数理解析研究所講究録
第2018巻 2017年 45-69

Quantum statistical decision theory with a restricted class of measurements.

大阪大学・基礎工学研究科　田中冬彦
Fuyuhiko Tanaka
Graduate School of Engineering Science
Osaka University

Abstract

Decades ago, Holevo established quantum statistical decision theory. However, this formulation arises from theoretical interests and still is unsatisfactory compared with the classical counterpart. From the viewpoint in modern statistics and aiming at more rich framework, we reformulate quantum statistical decision theory.

1 Introduction

Suppose that we have a quantum system described by a parametrized density operator, say, $\rho(\theta)$. Carefully prepared n-iid system, i.e., $\sigma(\theta) := \rho(\theta)^{\otimes n}$ is available as a source. Our main task is to perform a good measurement to the system $\sigma(\theta)$, to collect measurement outcome as data, and finally to estimate the unknown parameter θ based on the data. This is a typical setting of quantum estimation theory, where many authors have investigated. However, we introduce the following additional restriction.

In addition, suppose that we have only a restricted class of measurements, say, Γ.

From theoretical viewpoint, this assumption seems quite strange because most of all theoretical physicists only focus on fundamental accuracy, theoretical bound of estimation error. However, from a practical viewpoint, the above restriction is not

surprizing. For example, in a qubit experiment, only two-outcome projective valued-measurements for each system are available. In a poor laboratory, they can only prepare noisy measurements. In this case, such measurements are not projection any more. Our main concern here is how to formulate statistical inference with Γ given and how to find out the best/better strategy to obtain information on the microscopic system with Γ given.

In Section 2, we formulate our problems in a mathematical form. Most of them follows usual formulation in quantum statistical decision theory. However, we define additional concepts like implementable measurements. In Section 3, in our formulation, fundamental results are shown. An example is presented in Section 4.

2 Problem Setting

2.1 Density operator and positive-operator-valued measure

Let $\mathcal{H}$ be a separable Hilbert space. A linear operator on $\mathcal{H}$ is called a *density operator* if it is positive and of trace one. We assume that a prepared quantum system is described by a density operator ρ. Suppose that we perform a measurement for the prepared system. According to Holevo [14], any continuous probability distribution of a measurement outcome $\omega \in \Omega$ is given by the following form:

$$\omega \sim \mathrm{Tr}\rho\mathbf{M}(\mathrm{d}\omega),$$

where $\mathbf{M}$ is a map from Borel sets $\mathcal{B}(\Omega)$ to a positive operator on $\mathcal{H}$ satisfying

$$\mathbf{M}(\Omega) = I,\ \mathbf{M}(B) \geq 0, \forall B \in \mathcal{B}(\Omega),$$

$$\sum_{j=1}^{\infty} \mathbf{M}(B_j) = \mathbf{M}(\bigcup_{j=1}^{\infty} B_j),\ B_j \cap B_k = \emptyset \text{ if } j \neq k.$$

The map $\mathbf{M}$ is called a *positive-operator-valued measure (POVM)*. For the discrete probability distribution with Ω at most countably infinite, we may write

$$\omega \sim \mathrm{Tr}\rho\mathbf{M}_\omega,\ \sum_{\omega} \mathbf{M}_\omega = I, \mathbf{M}_\omega \geq 0, \forall \omega \in \Omega.$$

We symbolically write

$$\omega \sim \mathbf{M}(\mathrm{d}\omega)$$

even if a probability distribution does not have a continuous component. In physical experiments, there are various kinds of measurements like counting pulse, voltage, intensity of a current etc. Thus we also write a measurement as $(\mathbf{M}, \mathcal{B}(\chi))$, which implies that the outcome space is χ and the domain of $\mathbf{M}$ is the Borel sets $\mathcal{B}(\chi)$.

2.2 Decision space

Let Θ a compact metric space. We assume that $\sigma(\theta)$ as a map from Θ to the set of all density operators is one-to-one (identifiability) and regular [13]. The regularity is a slightly stronger condition than the continuity with respect to the trace norm. Usually the regularity is satisfied thus we omit the definition of the regularity. (See, e.g., Tanaka [23].) This map is called a *quantum statistical model.*

In statistical decision theory, statistical inference based on finite data (parameter estimation, hypothesis testing, etc.) is regarded as a *statistical decision.* For example, in parameter estimation, our decision is to determine an estimate of the unknown parameter θ based on a measurement outcome, say, $x \in X$. This decision is represented by a function from X to the parameter space Θ. For hypothesis testing, our decision whether we accept null hypothesis or reject depends on x. It is written as a function from X to a finite set $\{0, 1\}$. In this sense, any statistical inference as a decision is given by a function from the space of the measurement outcome to the set of possible decisions, which we call the *decision space.* This function is called a *decision function.* Generally we denote the decision space as U and assume that U is also a compact metric space.

2.3 Loss functions

In order to evaluate whether the decision, say $u(x)$ is better or not in a quantitative way, we introduce a loss function.

Definition 1. Let $w : \Theta \times U \to \mathbf{R} \cup \{+\infty\}$ be a lower semicontinuous function. We call w a *loss function* if it is bounded from below, $w(\theta, u) > -M > -\infty, \forall\theta, \forall u$ for a constant M.

For simplicity, we assume that $w(\theta, u) \geq 0$. One typical example of the loss function is the squared loss $w(\theta, u) = |\theta - u|^2$, which is often used in the parameter estimation problem.

