\theta_1) = \sup_{M:\text{POVM on}\,\mathcal{H}} \lim_{\epsilon\to 0} \frac{1}{\epsilon^2} D^M(\theta_1 + \epsilon\|\theta_1) = \frac{J_{\theta_1}}{2}. \tag{66}$$

The last equation is derived from (29). □

The preceding theorem holds for any adaptive estimator. As a simple extension, we can define an m-adaptive estimator that satisfies (62) when every $M_l(\vec{\omega}_{l-1})$ is a POVM on $\mathcal{H}^m$. As a corollary of theorem 16, we have the following.

Corollary 17. *If a sequence of m-adaptive estimators* $\vec{M} = \{\mathcal{E}_n\} = \{(M^n, T_n)\}$ *satisfies the weak consistency condition, then the inequalities*

$$\beta(\vec{M}, \theta_1, \epsilon) \leqslant \inf_{|\theta-\theta_1|>\epsilon} \sup_{M:\text{POVM on}\,\mathcal{H}^{\otimes m}} \frac{1}{m} D^M(\theta\|\theta_1) \tag{67}$$

$$\alpha(\vec{M}, \theta_1) \leqslant \frac{J_{\theta_1}}{2} \tag{68}$$

hold.

Now, we obtain the equation

$$\lim_{m\to\infty} \lim_{\epsilon\to 0} \sup_{\vec{M}:m-\text{AWC}} \frac{1}{\epsilon^2} \underline{\beta}(\vec{M}, \theta, \epsilon) = \frac{J_\theta}{2}. \tag{69}$$

The part of $\geqslant$ holds because an adaptive estimator attaining the bound is constructed in theorem 11, and the part of $\leqslant$ follows from (67) and the equation

$$\lim_{\epsilon\to 0} \sup_{M:\text{POVM on}\,\mathcal{H}^{\otimes m}} \frac{1}{\epsilon^2 m} D^M(\theta_1+\epsilon\|\theta_1) = \sup_{M:\text{POVM on}\,\mathcal{H}^{\otimes m}} \lim_{\epsilon\to 0} \frac{1}{\epsilon^2 m} D^M(\theta_1+\epsilon\|\theta_1) = \frac{J_{\theta_1}}{2}$$

which is proved in a similar manner to (66).

9. Difference in order among limits and supremums

Theorem 15 yields another operational comparison as

$$\sup_{\vec{M}:\text{SC at}\,\theta} \liminf_{\epsilon\to 0} \frac{1}{\epsilon^2} \underline{\beta}(\vec{M}, \theta, \epsilon) = \frac{J_\theta}{2} \tag{70}$$

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \sup_{\vec{M}:\text{SC at}\,\theta} \underline{\beta}(\vec{M}, \theta, \epsilon) = \frac{\tilde{J}_\theta}{2}. \tag{71}$$

Equation (70) equals (61) and equation (71) follows from the theorem below. Therefore, the difference between $\frac{J_\theta}{2}$ and $\frac{\tilde{J}_\theta}{2}$ can be regarded as the difference in the order of $\liminf_{\epsilon\to 0}$ and $\sup_{\vec{M}:\text{SC}}$.

Theorem 18. *We adopt assumption 1 in theorem 11 and* $D\big(\rho_{\theta'}\big\|\rho_{\theta_1}\big) < \infty$ *for* $\forall\theta' \in \Theta$. *For any* $\delta > 0$, *there exists a sequence* $\vec{M}^{m,\delta}_{\theta_0} = \big\{M^{m,\delta,n}_{\theta_0}\big\}$ *of m-adaptive estimators satisfying the strong consistency condition and the inequality*

$$\lim_{n\to\infty} \frac{-1}{nm} \log \mathrm{P}^{M^{m,\delta,n}_{\theta_0}}_{\theta_0}\{|\hat{\theta}-\theta_0| > \epsilon\} \geqslant (1-\delta)\inf\{D(\theta\|\theta_0)||\theta-\theta_0| > \epsilon\} - \frac{(1-\delta)(k-1)\log(m+1)}{m}.$$

However, using theorem 18, we obtain a stronger equation than (71):

$$\lim_{\epsilon\to 0} \lim_{m\to\infty} \sup_{\vec{M}:m\text{-ASC at}\,\theta} \frac{1}{\epsilon^2} \beta(\vec{M}, \theta, \epsilon) = \frac{\tilde{J}_\theta}{2} \tag{72}$$

where m-ASC at θ denotes m-adaptive and is strongly consistent at θ. This equation is in contrast with (69). Of course, the part of $\leqslant$ for (72) follows from (67). The part of $\geqslant$ for (72) is derived from the above theorem.

The following two lemmas are essential for our proof of theorem 18.

Lemma 19. *For two parameters θ_1 and θ_0, the inequality*

$$m D(\theta_0\|\theta_1) - (k-1)\log(m+1) \leqslant D^{E^m_{\theta_1}}(\theta_0\|\theta_1) \leqslant m D(\theta_0\|\theta_1) \tag{73}$$

holds, where the PVM $E^m_{\theta_1}$ on $\mathcal{H}^{\otimes m}$ is defined in appendix J.3. It is independent of θ_0.

This lemma was proved by Hayashi [27] and can be regarded as an improvement of Hiai and Petz's result [10]. However, Hiai and Petz's original version is sufficient for our proof of theorem 18. For the reader's convenience, the proof is presented in appendix J.3.

Lemma 20. *Let Y be a curved exponential family and X be an exponential family including Y. For a curved exponential family and an exponential family, see chapter 4 in Amari and Nagaoka [1] or Barndorff-Nielsen [39]. In this setting, for n-i.i.d. data, the MLE $T^{ML}_{X,n}(\omega^n)$ for the exponential family X is a sufficient statistic for the curved exponential family Y, where $\vec{\omega}_n := (\omega_1, \ldots, \omega_n)$. Using the map $T : X \to Y$, we can define an estimator $T \circ T^{ML}_{X,n}$, and for an estimator T_Y, there exists a map $T : X \to Y$ such that $T_Y = T \circ T^{ML}_{X,n}$. We can identify a map T from X to Y with a sequence of estimators $T \circ T^{ML}_{X,n}(\vec{\omega}_n)$. We define the map $T_{\theta_0} : X \to Y$ as*

$$T_{\theta_0} := \arg\min_{\theta \in Y} \{D(x\|\theta) | D(\theta\|\theta_0) \leqslant D(x\|\theta_0)\}. \tag{74}$$

When Y is an exponential family (i.e. flat), T_{θ_0} coincides with the projection to Y. Then, the sequence of estimators corresponding to the map T_{θ_0} satisfies the strong consistency at θ_0 and the equation

$$\lim_{n\to\infty} \frac{-1}{n} \log p^n_{\theta_0}\left\{\left\|T_{\theta_0} \circ T^{ML}_{X,n}(\vec{\omega}_n) - \theta_0\right\| > \epsilon\right\} = \inf_{\theta \in Y}\{D(\theta\|\theta_0) | \|\theta - \theta_0\| > \epsilon\} \tag{75}$$

holds.

Proof. It is well known that for any subset $X' \subset X$, the equation

$$\lim_{n\to\infty} -\frac{1}{n} \log p^n_{\theta_0}\left\{T^{ML}_{X,n}(\vec{\omega}_n) \in X'\right\} = \inf_{x \in X'} D(x\|\theta_0) \tag{76}$$

holds. For the reader's convenience, we present a proof of (76) in appendix K. Thus, equation (75) follows from (74) and (76). If Y is an exponential family, then the estimator $T_{\theta_0} \circ T^{ML}_{X,n}$ coincides with the MLE and satisfies the strong consistency. Otherwise, we choose a neighbourhood U of θ_0 so that we can approximate the neighbourhood U by the tangent space. The estimator $T_{\theta_0} \circ T^{ML}_{X,n}$ can be approximated by the MLE and satisfies the strong consistency at U. Thus, it also satisfies the strong consistency at θ_0. □

Proof of theorem 18. Let $M = \{M_i\}$ be a faithful POVM defined in section 7.2 such that the number of operators M_i is finite. For any m and any $\delta > 0$, we define the POVM $M^m_{\theta_0}$ to be the disjoint random combination of $M \times m$ and $E^m_{\theta_0}$ with the ratio $\delta : 1-\delta$. Note that a disjoint random combination is defined in section 4. From the definition of $M^m_{\theta_0}$, the inequality

$$(1-\delta) D^{E^m_{\theta_0}}(\theta\|\theta) \leqslant D^{M^m_{\theta_0}}(\theta\|\theta) \tag{77}$$

holds. Since the map $\theta \mapsto \mathrm{P}^M_\theta$ is one-to-one, the map $\theta \mapsto \mathrm{P}^{M^m_{\theta_0}}_\theta$ is also one-to-one. Since M and $E^m_{\theta_0}$ are finite resolutions of the identity, the one-parameter family $\left\{\mathrm{P}^{M^m_{\theta_0}}_\theta \middle| \theta \in \Theta\right\}$ is a subset of multi-nominal distributions X, which is an exponential family. Applying lemma 20, we have

$$\lim_{n\to\infty} \frac{-1}{nm} \log \mathrm{P}^{M^m_{\theta_0}\times n}_{\theta_0}\left\{\left|T_{\theta_0} \circ T^{ML}_{X,n}(\vec{\omega}_n) - \theta_0\right| > \epsilon\right\}$$
$$= \frac{1}{M} \inf_{\theta\in\Theta}\left\{D^{M^m_{\theta_0}}(\theta\|\theta_0) \| |\theta - \theta_0| > \epsilon\right\}$$

$$\geqslant \frac{(1-\delta)}{m} \inf\left\{ D^{E^m_{\theta_0}}(\theta\|\theta_0) \| \theta - \theta_0| > \epsilon \right\}$$
$$\geqslant (1-\delta) \inf\{D(\theta\|\theta_0) | |\theta - \theta_0| > \epsilon\} - \frac{(1-\delta)(k-1)\log(m+1)}{m}$$

where the first inequality follows from (77) and the second inequality follows from (73). □

Remark 4. In the case of the one-parameter equatorial spin 1/2 system state family, the map $\theta \mapsto \mathrm{P}_\theta^{E^m_{\theta_0}}$ is not one-to-one. Therefore, we must not treat $E^m_{\theta_0}$ but $M^m_{\theta_0}$.

10. Conclusions

It has been clarified that the SLD Fisher information J_θ gives the essential large deviation bound in the quantum estimation and the KMB Fisher information $\tilde{J}_\theta$ gives the large deviation bound of consistent superefficient estimators. Since estimators attaining the bound $\frac{\tilde{J}_\theta}{2}$ are unnatural, the bound $\frac{J_\theta}{2}$ is more important from the viewpoint of quantum estimation than the bound $\frac{\tilde{J}_\theta}{2}$. On the other hand, as mentioned in appendix A, concerning a quantum analogue of information geometry from the viewpoint of e-connections, KMB is the most natural among the quantum versions of the Fisher information. The interpretation of these two facts which seem to contradict each other, remains a problem. Similarly, it is a future problem to explain geometrically the relationship between the change of the orders of limits and the difference between the two quantum analogues of the Fisher information.

Acknowledgments

The author wishes to thank Professor H Nagaoka for encouragement and essential advice regarding this manuscript. He also wishes to thank Professor K Matsumoto for useful advice regarding this manuscript. He is grateful to Professor S Amari and Professor A Tomita and two anonymous referees, whose comments helped to considerably improve the presentation.