Definition 2. For a quantum statistical model $\sigma(\theta)$, a decision space U and a loss function $w(\theta, u)$, we call the triplet (σ, U, w) a *quantum statistical decision problem (QSD problem)* .

Classical statistical inference is usually formulated as a statistical decision problem. Likewise, quantum statistical inference including quantum estimation, quantum state discrimination is formulated as a QSD problem. (See, e.g., references in Kumagai and Hayashi [16] for recent works in this direction, although they deal with non-Bayesian hypothesis testing.) Interestingly enough, quantum state cloning [2, 4] and benchmark fidelity [7, 8, 9, 18], which are purely physical topics, are also described in the framework (See, e.g., Tanaka [22] for the relation between quantum benchmark and quantum estimation of pure states).

In classical statistics, the observation x is a random variable and its distribution is given as a member of a parametric model $\{p(x|\theta)\}$. We only have to specify a

decision function $\delta : x \mapsto \delta(x)$. In the quantum setting, however, we have to specify a measurement over the quantum system, which is described by a POVM. After we fix a POVM, the distribution of the observation x is specified by this POVM and the density operator of the quantum system. Then, we have to specify a decision function.

2.4 Decision POVM

Let a QSD problem (σ, U, w) and some statistical procedures be given. Each statistical procedure is given by a measurement $(\mathbf{M}, \mathcal{B}(\chi))$ for the system $\sigma(\theta)$ (we assume that it is written in a mathematical form.) and a decision function $u(x)$. Suppose that we want to compare statistical procedures. The performance of each statistical procedure is evaluated through the risk function, a generalization of the average estimation error. As we shall see later, the risk function depends only on the POVM over the decision space U constructed in the following manner. (See, e.g., Helstrom [11], Holevo [12].)

As an example, we consider the parameter estimation. Suppose that we perform a finite-valued measurement described by the POVM over χ, $(|\chi| < \infty)$.

$$x \sim \mathbf{E}_x$$

and use a specific estimate (e.g. the maximum likelihood estimate) $u(x)$. Then, the POVM over χ and the estimator $u(x)$ yield the following POVM,

$$\mathbf{N}_u = \sum_{x:u(x)=u} \mathbf{E}_x, \forall u \in \Theta.$$

Clearly the set of $\mathbf{N}_u$ is the POVM over the parameter space Θ. If we consider the quantum state discrimination, we take a POVM over a finite set.

Thus, every statistical procedure yields a POVM over the decision space U, which we call a *decision POVM*. Let us denote $\mathcal{P}o(U)$ as the set of all decision POVMs.

2.5 Risk function and the average risk

In order to evaluate the performance of a statistical procedure, we consider the average loss function called the *risk function*. For simplicity, we consider finite-valued measurements. Naive definition is as follows:

$$R(\theta; \mathbf{E}, u(x)) := \sum_x w(\theta, u(x)) \mathrm{Tr}\sigma(\theta)\mathbf{E}_x.$$

We rewrite the risk function as

$$\begin{aligned} R(\theta; \mathbf{E}, u(x)) &= \sum_u \sum_{x:u(x)=u} w(\theta, u(x)) \mathrm{Tr}\sigma(\theta)\mathbf{E}_x \\ &= \sum_u w(\theta, u) \mathrm{Tr}\sigma(\theta) \left(\sum_{x:u(x)=u} \mathbf{E}_x \right) \\ &= \sum_u w(\theta, u) \mathrm{Tr}\sigma(\theta)\mathbf{N}_u, \end{aligned}$$

where in the last line we set $\mathbf{N}_u = \sum_{x:u(x)=u} \mathbf{E}_x$. As mentioned before, we see that the risk function depends only on the decision POVM $\mathbf{N} = \{\mathbf{N}_u\}$. Thus, we write the risk function as

$$R(\theta; \mathbf{N}) = \sum_u w(\theta, u) \mathrm{Tr}\sigma(\theta)\mathbf{N}_u.$$

Formal definition is given below.

Definition 3. For a QSD problem $(\sigma(\theta), U, w)$, the *risk function for the decision POVM* $\mathbf{N} \in \mathcal{P}o(U)$ is

$$R(\theta; \mathbf{N}) = \int_U w(\theta, u) \mathrm{Tr}\sigma(\theta)\mathbf{N}(\mathrm{d}u).$$

Main concern in all of statistical problems is to make the risk function smaller by choosing a good decision POVM. However, it is impossible to achieve the smallest risk for every θ. Thus, in the present article, we adopt the Bayesian optimality. That

is, we assume a prior distribution $\pi(\mathrm{d}\theta) \in \mathcal{P}(\Theta)$ over the parameter space Θ and consider the minimization of the following average.

$$R_\pi(\mathbf{M}) := \int R(\theta; \mathbf{M})\pi(\mathrm{d}\theta).$$

Until now, we basically follow the original work done by Holevo [13]. Now we mention our own new approach. First, all of previous works in Holevo's framework only focus on the theoretical value for the average risk, which is in our notation, written as

$$\inf\{R_\pi(\mathbf{M}) : \mathbf{M} \in \mathcal{P}o(U)\}.$$

Certainly, these works are so valuable in a mathematical sense and they bring us deep insight of quantum correlation in a physical sense. However, even if we obtain an explicit form of the optimal decision POVM, which is almost useless for practitioners. For example, when the quantum state is given by a tensor product $\sigma(\theta) = \rho(\theta)^{\otimes n}$, the optimal POVM uses full quantum correlation over the n-system, which is almost impossible to implement except for smaller n, say, $n = 1, 2$. Thus, this formulation itself is meaningful for theoretical analysis but not for practical purpose. Thinking of the usefulness in classical statistical decision theory, the quantum counterpart is not satisfactory.