Appendix A. Brief summary of information-geometrical properties of J_θ, $\tilde{J}_\theta$ and $\check{J}_\theta$

The quantum analogues of the Fisher information J_θ, $\tilde{J}_\theta$ and $\check{J}_\theta$ are obtained from the inner products J_ρ, $\tilde{J}_\rho$ and $\check{J}_\rho$ on the linear space consisting of self-adjoint operators

$$\tilde{J}_\rho(A, B) := \mathrm{Tr}\, A\tilde{L}_B \qquad B = \int_0^1 \rho^t \tilde{L}_B \rho^{1-t}\, \mathrm{d}t$$
$$J_\rho(A, B) := \mathrm{Tr}\, AL_B \qquad B = \tfrac{1}{2}(L_B\rho + \rho L_B)$$
$$\check{J}_\rho(A, B) := \mathrm{Tr}\, A\check{L}_B \qquad B = \rho\check{L}_B$$

in the following way:

$$J_\theta = J_{\rho_\theta}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}, \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right) \qquad \tilde{J}_\theta = \tilde{J}_{\rho_\theta}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}, \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right) \qquad \check{J}_\theta = \check{J}_{\rho_\theta}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}, \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right).$$

In the multi-dimensional case, these are regarded as metrics as follows. For example, we can define a metric

$$\langle \partial_i, \partial_j \rangle = J_{\rho_\theta}\left(\frac{\partial\rho_\theta}{\partial\theta^i}, \frac{\partial\rho_\theta}{\partial\theta^j}\right) \tag{A.1}$$

on the tangent space at θ and the rhs of (A.1) is called the SLD Fisher matrix.

In a quantum setting, any information precessing is described by a trace-preserving CP map $C : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}')$. These inner products satisfy the monotonicity:

$$J_{\rho_\theta}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}, \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right) \geqslant J_{C(\rho_\theta)}\left(\frac{\mathrm{d}C(\rho_\theta)}{\mathrm{d}\theta}, \frac{\mathrm{d}C(\rho_\theta)}{\mathrm{d}\theta}\right)$$
$$\tilde{J}_{\rho_\theta}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}, \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right) \geqslant \tilde{J}_{C(\rho_\theta)}\left(\frac{\mathrm{d}C(\rho_\theta)}{\mathrm{d}\theta}, \frac{\mathrm{d}C(\rho_\theta)}{\mathrm{d}\theta}\right)$$
$$\check{J}_{\rho_\theta}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}, \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right) \geqslant \check{J}_{C(\rho_\theta)}\left(\frac{\mathrm{d}C(\rho_\theta)}{\mathrm{d}\theta}, \frac{\mathrm{d}C(\rho_\theta)}{\mathrm{d}\theta}\right)$$

for a one-parametric density family $\{\rho_\theta \in \mathcal{S}(\mathcal{H}) | \theta \in \Theta \subset \mathbb{R}\}$ [1]. These inequalities can be regarded as the quantum versions of (5). An inner product satisfying the above is called a monotone inner product. According to Petz [2], the inner product $\check{J}_\rho$ is the maximum one among normalized monotone inner products and the inner product J_ρ is the minimum one.

In the information geometry community, we usually discuss the torsions. As is known within this community, α-connection is a generalization of e-connection. The torsion of α-connection concerning the Fisher inner product vanishes in any distribution family [1]. In quantum setting, we can define the e-connections with respect to several quantum Fisher inner products. One may expect that in a quantum setting, its torsion vanishes in any density family. However, for only the inner product $\tilde{J}_\rho$, the torsion of e-connection vanishes in any density family [1]. Thus, the KMB Fisher information seems the most natural quantum analogue of the Fisher information, from an information-geometrical viewpoint.

Appendix B. Proof of (15)

From (14), we can calculate as

$$D(\rho_{\theta+\epsilon}\|\rho_\theta) = \mathrm{Tr}(\rho_{\theta+\epsilon}(\log \rho_{\theta+\epsilon} - \log \rho_\theta)) \cong \mathrm{Tr}\left(\rho_\theta + \frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\epsilon\right)\left(\frac{\mathrm{d}\log \rho_\theta}{\mathrm{d}\theta}\epsilon + \frac{1}{2}\frac{\mathrm{d}^2 \log \rho_\theta}{\mathrm{d}\theta^2}\epsilon^2\right)$$
$$= \mathrm{Tr}(\rho_\theta \tilde{L}_\theta)\epsilon + \left(\mathrm{Tr}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\tilde{L}_\theta\right) + \frac{1}{2}\mathrm{Tr}\left(\rho_\theta \frac{\mathrm{d}^2 \log \rho_\theta}{\mathrm{d}\theta^2}\right)\right)\epsilon^2. \tag{B.1}$$

Next, we calculate the above coefficients

$$\mathrm{Tr}(\rho_\theta \tilde{L}_\theta) = \int_0^1 \mathrm{Tr}\left(\rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t}\right)\mathrm{d}t = \mathrm{Tr}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\right) = 0. \tag{B.2}$$

Using (B.2) and (14), we have

$$\mathrm{Tr}\left(\rho_\theta \frac{\mathrm{d}^2 \log \rho_\theta}{\mathrm{d}\theta^2}\right) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\mathrm{Tr}\left(\rho_\theta \frac{\mathrm{d}\log \rho_\theta}{\mathrm{d}\theta}\right)\right) - \mathrm{Tr}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\frac{\mathrm{d}\log \rho_\theta}{\mathrm{d}\theta}\right) = -\mathrm{Tr}\left(\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}\tilde{L}_\theta\right) = -\tilde{J}_\theta. \tag{B.3}$$

From (B.1)–(B.3), we obtain

$$D(\rho_{\theta+\epsilon}\|\rho_\theta) \cong \tfrac{1}{2}\tilde{J}_\theta \epsilon^2.$$

Appendix C. Proof of lemma 3

We define the unitary operator U_ϵ as

$$b^2(\rho_\theta, \rho_{\theta+\epsilon}) = 2(1 - \mathrm{Tr}\,|\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}|) = \mathrm{Tr}(\sqrt{\rho} - \sqrt{\sigma}U_\epsilon)(\sqrt{\rho} - \sqrt{\sigma}U_\epsilon)^*.$$

Letting $W(\epsilon)$ be $\sqrt{\rho_{\theta+\epsilon}}U_\epsilon$, then we have

$$b^2(\rho_\theta, \rho_{\theta+\epsilon}) = \mathrm{Tr}(W(0) - W(\epsilon))(W(0) - W(\epsilon))^* \cong \mathrm{Tr}\left(-\frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0)\epsilon\right)\left(-\frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0)\epsilon\right)^* \cong \mathrm{Tr}\,\frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0)\frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0)^*\epsilon^2.$$

As is proved in the following discussion, the SLD L satisfies

$$\frac{\mathrm{d}W}{\mathrm{d}t}(0) = \frac{1}{2}LW(0). \tag{C.1}$$

Therefore, we have

$$b^2(\rho_\theta, \rho_{\theta+\epsilon}) \cong \mathrm{Tr}\,\tfrac{1}{4}LW(0)W(0)^*L\epsilon^2 = \tfrac{1}{4}\mathrm{Tr}\,L^2\rho_\theta\epsilon.$$

We obtain (38). It is sufficient to show (C.1).

From the definition of the Bures distance, we have

$$\begin{aligned} b^2(\rho_\theta, \rho_{\theta+\epsilon}) &= \min_{U:\text{unitary}} \mathrm{Tr}(\sqrt{\rho_\theta} - \sqrt{\rho_{\theta+\epsilon}}U)(\sqrt{\rho_\theta} - \sqrt{\rho_{\theta+\epsilon}}U)^* \\ &= 2 - \max_{U:\text{unitary}} \mathrm{Tr}\,\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}U^* + U\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta} \\ &= 2 - \mathrm{Tr}\,|\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}| + |\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta}| \\ &= 2 - \mathrm{Tr}(\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}U(\epsilon)^* + U(\epsilon)\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta}) \end{aligned}$$

which implies that $\sqrt{\rho_\theta}\sqrt{\rho_{\theta+\epsilon}}U(\epsilon)^* = U(\epsilon)\sqrt{\rho_{\theta+\epsilon}}\sqrt{\rho_\theta}$. Therefore, $W(0)W(\epsilon)^* = W(\epsilon)W(0)^*$. Taking the derivative, we have

$$W(0)\frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0)^* = \frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0)W(0)^*$$

which implies that there exists a self-adjoint operator L such that

$$\frac{\mathrm{d}W}{\mathrm{d}\epsilon}(0) = \frac{1}{2}LW(0).$$

Since $\rho_{\theta+\epsilon} = W(\epsilon)W(\epsilon)^*$, we have

$$\frac{\mathrm{d}\rho}{\mathrm{d}\theta}(\theta) = \frac{1}{2}(LW(0)W(0)^* + W(0)W(0)^*L).$$

Thus, the operator L coincides with the SLD.

Appendix D. Proof of (43)

Let $M = \{M_i\}$ be an arbitrary POVM. We choose the unitary U satisfying

$$U\sigma^{1/2}\rho^{1/2} = \sqrt{\rho^{1/2}\sigma\rho^{1/2}}.$$

Using the Schwarz inequality, we have

$$\begin{aligned} \sqrt{\mathrm{P}_\rho^M(\omega)}\sqrt{\mathrm{P}_\sigma^M(\omega)} &= \sqrt{\mathrm{Tr}\left(M_\omega^{1/2}\sigma^{1/2}U^*\right)^*\left(M_\omega^{1/2}\sigma^{1/2}U^*\right)}\sqrt{\mathrm{Tr}\left(M_\omega^{1/2}\rho^{1/2}\right)^*\left(M_\omega^{1/2}\rho^{1/2}\right)} \\ &\geqslant \mathrm{Tr}\left(M_\omega^{1/2}\sigma^{1/2}U^*\right)^*\left(M_\omega^{1/2}\rho^{1/2}\right) = \left|\mathrm{Tr}\,U\sigma^{1/2}M_\omega\rho^{1/2}\right|. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_\omega \left(\sqrt{\mathrm{P}_\rho^M(\omega)}\sqrt{\mathrm{P}_\sigma^M(\omega)}\right) &\geqslant \sum_\omega |\mathrm{Tr}\,U\sigma^{1/2}M_\omega\rho^{1/2}| \geqslant \left|\sum_\omega \mathrm{Tr}\,U\sigma^{1/2}M_\omega\rho^{1/2}\right| \\ &= |\mathrm{Tr}\,U\sigma^{1/2}\rho^{1/2}| = \mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}. \end{aligned}$$