Thus, we switch the strategy into more practical ones. We deal with the proper subset of decision POVMs $\mathcal{P} \subset \mathcal{P}o(U)$ and consider the following minimization problem.

$$\inf\{R_\pi(\mathbf{M}) : \mathbf{M} \in \mathcal{P}\}$$

However, it remains how we specify the proper subset $\mathcal{P} \subseteq \mathcal{P}o(U)$. Let us consider this in the next subsection.

2.6 Proper subset of decision POVMs

Suppose that we have a set of measurements Γ in a real experiment. Here we assume that each element in Γ is written in a mathematical form. We call Γ an *implementable class of measurements*. Theoretical analysis in quantum statistics should be done for each implementable class, which is the essence of our formulation. For example, in a poor laboratory, say, Tanaka-Labo (TL), they are able to prepare only noisy projective measurements (i.e., these measurements are not projective measurements). Then, Tanaka-Labo's implementable class Γ_{TL} consists of noisy projective measurements. Of course, detailed mathematical description of noisy projective measurements is necessary.

Theoretically, a noisy measurement is represented as a randomized measurement. As we shall see later, nonrandomized measurements are at least equal to or better than randomized ones. Now, we go to an abstract procedure.

2.6.1 Implementable decision POVMs

We construct a proper subset of decision POVMs based on Γ. First we fix an arbitrary implementable measurement $(\mathbf{M}, \mathcal{B}(\chi)) \in \Gamma$. Then for any Borel measurable function $u : \chi \to U$, we can construct a decision POVM $(\mathbf{M}', \mathcal{B}(U)) \in \mathcal{P}o(U)$. A function $f : X \to Y$ is called a *Borel measurable* if $f^{-1}(B) \in \mathcal{B}(X)$, $\forall B \in \mathcal{B}(Y)$. For each measurement, we obtain decision POVMs through this procedure. Let us denote Γ_U the whole set of the decision POVMs obtained by this change of variables. By definition Γ_U is a subset of $\mathcal{P}o(U)$.

Definition 4. Let a QSD problem $(\sigma(\theta), U, w)$ and an implementable class of measurements Γ be given. We call Γ_U a *naive set of implementable decision POVMs.*

Intuitively, the above subset of decision POVMs is very natural. However, according

to classical statistical decision theory, we have to include all randomized decisions [21]. For example, it is possible to use a randomly chosen estimator $u_j(x)$ among several estimators $u_1(x), u_2(x), \ldots$. It seems quite strange but there exists the case where a randomized decision is minimax in classical statistical decision problem.

A randomized procedure is mathematically represented by a Markov kernel $\kappa(du|x)$. It includes any nonrandomized function $u(x)$ as a special case. Thus, we write the whole set of decision POVMs constructed by a Markov kernel with an implementable measurement as $\kappa_U(\Gamma)$. Clearly $\Gamma_U \subseteq \kappa_U(\Gamma)$ holds.

Likewise, we can use a randomly chosen decision POVM among $\kappa_U(\Gamma)$. Mathematically it is represented by a closed convex hull of $\kappa_U(\Gamma)$. Technically speaking, we adopt the weak topology in $\mathcal{P}o(U)$ [13]. Thus we finally obtain the following definition.

Definition 5. Let a QSD problem $(\sigma(\theta), U, w)$ and an implementable class of measurements Γ be given. Let us call

$$\mathcal{P}_\Gamma := \overline{\mathrm{co}(\kappa_U(\Gamma))}$$

implementable decision POVMs generated by Γ, where $\overline{\mathrm{co}(A)}$ denotes a closed convex hull of a set A.

Lemma 1. Let U be a compact metric space and $\kappa(\mathrm{d}u)$ be a probability measure on U. For every $\epsilon > 0$ and continuous functions $f_1, \ldots, f_m \in C(U)$, there exists a number A and a finite subset $\{u^{(1)}, \ldots, u^{(A)}\} \subset U$ such that

$$\left| \frac{1}{A} \sum_{a=1}^{A} f_1(u^{(a)}) - \int_U f_1(u)\kappa(\mathrm{d}u) \right| < \epsilon,$$
$$\cdots$$
$$\left| \frac{1}{A} \sum_{a=1}^{A} f_m(u^{(a)}) - \int_U f_m(u)\kappa(\mathrm{d}u) \right| < \epsilon. \tag{1}$$

Proof. Let $\epsilon > 0$ be fixed. Let us consider sampling over U,

$$u^{(1)}, \ldots, u^{(A)} \overset{i.i.d.}{\sim} \kappa(\mathrm{d}u).$$

For notational convenience, we denote for each $j = 1, \ldots, m$,

$$\bar{f}_{j,A} := \frac{1}{A} \sum_{a=1}^{A} f_j(u^{(a)}), \qquad E[f_j] := \int_U f_j(u) \kappa(\mathrm{d}u).$$

Due to elementary probability theory, for a large number A, we obtain

$$P\left(\sum_{j=1}^{m} |\bar{f}_{j,A} - E[f_j]|^2 \geq \epsilon \right) < 1,$$

which implies that there exists a finite subset of U, $\{u^{(1)}, \ldots, u^{(A)}\}$ satisfying the inequalities (1). □

The following lemma describes the uniform approximation of a finite set of measurable functions. Since its proof is straightforward, it is omitted.