Appendix E. Proof of lemma 4

Let m and ϵ be an arbitrary positive integer and an arbitrary positive real number, respectively. There exists a sufficiently large integer N such that

$$\frac{1}{n}\log \mathrm{P}_\theta^{M^n}\left\{|\hat\theta-\theta|>\frac{\delta}{m}i\right\}\leqslant -\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}i\right)+\epsilon$$
$$\frac{1}{n}\log \mathrm{P}_{\theta+\delta}^{M^n}\left\{|\hat\theta-(\theta+\delta)|>\frac{\delta}{m}(m-i)\right\}\leqslant -\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i)\right)+\epsilon$$

for $i=0,\ldots,m$ and $\forall n\geqslant N$. From the monotonicity (42) and the additivity (39) of quantum affinity, we perform the following evaluation:

$$\begin{aligned}
-\frac{n}{8}I(\rho_\theta\|\rho_{\theta+\delta}) &= -\frac{1}{8}I\left(\rho_\theta^{\otimes n}\middle\|\rho_{\theta+\delta}^{\otimes n}\right)\\
&\leqslant \log\Bigg(\mathrm{P}_\theta^{M^n}\{\hat\theta\leqslant\theta\}^{\frac12}\mathrm{P}_{\theta+\delta}^{M^n}\{\hat\theta\leqslant\theta\}^{\frac12}+\mathrm{P}_\theta^{M^n}\{\theta+\delta<\hat\theta\}^{\frac12}\mathrm{P}_{\theta+\delta}^{M^n}\{\theta+\delta<\hat\theta\}^{\frac12}+\sum_{i=1}^{m}\mathrm{P}_\theta^{M^n}\\
&\quad\times\left\{\theta+\frac{\delta}{m}(i-1)<\hat\theta\leqslant\theta+\frac{\delta}{m}i\right\}^{\frac12}\mathrm{P}_{\theta+\delta}^{M^n}\left\{\theta+\frac{\delta}{m}(i-1)<\hat\theta\leqslant\theta+\frac{\delta}{m}i\right\}^{\frac12}\Bigg)\\
&\leqslant \log\Bigg(\mathrm{P}_{\theta+\delta}^{M^n}\{|\hat\theta-(\theta+\delta)|\geqslant\delta\}^{\frac12}+\mathrm{P}_\theta^{M^n}\{|\hat\theta-\theta|>\delta\}^{\frac12}\\
&\quad+\sum_{i=1}^{m}\mathrm{P}_\theta^{M^n}\left\{|\hat\theta-\theta|>\frac{\delta}{m}(i-1)\right\}^{\frac12}\mathrm{P}_{\theta+\delta}^{M^n}\left\{|\hat\theta-(\theta+\delta)|\geqslant\frac{\delta}{m}(m-i)\right\}^{\frac12}\Bigg)\\
&\leqslant \log\Bigg(\mathrm{P}_{\theta+\delta}^{M^n}\left\{|\hat\theta-(\theta+\delta)|>\frac{\delta}{m}(m-1)\delta\right\}^{\frac12}+\mathrm{P}_\theta^{M^n}\{|\hat\theta-\theta|>\delta\}^{\frac12}\\
&\quad+\sum_{i=1}^{m}\mathrm{P}_\theta^{M^n}\left\{|\hat\theta-\theta|>\frac{\delta}{m}(i-1)\right\}^{\frac12}\mathrm{P}_{\theta+\delta}^{M^n}\left\{|\hat\theta-(\theta+\delta)|>\frac{\delta}{m}(m-i-1)\right\}^{\frac12}\Bigg)\\
&\leqslant \log\Bigg(\exp\left(-\frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(m-1)\right)-\epsilon\right)\right)\\
&\quad+\exp\Big(-\frac{n}{2}(\underline{\beta}(\vec{M},\theta+\delta,\delta)-\epsilon)\Big)+\sum_{i=1}^{m}\exp\Bigg(-\frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)-\epsilon\right)\\
&\quad-\frac{n}{2}\left(\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-\epsilon\right)\Bigg)\Bigg)\\
&\leqslant \log(m+2)\exp\left(-\frac{n}{2}\min_{0\leqslant i\leqslant m}\left(\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)\right.\right.\\
&\quad\left.\left.+\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-2\epsilon\right)\right)\\
&= \log(m+2)-\frac{n}{2}\left(\min_{0\leqslant i\leqslant m}\underline{\beta}\left(\vec{M},\theta,\frac{\delta}{m}(i-1)\right)\right.\\
&\quad\left.+\underline{\beta}\left(\vec{M},\theta+\delta,\frac{\delta}{m}(m-i-1)\right)-2\epsilon\right)
\end{aligned}$$

where we assume that $\underline{\beta}(\vec{M}, \theta, a) = 0$ for any negative real number a. Taking the limit $n \to \infty$ after dividing by n, we have

$$\frac{1}{8} I(\rho_\theta \| \rho_{\theta+\delta}) \geqslant \frac{1}{2} \min_{0 \leqslant i \leqslant m} \left(\underline{\beta}\left(\vec{M}, \theta, \frac{\delta}{m}(i-1)\right) + \underline{\beta}\left(\vec{M}, \theta+\delta, \frac{\delta}{m}(m-i-1)\right) - 2\epsilon \right).$$

Since $\epsilon > 0$ is arbitrary, the inequality

$$\frac{1}{8} I(\rho_\theta \| \rho_{\theta+\delta}) \geqslant \frac{1}{2} \min_{0 \leqslant i \leqslant m} \left(\underline{\beta}\left(\vec{M}, \theta, \frac{\delta}{m}(i-1)\right) + \underline{\beta}\left(\vec{M}, \theta+\delta, \frac{\delta}{m}(m-i-1)\right) \right)$$

holds. Taking the limit $m \to \infty$, we obtain (44).

Appendix F. Unitary evolutions on the boson coherent system

In the system $\mathcal{H} = L^2(\mathbb{R})$, the unitary operator $U_1(\beta) := \exp(\beta a^* - \beta^* a)$ acts on the coherent state as

$$U_1(\beta)|\alpha\rangle = |\alpha - \beta\rangle$$

where α and β are complex numbers and a is the annihilation operator. Thus, we can verify that

$$U_1(\beta)\rho_\alpha U_1(\beta)^* = \rho_{\alpha-\beta}.$$

Now, we let a_i be the annihilation operator on the ith system. The unitary operator $U_n(\beta) := \prod_{i=1}^n \exp(-\beta a_i^* + \beta^* a_i)$ acts on the system $\mathcal{H}^{\otimes n}$ as

$$U_n(\beta)\rho_\theta^{\otimes n} U_n(\beta)^* = \rho_{\theta-\beta}^{\otimes n}.$$

In the two-mode system $\mathcal{H} \otimes \mathcal{H}$, the unitary $V_2(t) := \exp t(-a_2^* a_1 + a_1^* a_2)$ acts as

$$V_1(t)|\alpha_1\rangle \otimes |\alpha_2\rangle = |\alpha_1 \cos t + \alpha_2 \sin t\rangle \otimes |-\alpha_1 \sin t + \alpha_2 \cos t\rangle.$$

Thus, we can verify that

$$V_1(t)\rho_{\theta_1} \otimes \rho_{\theta_2} V_1(t)^* = \rho_{\theta_1 \cos t + \theta_2 \sin t} \otimes \rho_{-\theta_1 \sin t + \theta_2 \cos t}.$$

Therefore, the unitary $V_n := \prod_{i=1}^n \exp t_i(-a_i^* a_1 + a_1^* a_i)$ satisfies

$$V_n \rho_\theta^{\otimes n} V_n^* = \rho_{\sqrt{n}\theta} \otimes \rho_0^{\otimes(n-1)}$$

where $\cos t_i = \sqrt{\frac{i-1}{i}}$, $\sin t_i = \sqrt{\frac{1}{i}}$.

Appendix G. Proof of proposition 8

For a proof of proposition 8, we need the following lemma.

Lemma 21. *Let $g_n(\omega)$, $f_n(\omega)$ be functions on Ω. Assume that the functions $\beta_1(\omega) := \lim_{n\to\infty} \frac{-1}{n} \log f_n(\omega)$ and $\beta_2(\omega) := \lim_{n\to\infty} \frac{-1}{n} \log g_n(\omega)$ are continuous. If the inequality $g_n(\omega) \leqslant 1$ holds for any element $\omega \in \Omega$ and any positive integer n, and if there exists a subset $K \subset \Omega$ such that*

$$\lim_{n\to\infty} \frac{-1}{n} \log \left(\int_K f_n(\omega)\, \mathrm{d}\omega \right) > \min_{\omega\in\Omega}(\beta_1(\omega) + \beta_2(\omega))$$

the relation

$$\lim_{n\to\infty} \frac{-1}{n} \log \left(\int_\Omega f_n(\omega) g_n(\omega)\, \mathrm{d}\omega \right) = \min_{\omega\in\Omega}(\beta_1(\omega) + \beta_2(\omega))$$

holds.

Similar to lemma 4, lemma 21 is proved.

Now, we will prove proposition 8. From the definition of $M^{w,n}$ and the equation $\rho_0 = \frac{1}{\overline{N}+1}\sum_k \left(\frac{\overline{N}}{\overline{N}+1}\right)^k |k\rangle\langle k|$, we have

$$\log \mathrm{P}_0^{M^{s,n}}\{T_n > \epsilon\} = \log \sum_{k>n\epsilon^2} \left(\frac{\overline{N}}{\overline{N}+1}\right)^k = \log \left(\frac{\overline{N}}{\overline{N}+1}\right)^{[n\epsilon^2]}$$

where [ ] is a Gauss notation. Therefore, we obtain

$$\beta(\vec{M}^w, 0, \epsilon) = \epsilon^2 \log\left(1 + \frac{1}{\overline{N}}\right)$$

which implies (50).

Next, we prove the strong consistency condition and (51). We perform the following calculation:

$$\begin{aligned}\mathrm{P}_\theta^{M^{w,n}}\{T_n - \theta > \epsilon\} &= \sum_{k>(\theta+\epsilon)^2 n} \langle k| \int_{\mathbb{C}} \frac{1}{\pi\overline{N}} |\alpha\rangle\langle\alpha| \, \mathrm{e}^{-\frac{|\alpha-\sqrt{n}\theta|^2}{\overline{N}}} \mathrm{d}^2\alpha |k\rangle \\ &= \int_{\mathbb{C}} \frac{\sqrt{n}}{\pi\overline{N}} \mathrm{e}^{-n\frac{|\alpha-\theta|^2}{\overline{N}}} \sum_{k>(\theta-\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!} \mathrm{e}^{-n|\alpha|^2} \mathrm{d}^2\alpha. \end{aligned} \tag{G.1}$$

The equation

$$\lim_{n\to\infty} \frac{-1}{n} \log \frac{\sqrt{n}}{\pi\overline{N}} \mathrm{e}^{-n\frac{|\alpha-\theta|^2}{\overline{N}}} = \frac{|\alpha-\theta|^2}{\overline{N}} \tag{G.2}$$

holds. Also, as proved in the latter, the equations

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log \left(\sum_{k>(\theta+\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!} \mathrm{e}^{-n|\alpha|^2} \right) \\ &\quad = \left((\theta+\epsilon)^2 \log \frac{(\theta+\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta+\epsilon)^2 \right) 1((\theta+\epsilon)^2 - |\alpha|^2) \end{aligned} \tag{G.3}$$

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log \left(\sum_{k<(\theta-\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!} \mathrm{e}^{-n|\alpha|^2} \right) \\ &\quad = \left((\theta-\epsilon)^2 \log \frac{(\theta-\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta-\epsilon)^2 \right) 1(-(\theta-\epsilon)^2 + |\alpha|^2) \end{aligned} \tag{G.4}$$

hold, where $1(x)$ is defined as

$$1(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x < 0. \end{cases}$$

For any $\delta > 0$, there exists a real number K such that

$$\lim_{n\to\infty} -\frac{1}{n} \log \left(\int_{|\alpha|>K} \frac{\sqrt{n}}{\pi\overline{N}} \exp\left(-n\frac{|\alpha-\theta|^2}{\overline{N}}\right) \mathrm{d}x \right) = \frac{K-\theta}{\overline{N}} > \delta.$$

Now, we can apply lemma 21 to (G.1). From (G.2) and (G.3), the relations

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log \mathrm{P}^{M_\theta^{w,n}}\{T_n - \theta > \epsilon\} \\ &\quad = \min_{\alpha\in\mathbb{C}} \left(\frac{|\alpha-\theta|^2}{\overline{N}} \right. \end{aligned}$$

$$+\left((\theta+\epsilon)^2\log\frac{(\theta+\epsilon)^2}{|\alpha|^2}+|\alpha|^2-(\theta+\epsilon)^2\right)1((\theta+\epsilon)^2-|\alpha|^2)\Bigg)$$
$$=\min_{\alpha\in\mathbb{R}}\left(\frac{|\alpha-\theta|^2}{\overline{N}}\right.$$
$$+\left((\theta+\epsilon)^2\log\frac{(\theta+\epsilon)^2}{|\alpha|^2}+|\alpha|^2-(\theta+\epsilon)^2\right)1((\theta+\epsilon)^2-|\alpha|^2)\Bigg)$$
$$=\min_{s\in\mathbb{R}}\left(\frac{s^2}{\overline{N}}\right.$$
$$+\left((\theta+\epsilon)^2\log\frac{(\theta+\epsilon)^2}{(\theta-s)^2}+(\theta-s)^2-(\theta+\epsilon)^2\right)1((\theta+\epsilon)^2-(\theta-s)^2)\Bigg)$$

hold. If ϵ is sufficiently small for θ, we have the following approximation:

$$\lim_{n\to\infty}\frac{-1}{n}\log \mathrm{P}_\theta^{M^{w,n}}\{T_n-\theta>\epsilon\}\cong\min_s\frac{1+2\overline{N}}{\overline{N}}\left(s-\frac{2\overline{N}}{1+2\overline{N}}\epsilon\right)^2+\frac{\epsilon^2}{\overline{N}+\frac{1}{2}}.$$

Thus,

$$\lim_{\epsilon\to0}\lim_{n\to\infty}\frac{-1}{n\epsilon^2}\log \mathrm{P}_\theta^{M^{w,n}}\{T_n-\theta>\epsilon\}=\frac{1}{\overline{N}+\frac{1}{2}}. \tag{G.5}$$

The second convergence of the lhs of (G.5) is uniform in a sufficiently small neighbourhood U_{θ_0} of arbitrary $\theta_0\in\mathbb{R}^+\backslash\{0\}$.

Similar to (G.5), from (G.4), we can prove

$$\lim_{\epsilon\to0}\lim_{n\to\infty}\frac{-1}{n\epsilon^2}\log \mathrm{P}_\theta^{M^{w,n}}\{T_n-\theta<-\epsilon\}=\frac{1}{\overline{N}+\frac{1}{2}}. \tag{G.6}$$

Also, the second convergence of the lhs of (G.6) is uniform at a sufficiently small neighbourhood U_{θ_0} of arbitrary $\theta_0\in\mathbb{R}^+\backslash\{0\}$. Thus, (51) and the strong consistency condition are proved.

Next, we prove (G.3) and (G.4). Using the Stirling formula, we have

$$\lim_{n\to\infty}\frac{-1}{n}\log\frac{(n|\alpha|^2)^{[\delta n]}}{[\delta n]!}\mathrm{e}^{-n|\alpha|^2}=\left(\delta\log\frac{\delta}{|\alpha|^2}+|\alpha|-\delta^2\right)1(\delta-|\alpha|^2). \tag{G.7}$$

Since the relations

$$\frac{(n|\alpha|^2)^{([(\theta-\epsilon)^2n]-1)}}{([(\theta-\epsilon)^2n]-1)!}\mathrm{e}^{-n|\alpha|^2}\leqslant\sum_{k<(\theta-\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2}\leqslant[(\theta-\epsilon)^2n]\frac{(n|\alpha|^2)^{([(\theta-\epsilon)^2n]-1)}}{([(\theta-\epsilon)^2n]-1)!}\mathrm{e}^{-n|\alpha|^2}$$

hold, (G.4) follows from (G.7). If $(\theta+\epsilon)^2\leqslant|\alpha|^2$, the equation

$$\lim_{n\to\infty}\frac{-1}{n}\log\sum_{k>(\theta+\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2}=0 \tag{G.8}$$

holds. It implies (G.3) in the case of $(\theta+\epsilon)^2\leqslant|\alpha|^2$.

Next we prove (G.3) in the case of $(\theta+\epsilon)^2>|\alpha|^2$. In this case, we have

$$\sum_{Ln>k>(\theta+\epsilon)^2n}\frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2}\leqslant n(L-(\theta+\epsilon)^2)\frac{(n|\alpha|^2)^{[(\theta+\epsilon)^2n]}}{[(\theta+\epsilon)^2n]!}\mathrm{e}^{-n|\alpha|^2} \tag{G.9}$$

because $\left(\frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2}\right)\Big/\left(\frac{(n|\alpha|^2)^{(k+1)}}{(k+1)!}\mathrm{e}^{-n|\alpha|^2}\right)=\frac{k+1}{n|\alpha|^2}$. If L and N are sufficiently large for $|\alpha|^2$, we have

$$\sum_{k\geqslant Ln}\frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2}\leqslant\sum_{k\geqslant Ln}\mathrm{e}^{-k}=\frac{\mathrm{e}^{-nL}}{1-\mathrm{e}^{-1}} \tag{G.10}$$

because (G.7) implies that

$$\frac{(n|\alpha|^2)^{[\delta n]}}{[\delta n]!}\mathrm{e}^{-n|\alpha|^2} \leqslant \mathrm{e}^{-[\delta n]} \qquad \forall \delta \geqslant L \quad \forall n \geqslant N.$$

Since the relations

$$\begin{aligned}\frac{(n|\alpha|^2)^{[(\theta+\epsilon)^2 n]}}{[(\theta+\epsilon)^2 n]!}\mathrm{e}^{-n|\alpha|^2} &\leqslant \sum_{k>(\theta+\epsilon)^2 n} \frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2} \\ &\leqslant n(L-(\theta+\epsilon)^2)\frac{(n|\alpha|^2)^{[(\theta+\epsilon)^2 n]}}{[(\theta+\epsilon)^2 n]!}\mathrm{e}^{-n|\alpha|^2} + \frac{\mathrm{e}^{-nL}}{1-\mathrm{e}^{-1}}\end{aligned}$$

hold, we have

$$\begin{aligned}\left((\theta+\epsilon)^2 \log\frac{(\theta+\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta+\epsilon)^2\right) &\geqslant \lim_{n\to\infty}\frac{-1}{n}\log\left(\sum_{k>(\theta+\epsilon)^2 n}\frac{(n|\alpha|^2)^k}{k!}\mathrm{e}^{-n|\alpha|^2}\right) \\ &\geqslant \min\left\{\left((\theta+\epsilon)^2 \log\frac{(\theta+\epsilon)^2}{|\alpha|^2} + |\alpha|^2 - (\theta+\epsilon)^2\right), L\right\}.\end{aligned}$$

If we let L be a sufficiently large real number, we have (G.3).

Appendix H. Proof of theorem 11

In this proof, we use the function $\phi_{\theta,\check{\theta}}(s)$ defined in (K.1). First, we prove the following four facts.

(i) The faithful POVM M_f satisfies the inequalities

$$\beta(\vec{M}_f, \theta, \epsilon) > 0 \qquad \alpha(\vec{M}_f, \theta) > 0.$$

(ii) The relation

$$\lim_{\check{\theta}\to\theta}\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}} - \frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2\right)^{-1} = J_\theta \qquad \forall\theta\in\Theta$$

holds.

(iii) The equation

$$\lim_{s\to 0}\frac{\phi_{\theta,\check{\theta}}(s)-1}{s^2} = \frac{1}{2}\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}} - \frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2 \tag{H.1}$$

holds. The lhs converges uniformly w.r.t. $\theta, \check{\theta}$.

(iv) For any real number $\delta_2 > 0$, there exists a sufficiently small real number $\epsilon > 0$ such that if $|\mathrm{Tr}\,\rho_\theta L_{\check{\theta}} - \mathrm{Tr}\,\rho_{\theta'} L_{\check{\theta}}| \leqslant \epsilon(1-\delta_2)$ and $|\check{\theta}-\theta| < \sqrt{\epsilon}$, then $|\theta'-\theta| < \epsilon$.

Fact (i) is easily proved from the definition of M_f. Fact (iii) is proved by the relation

$$\sup_{\check{\theta},\theta}\left\|\frac{L_{\check{\theta}}}{J_{\check{\theta}}} - \frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right\| < \infty.$$

Fact (ii) is, also, proved by the relations

$$\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}} - \frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2 = \frac{\mathrm{Tr}\,\rho_\theta\left(L_{\check{\theta}}^2\right)}{J_{\check{\theta}}^2} - \frac{(\mathrm{Tr}\,\rho_\theta L_{\check{\theta}})^2}{J_{\check{\theta}}^2} \to J_\theta^{-1} \qquad \text{as} \quad \check{\theta}\to\theta.$$

Fact (iv) follows from the relation

$$\frac{\partial \operatorname{Tr} \rho_\theta L_{\check\theta}}{\partial \theta} \to 1 \qquad \text{as} \quad \check\theta \to \theta$$

which follows from fact (i).

Next, we prove the theorem from the preceding four facts. The inequality

$$\mathrm{P}_\theta^{M_\delta^{s,n}}\{\hat\theta \notin U_{\theta,\epsilon}\} \leqslant \mathrm{P}_\theta^{M_f\times\delta n}\{\hat\theta \in U_{\theta,\sqrt{\epsilon}}\} \sup_{\check\theta\in U_{\theta,\sqrt{\epsilon}}} \mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{\hat\theta \notin U_{\theta,\epsilon}\} + \mathrm{P}_\theta^{M_f\times\delta n}\{\hat\theta \notin U_{\theta,\sqrt{\epsilon}}\} \tag{H.2}$$

holds. As proved in the latter, the inequality

$$\begin{aligned}\liminf_{n\to\infty} -\frac{1}{n}\log \sup_{\check\theta\in U_{\theta,\sqrt{\epsilon}}} \mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\left\{T_{\check\theta}^n \notin U_{\theta,\epsilon}\right\} \\ \geqslant (1-\delta) g\left(\epsilon^2(1-\delta_2)^2\frac{1}{2}\left(\operatorname{Tr}\rho_\theta\left(\frac{L_{\check\theta}}{J_{\check\theta}} - \frac{\operatorname{Tr}\rho_\theta L_{\check\theta}}{J_{\check\theta}}\right)^2\right)^{-1}, \frac{\epsilon^2(1-\delta_2)^2}{2}\delta\right)\end{aligned} \tag{H.3}$$

holds, where the function $g(x, y)$ is defined as $g(x, y) := x - \log\left(1 + \frac{x}{2} + y\right)$. Therefore, we have

$$\begin{aligned}\underline{\beta}\left(\vec{M}_\delta^s, \theta, \epsilon\right) &= \liminf_{n\to\infty} -\frac{1}{n}\log \mathrm{P}_\theta^{M_\delta^{s,n}}\{\hat\theta \notin U_{\theta,\sqrt{\epsilon}}\} \\ &\geqslant \min\left\{(1-\delta) g\left(\epsilon^2(1-\delta_2)^2\frac{1}{2}\left(\operatorname{Tr}\rho_\theta\left(\frac{L_{\check\theta}}{J_{\check\theta}} - \frac{\operatorname{Tr}\rho_\theta L_{\check\theta}}{J_{\check\theta}}\right)^2\right)^{-1}, \frac{\epsilon^2(1-\delta_2)^2}{2}\delta\right),\right. \\ &\qquad \left. c\underline{\beta}(\{M_f\times\delta n\}, \theta, \sqrt{\epsilon})\right\}.\end{aligned} \tag{H.4}$$

From facts (i) and (ii), the equations

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}(\text{rhs of (H.4)}) \\ &\quad = \frac{1-\delta}{2}\left(\lim_{\check\theta\to\theta}(1-\delta_1)^2(1-\delta_2)^2\left(\operatorname{Tr}\rho_\theta\left(\frac{L_{\check\theta}}{J_{\check\theta}} - \frac{\operatorname{Tr}\rho_\theta L_{\check\theta}}{J_{\check\theta}}\right)^2\right)^{-1} - (1-\delta_2)^2\delta_3\right) \\ &\quad = \frac{1-\delta}{2}((1-\delta_1)^2(1-\delta_2)^2 J_\theta - (1-\delta_2)^2\delta_3)\end{aligned} \tag{H.5}$$

hold. The rhs of (H.5) converges locally uniformly w.r.t. θ. Let $\underline{\beta}_m\left(\vec{M}_\delta^s, \theta, \epsilon\right)$ be the rhs of (H.4) in the case of $\delta_2 = \delta_3 = \frac{1}{m}$. Therefore, we have

$$\lim_{m\to\infty}\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\underline{\beta}_m\left(\vec{M}_\delta^s, \theta, \epsilon\right) = \frac{1-\delta}{2}J_\theta$$

which implies that

$$\underline{\alpha}\left(\vec{M}_\delta^s, \theta\right) \geqslant \frac{1-\delta}{2}J_\theta.$$

If the converse inequality

$$\alpha\left(\vec{M}_\delta^s, \theta\right) \leqslant \frac{1-\delta}{2}J_\theta \tag{H.6}$$

holds, we can immediately derive relations (55) and show that the sequence of estimators $\vec{M}^s_\delta$ satisfies the second strong consistency condition.