Lemma 2. Let X be a topological space with Borel algebra $\mathcal{B}(X)$ and $q_1, \ldots, q_m$ be probability measures on $(X, \mathcal{B}(X))$. Suppose that measurable functions $h_1, \ldots, h_m$ are integrable with respect to $q_1, \ldots, q_m$ respectively. Then, for every $\epsilon > 0$, there exists a finite collection of mutually disjoint subsets $\{Z_1, \ldots, Z_S\} \subseteq \mathcal{B}(X)$ of X and points $\{z_1, \ldots, z_S\} \subseteq X$ satisfying the following:

$$\begin{aligned} &X = Z_1 \cup \cdots \cup Z_S, \\ &z_s \in Z_s, \ s = 1, \ldots, S, \\ &\left| \sum_{s=1}^{S} h_j(z_s) q_j(Z_s) - \int_X h_j(x) q_j(\mathrm{d}x) \right| < \epsilon, \ j = 1, \ldots, m. \end{aligned} \tag{2}$$

Using the above lemmas, we obtain a fundamental result.

Theorem 1. Let a QSD problem $(\sigma(\theta), U, w)$ and an implementable class of measurements Γ be given. Then,

$$\mathcal{P}_\Gamma = \overline{\mathrm{co}(\Gamma_U)}$$

holds.

Proof. Since $\mathcal{P}_\Gamma \supseteq \overline{\mathrm{co}(\Gamma_U)}$ holds, we show $\mathcal{P}_\Gamma \subseteq \overline{\mathrm{co}(\Gamma_U)}$. It is enough to show that $\kappa_U(\Gamma) \subseteq \overline{\mathrm{co}(\Gamma_U)}$.

According to Holevo [13], an arbitrary neighborhood of $\mathbf{N}$ in $\mathcal{P}o(U)$ is given by

$$\mathcal{U}(\mathbf{N}) = \Big\{ \mathbf{M} \in \mathcal{P}o(U) : \Big| \int_U f_1(u)\mathrm{Tr}\rho_1 \mathbf{M}(\mathrm{d}u) - \int_U f_1(u)\mathrm{Tr}\rho_1 \mathbf{N}(\mathrm{d}u) \Big| < \epsilon_1, \ldots, \\ \Big| \int_U f_m(u)\mathrm{Tr}\rho_m \mathbf{M}(\mathrm{d}u) - \int_U f_m(u)\mathrm{Tr}\rho_m \mathbf{N}(\mathrm{d}u) \Big| < \epsilon_m \Big\},$$

where $\epsilon_1, \ldots, \epsilon_m$ are positive numbers, $\rho_1, \ldots, \rho_m$ are density operators, and $f_1, \ldots, f_m$ are continuous functions. For simplicity, we set $\epsilon_1 = \cdots = \epsilon_m = \epsilon$ and we fix density operators $\rho_1, \ldots, \rho_m$, continuous functions $f_1, \ldots, f_m \in C(U)$. We show that $\mathcal{U}(\mathbf{N}) \cap \mathrm{co}(\Gamma_U) \neq \emptyset$ for an arbitrary neighborhood of $\mathbf{N} \in \kappa_U(\Gamma)$. It implies that $\mathbf{N}$ is a closure point of $\mathrm{co}(\Gamma_U)$, i.e., $\mathbf{N} \in \overline{\mathrm{co}(\Gamma_U)}$. We construct a decision POVM $\mathbf{M}_0 \in \mathcal{U}(\mathbf{N}) \cap \mathrm{co}(\Gamma_U)$ in the following manner.

First we fix $\mathbf{N} \in \kappa_U(\Gamma)$. Since $\mathbf{N}$ is decomposed into a Markov kernel and an implementable measurement, say, $\mathbf{M}(\mathrm{d}x) \in \Gamma$, we rewrite the terms related to $\mathbf{N}$ as follows. For $j = 1, \ldots, m$,

$$\begin{aligned} \int_U f_j(u)\mathrm{Tr}\rho_j \mathbf{N}(\mathrm{d}u) &= \int_X \left(\int_U f_j(u)\kappa(\mathrm{d}u|x) \right) \mathrm{Tr}\rho_j \mathbf{M}(\mathrm{d}x) \\ &= \int_X h_j(x) q_j(\mathrm{d}x). \end{aligned} \tag{3}$$

In the last line, we set

$$h_j(x) = \int_U f_j(u)\kappa(\mathrm{d}u|x), \ q_j(\mathrm{d}x) = \mathrm{Tr}\rho_j \mathbf{M}(\mathrm{d}x),$$

where q_j is a probability measure on X and h_j is a bounded function over X, thus, also q_j-integrable. Then, due to Lemma 2, we take a finite collection of subsets $\{Z_1, \ldots, Z_S\}$ and points $\{z_1, \ldots, z_s\}$ satisfying (2).

Next, for every $s = 1, \ldots, S$, the following integrals

$$h_j(z_s) = \int_U f_j(u)\kappa(\mathrm{d}u|z_s), \ j = 1, \ldots, m$$

are simultaneously approximated by mean of sample points on U due to Lemma 1. It might depend on s. Thus, we write $u_s^{(1)}, \ldots, u_s^{(A)}$ as those sample points. Then the following holds:

$$\left| \frac{1}{A} \sum_{a=1}^{A} f_j(u_s^{(a)}) - \int_U f_j(u) \kappa(\mathrm{d}u|z_s) \right| < \epsilon, \ j = 1, \ldots, m.$$

Note that we take the same number of points A for each s. Summing over s, we obtain the following inequality,

$$\begin{aligned} & \left| \sum_{s=1}^{S} h_j(z_s) q_j(Z_s) - \sum_{s=1}^{S} \left\{ \frac{1}{A} \sum_{a=1}^{A} f_j(u_s^{(a)}) \right\} q_j(Z_s) \right| \\ & \leq \sum_{s=1}^{S} q_j(Z_s) \left| h_j(z_s) - \frac{1}{A} \sum_{a=1}^{A} f_j(u_s^{(a)}) \right| \\ & < \epsilon \sum_{s=1}^{S} q_j(Z_s) \\ & = \epsilon. \end{aligned} \tag{4}$$

Finally, we define $\mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(A)} \in \Gamma_U$ in the following manner. Let us define measurable functions $u^{(a)} : X \to U$, $a = 1, \ldots, A$ such that

$$u^{(a)}(x) := \begin{cases} u_1^{(a)}, & x \in Z_1, \\ \quad \ldots & \\ u_S^{(a)}, & x \in Z_S. \end{cases}$$

Through this change of variables, from $\mathbf{M}(\mathrm{d}x)$, we define decision POVMs $\mathbf{M}^{(a)}(\mathrm{d}u) \in \Gamma_U$, $a = 1, \ldots, A$. Then, for $a = 1, \ldots, A$, we can rewrite as follows:

$$\begin{aligned} \int_U f_j(u) \mathrm{Tr} \rho_j \mathbf{M}^{(a)}(\mathrm{d}u) &= \int_X f_j(u^{(a)}(x)) \mathrm{Tr} \rho_j \mathbf{M}(\mathrm{d}x) \\ &= \sum_{s=1}^{S} \int_{Z_s} f_j(u^{(a)}(x)) \mathrm{Tr} \rho_j \mathbf{M}(\mathrm{d}x) \\ &= \sum_{s=1}^{S} \int_{Z_s} f_j(u_s^{(a)}) q_j(\mathrm{d}x) \\ &= \sum_{s=1}^{S} f_j(u_s^{(a)}) q_j(Z_s). \end{aligned} \tag{5}$$

We claim that $\mathbf{M}_0 = \frac{1}{A}\sum_{a=1}^{A}\mathbf{M}^{(a)}(\mathrm{d}u) \in \mathrm{co}(\Gamma_U)$ is in the neighborhood of $\mathbf{N}$. Indeed, by using (3), (4), and (5),

$$\begin{aligned}
&\left|\int_U f_j(u)\mathrm{Tr}\rho_j\mathbf{N}(\mathrm{d}u) - \int_U f_j(u)\mathrm{Tr}\rho_j\mathbf{M}_0(\mathrm{d}u)\right| \\
&= \left|\int_U f_j(u)\mathrm{Tr}\rho_j\mathbf{N}(\mathrm{d}u) - \int_U f_j(u)\mathrm{Tr}\rho_j\left(\frac{1}{A}\sum_{a=1}^{A}\mathbf{M}^{(a)}(\mathrm{d}u)\right)\right| \\
&= \left|\int_X h_j(x)q_j(\mathrm{d}x) - \sum_{s=1}^{S}\frac{1}{A}\sum_{a=1}^{A} f_j(u_s^{(a)})q_j(Z_s)\right| \\
&\leq \left|\int_X h_j(x)q_j(\mathrm{d}x) - \sum_{s=1}^{S} h_j(z_s)q_j(Z_s)\right| \\
&\quad + \left|\sum_{s=1}^{S} h_j(z_s)q_j(Z_s) - \sum_{s=1}^{S}\frac{1}{A}\sum_{a=1}^{A} f_j(u_s^{(a)})q_j(Z_s)\right| \\
&< 2\epsilon.
\end{aligned}$$

holds for every $j = 1, \ldots, m$.

□

By definition, every implementable decision POVM is (almost) implementable in an experiment. Note that $\mathcal{P}_\Gamma$ includes closure points, where pathological decision POVMs like the Cantor distribution appear.

3 Main Result

In classical statistics, almost all of statistical problems are formulated in statistical decision theory [5, 25]. As a mathematical problem, we can consider statistical methods and their theoretical justification. Directly or indirectly, theoretical results bring many real application.

As a straightforward extension, quantum statistical decision theory was established decades ago [12]. Each statistical problem in a quantum physics experiment is written as a QSD problem $(\sigma(\theta), U, w)$. However, main interests have been mathematical

formulation itself and effect of entanglements [2, 17]. In other words, Holevo's QSD theory and asymptotic theory in this framework [6, 10, 15, 24] lacks the original significance in classical counterpart, i.e., yielding useful statistical methods to experimenters.

Our claim is that we should consider each QSD problem with a restricted class of decision POVMs $\mathcal{P}_\Gamma$ according to Γ, where Γ reflects a real capability of preparing experiments in each laboratory. As we shall see in Section 4, considering decision POVMs among $\mathcal{P}_\Gamma$ yields directly useful statistical methods to experimenters. Many examples will be presented for another occasion. Here we mainly focus on foundational results.

3.1 Existence theorem of Bayes POVMs and minimax POVMs

Now we consider the best decision POVM among the restricted class, $\mathcal{P}_\Gamma$. Since $\mathcal{P}_\Gamma$ inherits good mathematical properties from $\mathcal{P}o(U)$, some theoretical results still hold if we replace $\mathcal{P}o(U)$ with $\mathcal{P}_\Gamma$. First we define two kinds of optimality.