In the following, relations (H.6) and (H.3) are proved. First, we prove (H.6). We can evaluate the probability $\mathrm{P}_\theta^{M^{s,n}_\delta}\{\hat\theta \in U_{\theta,\epsilon}\}$ as

$$\begin{aligned}
-\log \mathrm{P}_\theta^{M^{s,n}_\delta}\{\hat\theta \in U_{\theta,\epsilon}\} &= -\log \int \mathrm{P}_\theta^{M_f\times \delta n}(\mathrm{d}\check\theta)\mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\left\{T^n_{\check\theta} \notin U_{\theta,\epsilon}\right\} \\
&\leqslant -\int \mathrm{P}_\theta^{M_f\times \delta n}(\mathrm{d}\check\theta)\log\left(\mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\left\{T^n_{\check\theta} \notin U_{\theta,\epsilon}\right\}\right) \\
&\leqslant -\int \mathrm{P}_\theta^{M_f\times \delta n}(\mathrm{d}\check\theta)\frac{D^{L_{\check\theta}\times(1-\delta)n}(\theta+\xi\epsilon\|\theta)+h\left(\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}\right)}{\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}}
\end{aligned}$$

where $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n} := \mathrm{P}^{L_{\check\theta}\times(1-\delta)n}_{\theta+\xi\epsilon,n}\{T_{\check\theta^n} \notin U_{\theta,\epsilon}\}$, and similar to (35), we can prove the last inequality. For any $\delta_4 > 0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} -\frac{1}{n}\log \mathrm{P}_\theta^{\vec{M}^s_\delta}\{T_n \notin U_{\theta,\epsilon}\} \\
&\leqslant \limsup_{n\to\infty}\int_{\mathbb{R}} \mathrm{P}_\theta^{M_f\times\delta n}(\mathrm{d}\check\theta)(1-\delta)\min_{\xi=1-\delta_4,-(1-\delta_4)}\frac{(1-\delta)D^{L_{\check\theta}}(\theta+\xi\epsilon\|\theta)+\frac{h\left(\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}\right)}{n}}{(1-\delta)\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}} \\
&= (1-\delta)\min_{\xi=1-\delta_4,-(1-\delta_4)} D^{L_{\check\theta}}(\theta+\xi\epsilon\|\theta) = \frac{1-\delta}{2}J_\theta.
\end{aligned}$$

The last equation is derived from Lebesgue's convergence theorem and the fact that the probability $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}$ tends to 1 uniformly w.r.t. $\check\theta$, as follows from assumptions 1 and 3.

The reason for the applicability of Lebesgue's convergence theorem is given as follows. Since $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}$ tends to 1 uniformly w.r.t. $\check\theta$, there exists $N, R > 0$ such that $\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n} > \frac{1}{R}$, $\forall\check\theta\in\Theta, n \geqslant N$. Thus, we have

$$\frac{D^{L_{\check\theta}\times(1-\delta)n}(\theta+\xi\epsilon\|\theta)+h\left(\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}\right)}{\mathrm{P}^{L_{\check\theta}}_{\theta+\xi\epsilon,n}} \leqslant \frac{R}{1-\delta}\left((1-\delta)D(\theta+\epsilon\xi\|\theta)+2\right) < \infty.$$

Therefore, we can apply Lebesgue's convergence theorem. Thus, the relations

$$\begin{aligned}
\alpha\left(\vec{M}^s_\delta,\theta\right) &= \limsup_{\epsilon\to 0}\limsup_{n\to\infty} -\frac{1}{n\epsilon^2}\log \mathrm{P}_\theta^{\vec{M}^s_\delta}\{T_n \notin U_{\theta,\epsilon}\} \\
&\leqslant (1-\delta)\limsup_{\epsilon\to 0}\frac{1}{\epsilon^2}\min_{\xi=1-\delta_4,-(1-\delta_4)} D^{L_{\check\theta}}(\theta+\xi\epsilon\|\theta) \\
&= (1-\delta)(1-\delta_4)^2\tfrac{1}{2}J_\theta
\end{aligned}$$

hold. Since $\delta_4 > 0$ is arbitrary, inequality (H.6) holds.

Next, we prove the inequality (H.3). Assume that $|\check\theta-\theta| \leqslant \epsilon$ and define

$$\Lambda(\xi,\check\theta,\theta) := \sup_{\eta\in\mathbb{R}}(\eta\xi - \log\phi_{\theta,\check\theta}(\eta)).$$

Then, the inequalities

$$\mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{\check\theta \notin U_{\theta,\epsilon}\} \leqslant \mathrm{P}_\theta^{L_{\check\theta}\times(1-\delta)n}\{|\mathrm{Tr}\,\rho_{\hat\theta}L_{\check\theta} - \mathrm{Tr}\,\rho_\theta L_{\check\theta}| \leqslant (1-\delta_2)\epsilon\} \tag{H.7}$$

$$\leqslant 2\exp(-(1-\delta)n\min\{\Lambda((1-\delta_2)\epsilon,\check\theta,\theta),\Lambda(-(1-\delta_2)\epsilon,\check\theta,\theta)\}) \tag{H.8}$$

hold, where (H.7) is derived from fact (iv), and (H.8) is derived from Markov's inequality.

Thus,

$$\lim_{n\to\infty} -\frac{1}{n}\log \sup_{\check{\theta}\in U_{\theta,\sqrt{\epsilon}}} \mathrm{P}_{\theta}^{L_{\check{\theta}}\times(1-\delta)n}\{\check{\theta}\notin U_{\theta,\epsilon}\}$$
$$\geqslant (1-\delta)\inf_{\check{\theta}\in U_{\theta,\sqrt{\epsilon}}} \min\{\Lambda((1-\delta_2)\epsilon,\check{\theta},\theta),\Lambda(-(1-\delta_2)\epsilon,\check{\theta},\theta)\}. \tag{H.9}$$

We let $\epsilon>0$ be a sufficiently small real number for arbitrary $\delta_3>0$ and define η by

$$\eta := \epsilon(1-\delta_2)\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2\right)^{-1}.$$

Then, the inequalities

$$\Lambda(\pm(1-\delta_2)\epsilon,\check{\theta},\theta)\geqslant \pm(1-\delta_2)\epsilon(\pm\eta)-\log\phi_{\theta,\check{\theta}}(\pm\eta)$$
$$\geqslant \epsilon^2(1-\delta)^2\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2\right)^{-1}$$
$$-\log\left(1+\frac{\epsilon^2(1-\delta)^2}{2}\left(\left(\mathrm{Tr}\,\rho_\theta\left(\frac{L_{\check{\theta}}}{J_{\check{\theta}}}-\frac{\mathrm{Tr}\,\rho_\theta L_{\check{\theta}}}{J_{\check{\theta}}}\right)^2\right)^{-1}+\delta_3\right)\right) \tag{H.10}$$

hold, where (H.10) follows from fact (iii). The uniformity of (H.1) (fact (iii)) and the boundness of the rhs of (H.1) (assumption 3) guarantee that the choice of $\epsilon>0$ is independent of $\theta,\check{\theta}$. From (H.9) and (H.10), we obtain (H.4) because the function $x\mapsto g(x,y)$ where $y,x\geqslant 0$.

Appendix I. Proof of theorem 13

If the true state is ρ_{θ_1}, the inequalities

$$\mathrm{P}_{\theta_1}^{M_{\theta_1}^{w,n}}\{T_n\notin U_{\theta_1,\epsilon}\}\leqslant \mathrm{P}_{\theta_1}^{M_f\times\sqrt{n}}\{\check{\theta}\notin U_{\theta_1,\epsilon}\}\sup_{\check{\theta}\notin U_{\theta_1,\epsilon}}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}\left\{\mathrm{e}^{n(1-\delta_{n-\sqrt{n}})D(\check{\theta}\|\theta_1)}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)<\mathrm{P}_{\check{\theta}}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)\right\}$$
$$\leqslant 1\times\sup_{\check{\theta}\notin U_{\theta_1,\epsilon}}\mathrm{e}^{-n(1-\delta_{n-\sqrt{n}})D(\check{\theta}\|\theta_1)}$$

hold. Since $(1-\delta_{n-\sqrt{n}})\to 1$, we have

$$\lim_{n\to\infty}-\frac{1}{n}\log\mathrm{P}_{\theta_1}^{M_{\theta_1}^{w,n}}\{T_n\notin U_{\theta_1,\epsilon}\}=\inf_{\check{\theta}\notin U_{\theta_1,\epsilon}}D(\check{\theta}\|\theta_1).$$

Thus, equation (56) is proved. Then, it implies (57).

Next, we show the weak consistency of $\vec{M}_{\theta_1}^{w}$. Assume that the true state ρ_θ is not ρ_{θ_1}. Then, we have

$$\mathrm{P}_{\theta}^{M_{\theta_1}^{w,n}}\{T_n\notin U_{\theta,\epsilon_n}\}\leqslant \mathrm{P}_{\theta}^{M_f\times\sqrt{n}}\{\check{\theta}\notin U_{\theta,\epsilon_n}\}+\mathrm{P}_{\theta}^{M_f\times\sqrt{n}}\{\check{\theta}\in U_{\theta,\epsilon_n}\}$$
$$\times\sup_{\check{\theta}\in U_{\theta,\epsilon_n}}\mathrm{P}_{\theta}^{E_{\theta_1}^{n-\sqrt{n}}}\left\{\mathrm{e}^{n(1-\delta_{n-\sqrt{n}})D(\check{\theta}\|\theta_1)}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)\geqslant\mathrm{P}_{\check{\theta}}^{E_{\theta_1}^{n-\sqrt{n}}}(\omega)\right\} \tag{I.1}$$

where $\epsilon_n := \frac{D(\theta\|\theta_1)}{2\left|\mathrm{Tr}\,\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}(\log\rho_\theta-\log\rho_{\theta_1})\right|}\delta_n$. Since $\delta_n=\frac{1}{n^{\frac{1}{5}}}$, the convergence $\mathrm{P}_{\theta}^{M_f\times\sqrt{n}}\{\check{\theta}\notin U_{\theta,\epsilon_n}\}\to 0$ holds. Also, the relation $U_{\theta,\epsilon_n}\subset U_{\theta,\epsilon_{n-\sqrt{n}}}$ holds. If we can prove

$$\sup_{\check{\theta}\in U_{\theta,\epsilon_n}}\mathrm{P}_{\theta}^{E_{\theta_1}^{n}}\left\{\mathrm{e}^{n(1-\delta_n)D(\check{\theta}\|\theta_1)}\mathrm{P}_{\theta_1}^{E_{\theta_1}^{n}}(\omega)\geqslant\mathrm{P}_{\check{\theta}}^{E_{\theta_1}^{n}}(\omega)\right\}\to 0 \tag{I.2}$$

we obtain

$$\mathrm{P}_\theta^{M_{\theta_1}^{w,n}}\{T_n \notin U_{\theta,\epsilon_n}\} \to 0. \tag{I.3}$$

This condition (I.3) is stronger than the weak consistency condition. Thus, it is sufficient to show (I.2).