Definition 6. Let a QSD problem $(\sigma(\theta), U, w)$, an implementable class of measurements Γ and a prior distribution $\pi \in \mathcal{P}(\Theta)$ be given. If a decision POVM $\mathbf{M}_* \in \mathcal{P}_\Gamma$ achieves the infimum of the average risk

$$\inf_{\mathbf{M}\in\mathcal{P}_\Gamma} R_\pi(\mathbf{M}) = R_\pi(\mathbf{M}_*),$$

then it is called a $\mathcal{P}_\Gamma$-*Bayes decision POVM (w.r.t. π)*. If a decision POVM $\mathbf{M}_* \in \mathcal{P}_\Gamma$ achieves the infimum of the worst risk

$$\inf_{\mathbf{M}\in\mathcal{P}_\Gamma} \sup_\theta R(\theta;\mathbf{M}) = \sup_\theta R(\theta;\mathbf{M}_*),$$

then it is called a $\mathcal{P}_\Gamma$-*minimax decision POVM.*

When implementable decision POVMs are clear, we often omit $\mathcal{P}_\Gamma$. Theoretical results with respect to existence were given by the author [23]. We summarize them

for readers' convenience.

Theorem 2. Let a QSD problem $(\sigma(\theta), U, w)$ and an implementable class of measurements Γ be given.

(i) For every prior distribution $\pi \in \mathcal{P}(\Theta)$, there exists a $\mathcal{P}_\Gamma$-Bayes decision POVM.

(ii) There exists a $\mathcal{P}_\Gamma$-minimax decision POVM.

(iii) $\mathcal{P}_\Gamma$-minimax decision POVM is written as a $\mathcal{P}_\Gamma$-Bayes decision POVM with respect to a prior distribution.

Proof. For (i) and (ii), each proof follows the same route as $\mathcal{P}_\Gamma = \mathcal{P}o(U)$, which is shown by Holevo [13] (See also Ozawa [19]) and by Bogomolov [1] respectively. (iii) is a consequence of quantum minimax theorem [23]. □

3.2 Extremal point theorem

From the definition, a subset of decision POVM $\mathcal{P}_\Gamma$ is generally too abstract. Indeed, it includes pathological POVMs like the Cantor distribution. However, our following result implies that the $\mathcal{P}_\Gamma$- Bayes decision POVM is not pathological any more in usual cases.

Definition 7. Let a QSD problem $(\sigma(\theta), U, w)$ be given. For every subset $\mathcal{K}$ of $\mathcal{P}o(U)$, a decision POVM $\mathbf{M} \in \mathcal{K}$ is called an *extremal decision POVM of* $\mathcal{K}$ if it satisfies the following

$$\mathbf{M} = \sum_{a=1}^{A} \lambda_a \mathbf{M}_a,\ \lambda_1 > 0, \ldots, \lambda_A > 0, \sum_{a=1}^{A} \lambda_a = 1; \mathbf{M}_1, \ldots, \mathbf{M}_A \in \mathcal{K}$$
$$\Rightarrow \mathbf{M}_1 = \cdots = \mathbf{M}_A = \mathbf{M}.$$

Let us denote the whole extremal decision POVMs of $\mathcal{K}$ as $ex(\mathcal{K})$.

If we consider a convex subset $\mathcal{K}$, then the above definition reduces to the usual definition, i.e., $A = 2$. Unfortunately, this simple definition is not suitable for noncon-

vex sets. Suppose that we have only four decision POVMs, $\mathcal{K} = \{\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4\}$ and $\mathbf{M}_4$ has the unique representation of $\mathbf{M}_4 = \frac{1}{3}(\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3)/3$. Then $\mathbf{M}_4$ is not written as a mixture of any pair of decision POVMs in $\mathcal{K}$. If we fixed $A = 2$ in the above definition, then $\mathbf{M}_4$ would be an extremal point, which is not desired in Theorem 4.

Theorem 3. If we take the average risk $R_\pi(\mathbf{M})$ as a functional over the $\mathcal{P}_\Gamma$, then it achieves the minimum at least on the extremal points of $\mathcal{P}_\Gamma$. That is,

$$\min_{\mathbf{M}\in\mathcal{P}_\Gamma} R_\pi(\mathbf{M}) = \min_{\mathbf{M}\in ex(\mathcal{P}_\Gamma)} R_\pi(\mathbf{M})$$

holds.

Proof. This proof follows the same route as $\mathcal{P}_\Gamma = \mathcal{P}o(U)$, which is shown by Holevo [13]. □

Theorem 3 tells us that it is enough to seek the Bayes decision POVM in $ex(\mathcal{P}_\Gamma)$ instead of $\mathcal{P}_\Gamma$. On the other hand, it tells us nothing about these extremal points $ex(\mathcal{P}_\Gamma)$. By Theorem 1, $\mathcal{P}_\Gamma = \overline{\mathrm{co}(\Gamma_U)}$ holds and it is natural to expect that $ex(\mathcal{P}_\Gamma)$ is almost equal to the extremal points of Γ_U. However, we must be careful for taking closure. The mathematical operation of taking a closed convex hull of a subset A implies that weak closure points are added to $\mathrm{co}(A)$. When we find out the Bayes decision POVM (basically by a numerical method), these additional points would be troublesome. However, the following theorem assures that no additional point appear as an extremal point after taking closure.

Theorem 4. Let a QSD problem $(\sigma(\theta), U, w)$ and an implementable class of measurements Γ be given. Then,

$$ex(\mathcal{P}_\Gamma) \subseteq ex(\overline{\Gamma_U})$$

holds.

The following lemma is useful to show Theorem 4.

Lemma 3. Let A and B be subsets of a linear space. Then, $B \subseteq A \Rightarrow B \cap ex(A) \subseteq ex(B)$.