From lemma 14, the relations

$$\begin{aligned}
&\mathrm{P}_\theta^{E_{\theta_1}^n}\left\{\mathrm{e}^{n(1-\delta_n)D(\check\theta\|\theta_1)}\mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega) \geqslant \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega)\right\} \\
&\quad= \mathrm{P}_\theta^{E_{\theta_1}^n}\left\{\frac{1}{n}\left(-\log \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) + \log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega)\right) + D(\check\theta\|\theta_1) \geqslant \delta_n D(\check\theta\|\theta_1)\right\} \\
&\quad= \mathrm{P}_\theta^{E_{\theta_1}^n}\left\{\frac{1}{n}\left(-\log \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) + \log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega)\right) + \mathrm{Tr}\,\rho_\theta\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right)\right. \\
&\qquad\left.\geqslant \delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right)\right\} \\
&\quad\leqslant \mathrm{P}_\theta^{E_{\theta_1}^n}\left\{-\frac{1}{n}\log \mathrm{P}_{\check\theta}^{E_{\theta_1}^n}(\omega) + \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta} \geqslant \delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})\right. \\
&\qquad\left.\times\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right)\right\} + \mathrm{P}_\theta^{E_{\theta_1}^n}\left\{\frac{1}{n}\log \mathrm{P}_{\theta_1}^{E_{\theta_1}^n}(\omega) - \mathrm{Tr}\,\rho_\theta\log\rho_{\theta_1} \geqslant \delta_n D(\check\theta\|\theta_1)\right. \\
&\qquad\left.+\mathrm{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right)\right\} \\
&\quad\leqslant \exp\left[-\left(n\sup_{0\leqslant t\leqslant 1}(\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right) - \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta})t\right.\right. \\
&\qquad\left.\left. - t\frac{(k+1)\log(n+1)}{n} - \log\mathrm{Tr}\,\rho_\theta\rho_{\check\theta}^{-t}\right)\right] + \exp\left[-\left(n\sup_{0\leqslant t}\left(\delta_n D(\check\theta\|\theta_1)\right.\right.\right. \\
&\qquad\left.\left.\left. + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right) + \mathrm{Tr}\,\rho_\theta\log\rho_{\theta_1}\right)t - \log\mathrm{Tr}\,\rho_\theta\rho_{\theta_1}^{t}\right)\right]
\end{aligned} \tag{I.4}$$

hold. In the following, we assume that $|\theta - \check\theta| \leqslant \epsilon_n$. Since $\epsilon_n = \frac{D(\theta\|\theta_1)}{2\left|\mathrm{Tr}\frac{\mathrm{d}\rho_\theta}{\mathrm{d}\theta}(\log\rho_\theta - \log\rho_{\theta_1})\right|}\delta_n$, we can derive $\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) \leqslant \frac{1}{2}D(\theta\|\theta_1)\delta_n + O\left(\delta_n^2\right)$. Substituting $t = s\delta_n$, we have

$$\begin{aligned}
&\sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{n\delta_n^2}\left(n\sup_{0\leqslant t\leqslant 1}\left(\delta_n D(\check\theta\|\theta_1) + \mathrm{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right) - \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta}\right)t\right. \\
&\qquad\left. - t\frac{(k+1)\log(n+1)}{n} - \log\mathrm{Tr}\,\rho_\theta\rho_{\check\theta}^{-t}\right) \\
&\quad\geqslant \sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{\delta_n^2}\left(\left(\frac{1}{2}D(\theta\|\theta_1)\delta_n + O\left(\delta_n^2\right) - \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta}\right)s\delta_n - s\delta_n\frac{(k+1)\log(n+1)}{n}\right. \\
&\qquad\left. + \mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta}s\delta_n - \frac{1}{2}(\mathrm{Tr}\,\rho_\theta(\log\rho_{\check\theta})^2 - (\mathrm{Tr}\,\rho_\theta\log\rho_{\check\theta})^2)s^2\delta_n^2 + O\left(\delta_n^3\right)\right) \\
&\quad\geqslant \sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{\delta_n^2}\left(\frac{1}{2}D(\theta\|\theta_1)s\delta_n^2 + O\left(\delta_n^3\right) - s\delta_n\frac{(k+1)\log(n+1)}{n}\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{1}{2}(\operatorname{Tr}\rho_\theta(\log\rho_{\check\theta})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\check\theta})^2)s^2\delta_n^2 + O\left(\delta_n^3\right)\Bigg) \\
&\to \frac{1}{2}D(\theta\|\theta_1)s - \frac{1}{2}(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)s^2 \qquad (\text{as } n\to\infty) \\
&= -\frac{1}{2}\left(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2\right) \\
&\quad \times\left(s - \frac{D(\theta\|\theta_1)}{2(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)}\right)^2 \\
&\quad + \frac{D(\theta\|\theta_1)^2}{8(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\lim_{n\to\infty}\sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{n\delta_n^2}\Bigg(n\sup_{0\leqslant t\leqslant 1}\Bigg((\delta_n D(\check\theta\|\theta_1) + \operatorname{Tr}(\rho_\theta - \rho_{\check\theta})(\log\rho_{\check\theta} - \log\rho_{\theta_1}) - \operatorname{Tr}\rho_\theta\log\rho_{\check\theta})t \\
&\qquad - t\frac{(k+1)\log(n+1)}{n} - \log\operatorname{Tr}\rho_\theta\rho_{\check\theta}^{-t}\Bigg)\Bigg) \\
&\qquad \geqslant \frac{D(\theta\|\theta_1)^2}{8(\operatorname{Tr}\rho_\theta(\log\rho_\theta)^2 - (\operatorname{Tr}\rho_\theta\log\rho_\theta)^2)} > 0. \qquad \text{(I.5)}
\end{aligned}
$$

Also, we obtain

$$
\begin{aligned}
&\sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{n\delta_n^2}\Bigg(n\sup_{0\leqslant t}\Bigg(\left(\delta_n D(\check\theta\|\theta_1) + \operatorname{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right) + \operatorname{Tr}\rho_\theta\log\rho_{\theta_1}\right)t \\
&\qquad - \log\operatorname{Tr}\rho_\theta\rho_{\theta_1}^t\Bigg)\Bigg) \\
&\qquad \geqslant \sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{\delta_n^2}\Bigg(\left(\frac{1}{2}D(\theta\|\theta_1)\delta_n + O\left(\delta_n^2\right) + \operatorname{Tr}\rho_\theta\log\rho_{\theta_1}\right)s\delta_n - \operatorname{Tr}\rho_\theta\log\rho_{\theta_1}s\delta_n \\
&\qquad - \frac{1}{2}\left(\operatorname{Tr}\rho_\theta\left(\log\rho_{\theta_1}\right)^2 - \left(\operatorname{Tr}\rho_\theta\log\rho_{\theta_1}\right)^2\right)s^2\delta_n^2 + O\left(\delta_n^3\right)\Bigg) \\
&\qquad = \sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{\delta_n^2}\Bigg(\left(\frac{1}{2}D(\theta\|\theta_1)s - \frac{1}{2}\left(\operatorname{Tr}\rho_\theta(\log\rho_{\theta_1})^2 - \left(\operatorname{Tr}\rho_\theta\log\rho_{\theta_1}\right)^2\right)s^2\right)\delta_n^2 \\
&\qquad + O\left(\delta_n^3\right)\Bigg) \\
&\qquad \to \frac{1}{2}D(\theta\|\theta_1)s - \frac{1}{2}\left(\operatorname{Tr}\rho_\theta\left(\log\rho_{\theta_1}\right)^2 - \left(\operatorname{Tr}\rho_\theta\log\rho_{\theta_1}\right)^2\right)s^2 \qquad (\text{as } n\to\infty).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\lim_{n\to\infty}\sup_{\check\theta\in U_{\theta,\epsilon_n}}\frac{1}{n\delta_n^2}\Bigg(n\sup_{0\leqslant t}\left(\left(\delta_n D(\check\theta\|\theta_1) + \operatorname{Tr}(\rho_\theta - \rho_{\check\theta})\left(\log\rho_{\check\theta} - \log\rho_{\theta_1}\right) + \operatorname{Tr}\rho_\theta\log\rho_{\theta_1}\right)t\right. \\
&\qquad \left. - \log\operatorname{Tr}\rho_\theta\rho_{\theta_1}^t\right)\Bigg) \geqslant \frac{D(\theta\|\theta_1)^2}{8(\operatorname{Tr}\rho_\theta(\log\rho_{\theta_1})^2 - (\operatorname{Tr}\rho_\theta\log\rho_{\theta_1})^2)} > 0. \qquad \text{(I.6)}
\end{aligned}
$$

Since $n\delta_n^2\to\infty$, relation (I.2) follows from (I.4)–(I.6).

Appendix J. Pinching map and group-theoretical viewpoint

J.1. Pinching map in non-asymptotic setting

In the following, we prove lemma 14 and construct the PVM E_θ^n after some discussions concerning the pinching map in the non-asymptotic setting and group representation theory. In this subsection, we present some definitions and discussions of the non-asymptotic setting.

A state ρ is called *commutative* with a PVM $E(=\{E_i\})$ on $\mathcal{H}$ if $\rho E_i = E_i \rho$ for any index i. For PVMs $E(=\{E_i\}_{i\in I})$, $F(=\{F_j\}_{j\in J})$, the notation $E \leqslant F$ means that for any index $i \in I$ there exists a subset $(F/E)_i$ of the index set J such that $E_i = \sum_{j\in(F/E)_i} F_j$. For a state ρ, we denote by $E(\rho)$ the spectral measure of ρ which can be regarded as a PVM. The pinching map $\mathcal{E}_E$ with respect to a PVM E is defined as

$$\mathcal{E}_E : \rho \mapsto \sum_i E_i \rho E_i \tag{J.1}$$

which is an affine map from the set of states to itself. Note that the state $\mathcal{E}_E(\rho)$ is commutative with a PVM E. If a PVM $F = \{F_j\}_{j\in J}$ is commutative with a PVM $E = \{E_i\}_{i\in I}$, we can define the PVM $F \times E = \{F_j E_i\}_{(i,j)\in I\times J}$, which satisfies $F \times E \geqslant E$ and $F \times E \geqslant F$. For any PVM E, the supremum of the dimension of E_i is denoted by $w(E)$.

Lemma 22. *Let E be a PVM such that $w(E) < \infty$. If states σ and ρ are commutative with the PVM E, and if a PVM F satisfies $E \leqslant F$, $E(\sigma) \leqslant F$, then we have*

$$D(\rho\|\sigma) - \log w(E) \leqslant D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) \leqslant D(\rho\|\sigma).$$

This lemma follows from lemmas 23 and 24 below.

Lemma 23. *Let ρ and σ be states. If a PVM F satisfies $E(\sigma) \leqslant F$, then*

$$D(\rho\|\sigma) = D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) + D(\rho\|\mathcal{E}_F(\rho)). \tag{J.2}$$

Proof. Since $E(\sigma) \leqslant F$ and F is commutative with σ, we have $\operatorname{Tr}\mathcal{E}_F(\rho)\log\mathcal{E}_F(\sigma) = \operatorname{Tr}\rho\log\sigma$. Since ρ is commutative with $\log\rho$, we have $\operatorname{Tr}\mathcal{E}_F(\rho)\log\rho = \operatorname{Tr}\rho\log\rho$. Therefore, we obtain the following:

$$\begin{aligned} D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) - D(\rho\|\sigma) &= \operatorname{Tr}\mathcal{E}_F(\rho)(\log\mathcal{E}_F(\rho) - \log\mathcal{E}_F(\sigma)) - \operatorname{Tr}\rho(\log\rho - \log\sigma) \\ &= \operatorname{Tr}\mathcal{E}_F(\rho)(\log\mathcal{E}_F(\rho) - \log\rho). \end{aligned}$$

This proves (J.2). □

Lemma 24. *Let E and F be PVMs such that $E \leqslant F$. If a state ρ is commutative with E, we have*

$$D(\rho\|\mathcal{E}_F(\rho)) \leqslant \log w(E). \tag{J.3}$$

Proof. Let $a_i := \operatorname{Tr}E_i\rho E_i$ and $\rho_i := \frac{1}{a_i}E_i\rho E_i$. Then, we have $\rho = \sum_i a_i\rho_i$, $\mathcal{E}_F(\rho) = \sum_i a_i\mathcal{E}_F(\rho_i)$, $\sum_i a_i = 1$. Therefore,

$$\begin{aligned} D(\rho\|\mathcal{E}_F(\rho)) &= \sum_i \operatorname{Tr}E_i\rho(\log\rho - \log\mathcal{E}_F(\rho)) \\ &= \sum_i \operatorname{Tr}E_i\rho E_i(E_i\log\rho E_i - E_i\log\mathcal{E}_F(\rho)E_i) \\ &= \sum_i a_i D(\rho_i\|\mathcal{E}_F(\rho_i)) \leqslant \sup_i D(\rho_i\|\mathcal{E}_F(\rho_i)) \\ &= \sup_i(\operatorname{Tr}\rho_i\log\rho_i - \operatorname{Tr}\mathcal{E}_F(\rho_i)\log\mathcal{E}_F(\rho_i)) \end{aligned}$$

$$\leqslant -\sup_i \operatorname{Tr} \mathcal{E}_F(\rho_i) \log \mathcal{E}_F(\rho_i) \leqslant \sup_i \log \dim E_i = \log w(E).$$

Thus, we obtain inequality (J.3). □

Let us consider another type of inequality.