Now we show Theorem 4.

Proof. By Theorem 1, $\mathcal{P}_\Gamma = \overline{\mathrm{co}(\Gamma_U)} = \overline{\mathrm{co}(\overline{\Gamma_U})}$ holds. In Lemma 3, setting $A = \mathcal{P}_\Gamma$ and $B = \overline{\Gamma_U}$, we obtain $\overline{\Gamma_U} \cap ex(\mathcal{P}_\Gamma) \subseteq ex(\overline{\Gamma_U})$.

Since $\overline{\Gamma_U}$ is a closed, hence, a compact, subset of $\mathcal{P}o(U)$, all extremal points of $\mathcal{P}_\Gamma$ arise from points in $\overline{\Gamma_U}$. (Chap. V, section 8.3, lemma 5, in Dunford and Schwartz [3]). That is, $ex(\mathcal{P}_\Gamma) \subseteq \overline{\Gamma_U}$. Therefore, $ex(\mathcal{P}_\Gamma) \subseteq ex(\overline{\Gamma_U})$.

□

Due to Theorem 1, if $\overline{\Gamma_U}$ is convex, clearly $\overline{\Gamma_U} = \mathcal{P}_\Gamma$. If $\overline{\Gamma_U}$ is nonconvex, then Γ_U is also nonconvex (Contraposition of "A is convex $\Rightarrow$ $\bar{A}$ is convex.") and $ex(\overline{\Gamma_U})$ is not necessarily equal to $ex(\mathcal{P}_\Gamma)$. Theorem 4 also tells us that a pathological decision POVM does not appear as the unique Bayes decision POVM in usual cases. If so, such a decision POVM must be the extremal point and included in $ex(\overline{\Gamma_U}) \subseteq \overline{\Gamma_U}$, which is not likely to happen in a real experiment.

3.2.1 Finite-dimensional case

When the dimension of the underlying Hilbert space is finite, the above result becomes very simple.

Theorem 5. Let a QSD problem $(\sigma(\theta), U, w)$ on the finite-dimensional Hilbert space be given. Let an implementable class of measurements Γ be a closed subset of projective valued measurements (PVM). Then,

$$ex(\overline{\Gamma_U}) = \Gamma_U$$

holds.

Note that in a finite-dimensional Hilbert space, every PVM is identified with a finite set of mutually orthogonal projections summing to the identity.

Proof. By definition, every element of Γ_U is given by a finite-set of orthogonal projection matrix $\{\mathbf{E}_{u_1}, \ldots, \mathbf{E}_{u_k}\}$. Since Γ is closed, Γ_U is also closed. Every projective measurement in Γ_U is an extremal point (See, Chap. 1, Prop. 6.2, Holevo [14]). Thus, $ex(\overline{\Gamma_U}) = ex(\Gamma_U) = \Gamma_U$. □

This theorem assures that the $\mathcal{P}_\Gamma$-Bayes decision POVM is necessarily in Γ_U if an implementable class of measurements Γ is a closed subset of PVM. Thus, it is constructed by a PVM in Γ and a Bayes estimate. At least we can find the $\mathcal{P}_\Gamma$-Bayes decision POVM by a brute force method.

4 Example

We set $n = 4m$, $m = 1, 2, \ldots$. Suppose that we have $4m$-iid quantum system described by $\sigma(\theta) := \rho(\theta)^{\otimes 4m}$, where

$$\rho(\theta) := \frac{1}{2}\begin{pmatrix} 1+\theta_1 & \theta_2 \\ \theta_2 & 1-\theta_1 \end{pmatrix}.$$

We set $\mathrm{Im}\rho_{12} = \mathrm{Im}\rho_{21} = 0$ for simplicity.

We consider parameter estimation from $4m$ iid sample. The parameter space and the decision space are given by

$$U = \Theta = \{\theta = (\theta_1, \theta_2) \in \mathbf{R}^2 : \|\theta\| \leq 1\}.$$

We adopt the squared error as the loss function, $w(\theta, u) = \|\theta - u\|^2$. We assume a family of prior distributions $\pi(\mathrm{d}\theta) \propto r^{a+1}\mathrm{d}r\mathrm{d}\varphi$, where $a \geq 0$ is a positive constant and (r, φ) is the polar coordinate with $0 \leq r \leq 1$ and $0 \leq \varphi \leq 2\pi$, where $\theta_1 = r\cos\varphi$, $\theta_2 = r\sin\varphi$. We consider the QSD problem $(\sigma(\theta), U, w)$ with a prior distribution π.

Next let us consider implementable measurements. We introduce some notations to write them. First of all, we define the following observable $\widetilde{W}(\phi) := (\cos\phi)X + (\sin\phi)Z$ for $\phi \in [0, 2\pi)$, where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Then let us define PVM $\mathbf{W}(\phi)$ as the spectral decomposition of $\widetilde{W}(\phi)$. We symbolically write

$$\mathbf{W}(\phi_1) \otimes \mathbf{W}(\phi_2),$$

which implies that we perform the PVM $\mathbf{W}(\phi_1)$ for the first system and $\mathbf{W}(\phi_2)$ for the second system.