Lemma 25. *Let E be a PVM such that $w(E) < \infty$. If the state ρ is commutative with E, and if a PVM M satisfies that $M \geqslant E$, we have*

$$\rho \leqslant \mathcal{E}_M(\rho) w(E) \tag{J.4}$$

$$\rho^{-t} \geqslant \mathcal{E}_M(\rho)^{-t} w(E)^{-t} \tag{J.5}$$

for $1 \leqslant t \leqslant 0$.

Proof. It is sufficient for (J.4) to show

$$\rho \leqslant k \mathcal{E}_M(\rho) \tag{J.6}$$

for any state ρ and any PVM M on a k-dimensional Hilbert space $\mathcal{H}$. Now, it is sufficient to prove (J.6) in the pure state case. For any $\phi, \psi \in \mathcal{H}$, we have

$$\langle \psi | k \mathcal{E}_M(|\phi\rangle\langle\phi|) - |\phi\rangle\langle\phi||\psi\rangle = k \sum_{i=1}^k \langle\psi|M_i|\phi\rangle\langle\phi|M_i|\psi\rangle - \left| \sum_{i=1}^k \langle\psi|M_i|\phi\rangle \right|^2 \geqslant 0.$$

The last inequality follows from Schwartz inequality for vectors $\{\langle\psi|M_i|\phi\rangle\}_{i=1}^k$ and $\{1\}_{i=1}^k$. It is well known that the function $u \mapsto -u^{-t}$ $(0 \leqslant t \leqslant 1)$ is an operator monotone function [40]. Thus, (J.4) implies (J.5).

Lemma 26. *If a PVM M is commutative with a state σ and $w(M) = 1$, we have*

$$\mathrm{P}_\rho^M \left\{ \log \mathrm{P}_\sigma^M(\omega) \geqslant a \right\} \leqslant \exp\left(-\sup_{0 \leqslant t} (at - \log \operatorname{Tr} \rho \sigma^t) \right) \tag{J.7}$$

for any state ρ.

Proof. From Markov's inequality, we have

$$p\{X \geqslant a\} \leqslant \exp -\Lambda_t(X, p, a) \qquad \Lambda_t(X, p, a) := at - \log \int \mathrm{e}^{tX(\omega)} p(\mathrm{d}\omega). \tag{J.8}$$

Since $w(M) = 1$, the relation $\sum_\omega \mathrm{P}_\rho^M(\omega) \mathrm{P}_\sigma^M(\omega)^t = \operatorname{Tr} \mathcal{E}_M(\rho) \mathcal{E}_M(\sigma)^t$ holds. It yields

$$\Lambda_t \left(\log \mathrm{P}_\sigma^M, \mathrm{P}_\rho^M, a \right) = at - \log \operatorname{Tr} \mathcal{E}_M(\rho) \mathcal{E}_M(\sigma)^t = at - \log \operatorname{Tr} \rho \sigma^t.$$

Thus, we obtain (J.7). □

Lemma 27. *Assume that E and M are PVMs such that $w(E) < \infty$, $w(M) = 1$ and $M \geqslant E$. If the states ρ and ρ' are commutative with E, we have*

$$\mathrm{P}_\rho^M \left\{ -\log \mathrm{P}_{\rho'}^M(\omega) \geqslant a \right\} \leqslant \exp\left(-\sup_{0 \leqslant t \leqslant 1} ((a - \log w(E))t - \log \operatorname{Tr} \rho \rho'^{-t}) \right). \tag{J.9}$$

Proof. If $0 \leqslant t \leqslant 1$, we have

$$\begin{aligned} \Lambda_t\left(-\log \mathrm{P}_{\rho'}^M, \mathrm{P}_\rho^M, a\right) &= at - \log \operatorname{Tr} \mathcal{E}_M(\rho) \mathcal{E}_M(\rho')^{-t} = at - \log \operatorname{Tr} \rho \mathcal{E}_M(\rho')^{-t} \\ &\geqslant at - \log w(E)^t \operatorname{Tr} \rho \rho'^{-t} \qquad \text{(J.10)} \\ &\geqslant (a - \log w(E))t - \log \operatorname{Tr} \rho \rho'^{-t} \qquad \text{(J.11)} \end{aligned}$$

where (J.10) follows from lemma 25. Therefore, from (J.8) and (J.11), we obtain (J.9). □

J.2. Group representation and its irreducible decomposition

In this subsection, we consider the relation between irreducible representations and PVMs for the purpose of constructing the PVM E_θ^n and a proof of lemma 14. Let V be a finite-dimensional vector space over the complex numbers $\mathbb{C}$. A map π from a group G to the generalized linear group of a vector space V is called a *representation* on V if the map π is homomorphic, i.e. $\pi(g_1)\pi(g_2) = \pi(g_1 g_2), \forall g_1, g_2 \in G$. The subspace W of V is called *invariant* with respect to a representation π if the vector $\pi(g)w$ belongs to the subspace W for any vector $w \in W$ and any element $g \in G$. The representation π is called *irreducible* if there is no proper nonzero invariant subspace of V with respect to π. Let π_1 and π_2 be representations of a group G on V_1 and V_2, respectively. The *tensored* representation $\pi_1 \otimes \pi_2$ of G on $V_1 \otimes V_2$ is defined as $(\pi_1 \otimes \pi_2)(g) = \pi_1(g) \otimes \pi_2(g)$, and the *direct sum* representation $\pi_1 \oplus \pi_2$ of G on $V_1 \oplus V_2$ is also defined as $(\pi_1 \oplus \pi_2)(g) = \pi_1(g) \oplus \pi_2(g)$.

In the following, we treat a representation π of a group G on a finite-dimensional Hilbert space $\mathcal{H}$. The following fact is crucial in later arguments. There exists an irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ such that the irreducible components are orthogonal to one another if for any element $g \in G$ there exists an element $g^* \in G$ such that $\pi(g)^* = \pi(g^*)$, where $\pi(g)^*$ denotes the adjoint of the linear map $\pi(g)$. We can regard the irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ as the PVM $\left\{P_{\mathcal{H}_i}\right\}_{i=1}^l$, where $P_{\mathcal{H}_i}$ denotes the projection to $\mathcal{H}_i$. If two representations π_1 and π_2 satisfy the preceding condition, the tensored representation $\pi_1 \otimes \pi_2$ also satisfies it. Note that in general, an irreducible decomposition of a representation satisfying the preceding condition is not unique. In other words, we cannot uniquely define the PVM from such a representation.

J.3. Construction of PVM E_θ^n and the tensored representation

In this subsection, we construct the PVM E_θ^n after the discussion of the tensored representation. Let the dimension of the Hilbert space $\mathcal{H}$ be k. Concerning the natural representation $\pi_{\mathrm{SL}(\mathcal{H})}$ of the special linear group $\mathrm{SL}(\mathcal{H})$ on $\mathcal{H}$, we consider its nth tensored representation $\pi_{\mathrm{SL}(\mathcal{H})}^{\otimes n} := \underbrace{\pi_{\mathrm{SL}(\mathcal{H})} \otimes \cdots \otimes \pi_{\mathrm{SL}(\mathcal{H})}}_{n}$ on the tensored space $\mathcal{H}^{\otimes n}$. For any element $g \in \mathrm{SL}(\mathcal{H})$, the relation $\pi_{\mathrm{SL}(\mathcal{H})}(g)^* = \pi_{\mathrm{SL}(\mathcal{H})}(g^*)$ holds where the element $g^* \in \mathrm{SL}(\mathcal{H})$ denotes the adjoint matrix of the matrix g. Consequently, there exists an irreducible decomposition of $\pi_{\mathrm{SL}(\mathcal{H})}^{\otimes n}$ regarded as a PVM and we denote the set of such PVMs by $Ir^{\otimes n}$.

From Weyl's dimension formula ((7.1.8) or (7.1.17) in Weyl [41] and Goodman and Wallach [42]), the nth symmetric tensored space is the maximum-dimensional space in the irreducible subspaces with respect to the nth tensored representation $\pi_{\mathrm{SL}(\mathcal{H})}^{\otimes n}$. Its dimension equals the repeated combination ${}_k H_n$ evaluated by ${}_k H_n = \binom{n+k-1}{k-1} = \binom{n+k-1}{n} = {}_{n+1} H_{k-1} \leqslant (n+1)^{k-1}$. Thus, any element $E^n \in Ir^{\otimes n}$ satisfies

$$w(E^n) \leqslant (n+1)^{k-1}. \tag{J.12}$$

Lemma 28. *A PVM $E^n \in Ir^{\otimes n}$ is commutative with the nth tensored state $\rho^{\otimes n}$ of any state ρ on $\mathcal{H}$.*

Proof. If $\det \rho \neq 0$, this lemma is trivial based on the fact that $\det(\rho)^{-1}\rho \in \mathrm{SL}(\mathcal{H})$. If $\det \rho = 0$, there exists a sequence $\{\rho_i\}_{i=1}^\infty$ such that $\det \rho_i \neq 0$ and $\rho_i \to \rho$ as $i \to \infty$. We have $\rho_i^{\otimes n} \to \rho^{\otimes n}$ as $i \to \infty$. Because a PVM $E^n \in Ir^{\otimes n}$ is commutative with $\rho_i^{\otimes n}$, it is also commutative with $\rho^{\otimes n}$. □

Definition 29. *We can define the PVM $E^n \times E(\rho^{\otimes n})$ for any PVM $E^n \in Ir^{\otimes n}$. Now we define the PVM E^n_θ satisfying $w\left(E^n_\theta\right) = 1$, $E^n_\theta \geqslant E^n \times E\left(\rho^{\otimes n}_\theta\right)$ for a PVM $E^n \in Ir^{\otimes n}$. Note that the E^n_θ is not unique.*

Proof of lemma 14. From lemmas 26 and 27, (J.12) and the definition of E^n_θ, we obtain lemma 14. □

Proof of lemma 19. From lemma 22, (J.12) and the definition of E^n_θ, we obtain lemma 19. □

Appendix K. Large deviation theory for an exponential family

In this section, we summarize the large deviation theory for an exponential family. A d-dimensional probability family is called an exponential family if there exist linearly independent real-valued random variables $F_1, \ldots, F_d$ and a probability distribution p on the probability space Ω such that the family consists of the probability distribution

$$p_\theta(\mathrm{d}\omega) := \exp\left(\sum_{i=1}^d \theta^i F_i(\omega) - \psi(\theta)\right) p(\mathrm{d}\omega)$$

$$\psi(\theta) := \log \int_\Omega \exp\left(\sum_{i=1}^d \theta^i F_i(\omega)\right) p(\mathrm{d}\omega).$$

In this family, the parametric space is given by $\Theta := \{\theta \in \mathbb{R}^d | 0, < \psi(\theta) < \infty\}$, the parameter θ is called the natural parameter and the function $\psi(\theta)$ is called the potential. We define the dual potential $\phi(\theta)$ and the dual parameter $\eta(\theta)$, called the expectation parameter, as

$$\eta_i(\theta) := \frac{\partial \psi(\theta)}{\partial \theta^i} = \log \int_\Omega F_i(\omega) p_\theta(\mathrm{d}\omega) \qquad \phi(\theta) := \max_{\theta'} \left(\sum_{i=1}^d \theta'^i \eta_i(\theta) - \psi(\theta')\right).$$

From (K.1), we have

$$\phi(\theta) = \sum_{i=1}^d \theta^i \eta_i(\theta) - \psi(\theta).$$

In this family, the sufficient statistics are given by $F_1(\omega), \ldots, F_d(\omega)$. The MLE $\hat{\theta}(\omega)$ is given by $\eta_i(\hat{\theta}(\omega)) = F_i(\omega)$. The KL divergence $D(\theta \| \theta_0) := D\left(p_\theta \middle\| p_{\theta_0}\right)$ is calculated by

$$\begin{aligned} D(\theta \| \theta_0) &= \int_\Omega \log \frac{p_\theta(\omega)}{p_{\theta_0}(\omega)} p_\theta(\mathrm{d}\omega) = \int_\Omega \sum_i \left(\theta^i - \theta^i_0\right) F_i(\omega) + \psi(\theta_0) - \psi(\theta) p_\theta(\mathrm{d}\omega) \\ &= \sum_i \left(\theta^i - \theta^i_0\right) \eta_i(\omega) + \psi(\theta_0) - \psi(\theta) = \phi(\theta) + \psi(\theta_0) - \sum_i \theta^i_0 \eta_i(\omega) \\ &= \max_{\theta'} \left(\sum_i \theta'^i \eta_i(\theta) - \psi(\theta')\right) + \psi(\theta_0) - \sum_i \theta^i_0 \eta_i(\theta) \\ &= \max_{\theta'} \sum_i (\theta'^i - \theta'^{i_0}) \eta_i(\theta) - \log \int_\Omega \exp\left(\sum_i \left(\theta^i - \theta^i_0\right) F_i(\omega)\right) p_\theta(\mathrm{d}\omega). \end{aligned}$$

Next, we discuss the n-i.i.d. extension of the family $\{p_\theta | \theta \in \Theta\}$. For the data $\vec{\omega}_n := (\omega_1, \ldots, \omega_n) \in \Omega^n$, the probability distribution $p^n_\theta(\vec{\omega}_n) := p_\theta(\omega_1) \cdots p_\theta(\omega_n)$ is

given by

$$p_\theta^n(\vec{\omega}_n) = \exp\left(n\sum_i \theta^i F_{n,i}(\vec{\omega}_n) - n\psi(\theta)\right) p^n(\mathrm{d}\vec{\omega}_n)$$
$$p^n(\mathrm{d}\vec{\omega}_n) := p(\mathrm{d}\omega_1)\cdots p(\mathrm{d}\omega_n)$$
$$F_{n,i}(\vec{\omega}_n) := \frac{1}{n}\sum_{k=1}^n F_i(\omega_k).$$

Since the expectation parameter of the probability family $\left\{p_\theta^n \middle| \theta \in \Theta\right\}$ is given by $n\eta_i(\theta)$, the MLE $\hat{\theta}_n(\vec{\omega}_n)$ is given by

$$n\eta_i(\hat{\theta}_n(\vec{\omega}_n)) = nF_{n,i}(\vec{\omega}_n). \tag{K.1}$$

Applying Cramér's theorem [36] to the random variables $F_1, \ldots, F_d$ and the distribution p_{θ_0}, for any subset $S \subset \mathbb{R}^d$ we have

$$\inf_{\eta\in\overline{S}} \sup_{\theta'\in\mathbb{R}^d} \left(\sum_i \theta'^i \left(\eta_i - \mathrm{E}_{\theta_0}(F_i)\right) - \psi_{\theta_0}(\theta')\right) \leqslant \lim_{n\to\infty} \frac{-1}{n} \log p_{\theta_0}^n\{\vec{F}_n \in S\}$$
$$\leqslant \inf_{\eta\in\mathrm{int}S} \sup_{\theta'\in\mathbb{R}^d} \left(\sum_i \theta'^i \left(\eta_i - \mathrm{E}_{\theta_0}(F_i)\right) - \psi_{\theta_0}(\theta')\right)$$

where

$$\mathrm{E}_{\theta_0}(F_i) := \int_\Omega F_i(\omega)\, p_\theta(\mathrm{d}\omega)$$
$$\psi_{\theta_0}(\theta) := \int_\Omega \exp\left(\sum_i \theta^i F_i(\omega)\right) p_\theta(\mathrm{d}\omega)$$
$$\vec{F}_n(\vec{\omega}_n) := (F_{n,1}(\vec{\omega}_n), \ldots, F_{n,d}(\vec{\omega}_n))$$

and int S denotes the interior of S, which is consistent with $(\overline{S^c})^c$. Since

$$\sup_{\theta'\in\mathbb{R}^d} \left(\sum_i \theta'^i \left(\eta_i - \mathrm{E}_{\theta_0}(F_i)\right) - \psi_{\theta_0}(\theta')\right)$$
$$= \sup_{\theta'\in\mathbb{R}^d} \left(\sum_i \theta'^i \left(\eta_i - \eta_i(\theta_0)\right) - \psi(\theta')\right) + \psi(\theta_0) = D(\theta\|\theta_0)$$

and the map $\theta \mapsto D(\theta\|\theta_0)$ is continuous, it follows from (K.1) that

$$\lim_{n\to\infty} \frac{-1}{n} \log p_{\theta_0}^n\{\hat{\theta}_n \in \Theta'\} = \inf_{\theta\in\Theta'} D(\theta\|\theta_0)$$

for any subset $\Theta' \subset \Theta$, which is equivalent to (76). Conversely, if an estimator $\{T_n(\vec{\omega}_n)\}$ satisfies the weak consistency

$$\lim_{n\to\infty} p_\theta^n\{\|T_n(\vec{\omega}_n) - \theta\| > \epsilon\} \to 0 \qquad \forall \epsilon > 0 \quad \forall \theta \in \Theta$$

then, similar to (33), we can prove

$$\lim_{n\to\infty} \frac{-1}{n} \log p_{\theta_0}^n\{T_n(\vec{\omega}_n) \in \Theta'\} \leqslant \inf_{\theta\in\Theta'} D(\theta\|\theta_0).$$

Therefore, we can conclude that the MLE is optimal in the large deviation sense for exponential families.

References

[1] Amari S and Nagaoka H 2000 *Methods of Information Geometry* (Providence, RI/Oxford: AMS/Oxford University Press)
[2] Petz D 1996 *Linear Algebr. Appl.* **224** 81
[3] Petz D and Toth G 1993 *Lett. Math. Phys.* **27** 205
[4] Petz D and Sudár C 1998 Extending the Fisher metric to density matrices *Geometry in Present Day Science* vol 21, ed O E Barndorff-Nielsen and E B V Jensen (Singapore: World Scientific)
[5] Holevo A S 1998 *IEEE Trans. Inform. Theory* **44** 269 (*Preprint* quant-ph/9611023)
[6] Schumacher B and Westmoreland M D 1997 *Phys. Rev.* A **56** 131
[7] Schumacher B 1995 *Phys. Rev.* A **51** 2738
[8] Winter A 2000 Coding theorems of quantum information theory *PhD Dissertation* Uni Bielefeld (*Preprint* quant-ph/9907077)
[9] Hayashi M 2002 Exponents of quantum fixed-length pure state source coding *Phys. Rev.* A (to be published) (*Preprint* quant-ph/0202002)
[10] Hiai F and Petz D 1991 *Commun. Math. Phys.* **143** 99–114
[11] Ogawa T and Nagaoka H 2000 *IEEE Trans. Inform. Theory* **46** 2428 (*Preprint* quant-ph/9808063)
[12] Kubo R, Toda M and Hashitsume N 1991 *Statistical Physics II, Nonequilibrium Statistical Mechanics (Springer Series on Solid State Sciences)* (Berlin: Springer)
[13] Helstrom C W 1967 *Phys. Lett.* A **25** 101
[14] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[15] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[16] Gill R and Massar S 2000 *Phys. Rev.* A **61** 042312
[17] Hayashi M 1997 A linear programming approach to attainable Cramer–Rao type bound *Quantum Communication, Computing, and Measurement* ed O Hirota, A S Holevo and C M Caves (New York: Plenum) pp 99–108
[18] Hayashi M 1997 *Preprint* quant-ph/9704044
[19] Bahadur R R 1960 *Sankhyā* **22** 229
[20] Bahadur R R 1967 *Ann. Math. Stat.* **38** 303
[21] Bahadur R R 1971 *Some Limit Theorems in Statistics (Regional Conf. Series in Applied Mathematics, no 4)* (Philadelphia, PA: SIAM)
[22] Nagaoka H 1988 An asymptotically efficient estimator for a one-dimensional parametric model of quantum statistical operators *Proc. 1988 IEEE Int. Symp. on Information Theory* p 198
[23] Nagaoka H 1992 On the relation Kullback divergence and Fisher information—from classical systems to quantum systems *Proc. Society Information Theory and its Applications in Japan* p 63 (in Japanese)
[24] Hayashi M and Matsumoto K 2002 Quantum universal variable-length source coding *Phys. Rev.* A (to be published) (*Preprint* quant-ph/0202001)
[25] Čencov N N 1982 *Statistical Decision Rules and Optimal Inference* (Providence, RI: American Mathematical Society) (Originally published in Russian 1972) (Moscow: Nauka)
[26] Fu J C 1973 *Ann. Stat.* **1** 745
[27] Hayashi M 2001 *J. Phys. A: Math. Gen.* **34** 3413 (*Preprint* quant-ph/9704040)
[28] Keyl K and Werner R F 2001 *Phys. Rev.* A **64** 052311 (*Preprint* quant-ph/0102027)
[29] Lindblad G 1975 *Commun. Math. Phys.* **40** 147
[30] Uhlmann A 1977 *Commun. Math. Phys.* **54** 21
[31] Uhlmann A 1993 *Rep. Math. Phys.* **33** 253–63
[32] Matsumoto K 1995 Geometry of quantum states *Master's Thesis* Tokyo University (in Japanese)
[33] LeCam L 1986 *Asymptotic Methods in Statistical Decision Theory* (New York: Springer)
[34] Jozsa R 1994 *J. Mod. Opt.* **41** 2315
[35] Fuchs C A 1995 Distinguishability and accessible information in quantum theory *PhD Dissertation* University of New Mexico (*Preprint* quant-ph/9601020)
[36] Bucklew J A 1990 *Large Deviation Techniques in Decision, Simulation and Estimation* (New York: Wiley)
[37] Hayashi M and Matsumoto K 2000 *IEICE J* **83** 629 (in Japanese)
[38] Hayashi M and Matsumoto K 1998 Statistical model with a option for measurements and quantum mechanics *RIMS Koukyuroku* **1055** 96–110 (in Japanese)
[39] Barndorff-Nielsen O E 1978 *Information and Exponential Families in Statistical Theory* (New York: Wiley)
[40] Bhatia R 1997 *Matrix Analysis* (New York: Springer)
[41] Weyl H 1939 *The Classical Groups, Their Invariants and Representations* (Princeton, NJ: Princeton University Press)
[42] Goodman R and Wallach N 1998 *Representations and Invariants of the Classical Groups* (Cambridge: Cambridge University Press)