For each ϕ, the experimenter has to prepare the corresponding measurement described by $\mathbf{W}(\phi)$. Suppose that this preparation costs so much and m-repetition of the same measurement, $\mathbf{W}(\phi)^{\otimes m}$ is not so troublesome. Usually, in this situation, an experimenter prefers to perform two kinds of PVM, say, $\phi = 0, \pi/2$, for each $2m$-iid system. However, from the viewpoint of statistical estimation using *finite* data (i.e., $m < \infty$), there is no strong reason to adopt only two kinds of PVM. Thus, we admit at most four kinds of PVM, say $\phi = 0, \pi/4, \pi/2, 3\pi/4$ for our choice. General form is given by

$$\mathbf{W}(\phi_1)^{\otimes m_1} \bigotimes \mathbf{W}(\phi_2)^{\otimes m_2} \bigotimes \mathbf{W}(\phi_3)^{\otimes m_3} \bigotimes \mathbf{W}(\phi_4)^{\otimes m_4},$$

where m_1, m_2, m_3, m_4 is nonnegative integers satisfying $m_1 + m_2 + m_3 + m_4 = 4m$. For simplicity, we set $m_1 = m_2 = m_3 = m_4 = m$. After the above consideration, we take our implementable class as

$$\Gamma_4 := \left\{ \bigotimes_{j=1}^{4} \mathbf{W}(\phi_j)^{\otimes m} : \ \phi_1, \ldots, \phi_4 \in [0, 2\pi) \right\}.$$

Now our goal is to find the best parameter set $\phi = (\phi_1, \cdots, \phi_4)$. It corresponds to finding the optimal measurement *among possible class of measurements* in the average risk. Indeed, due to Theorem 3 and Theorem 5,

$$\min_{\mathbf{M} \in \mathcal{P}_{\Gamma_4}} R_\pi(\mathbf{M}) = \min_{\mathbf{M} \in \Gamma_{4,U}} R_\pi(\mathbf{M})$$

holds. By using Bayesian estimate, the right hand side reduces to the minimization problem with respect to the four parameter $(\phi_1, \cdots, \phi_4)$. For the squared loss, we explicitly write

$$u_\pi(x;\phi) = \frac{\int \theta p(x|\theta;\phi)\pi(\theta)\mathrm{d}\theta}{\int p(x|\theta;\phi)\pi(\theta)\mathrm{d}\theta},$$
$$\min_{\mathbf{M}\in\Gamma_{4,U}} R_\pi(\mathbf{M}) = \min_{\phi} \int E_{\theta,\phi}\left[\|\theta - u_\pi(x;\phi)\|^2\right]\pi(\theta)\mathrm{d}\theta,$$

where $u_\pi = (u_\pi^1, u_\pi^2)$ is the Bayes estimate with respect to π (for derivation, see, e.g., Robert [20], section 2-5.) and $E_{\theta,\phi}[\cdot]$ denotes the expectation with respect to the measurement outcome x, whose distribution $p(x|\theta;\phi)$ is determined by the density operator $\rho(\theta)^{\otimes 4m}$ and projective measurements $\bigotimes_{j=1}^4 \mathbf{W}(\phi_j)^{\otimes m}$. At least, by brute force method, we can find the best parameter ϕ numerically.

4.1 Monte Carlo optimization

It is necessary to evaluate the following quantities:

$$u_\pi(x;\phi) = \int \theta\pi(\theta|x;\phi)\mathrm{d}\theta = \frac{\int \theta p(x|\theta;\phi)\pi(\theta)\mathrm{d}\theta}{\int p(x|\theta;\phi)\pi(\theta)\mathrm{d}\theta},$$
$$R_\phi(r,\varphi) = E_{\theta,\phi}[\|\theta - u_\pi(x;\phi)\|^2] = E_{\theta,\phi}[(r\cos\varphi - u_\pi^1)^2 + (r\sin\varphi - u_\pi^2)^2],$$
$$r_\pi(\phi) = \iint R_\phi(r,\varphi)(a+2)r^{a+1}\mathrm{d}r\frac{\mathrm{d}\varphi}{2\pi}.$$

In the first line, $\pi(\theta|x)$ is the posterior distribution. We use the standard Monte Carlo method to evaluate each expectation or integral. Each iteration number of the Monte Carlo loop is set between $10^3 \sim 2\times 10^4$ respectively according to its accuracy. Thus, we need to perform, say $10^4 \times 10^4 \times 10^4 = 10^{12}$ iteration for one fixed measurement, $\bigotimes_{j=1}^4 \mathbf{W}(\phi_j)^{\otimes m}$. Stochastic optimization over the full parameter region $[0, 2\pi)^{\times 3}$ is inefficient. (Note that we fix $\phi_1 = 0$ due to symmetry.) Indeed, we have not finished yet detailed numerical studies. We only mention what we have done so far. Detailed analysis will be presented for another occasion.

4.2 Best 2-PVMs with flat prior

Although it seems trivial, we present the best 2-PVMs in order to show how our method really works. We call the pair of observables $(\widetilde{W}(\phi), \widetilde{W}(\phi+\pi/2))$, $\phi \in [0, 2\pi)$ an *orthogonal pair of observables*. Let us consider the following implementable class

$$\Gamma_2 = \left\{ \mathbf{W}(\phi_1)^{\otimes 2m} \otimes \mathbf{W}(\phi_2)^{\otimes 2m} : \ \phi_1, \phi_2 \in [0, 2\pi) \right\}.$$

Due to symmetry, we can set $\phi_1 = 0$. Then seeking the optimal decision POVMs in $\mathcal{P}_{\Gamma_2}$ reduces to one parameter optimization. It is easily seen that the best PVMs in Γ_2 is $\mathbf{W}(\phi)^{\otimes 2m} \otimes \mathbf{W}(\phi+\pi/2)^{\otimes 2m}$. In other words, $2m$-repetition of an orthogonal pair of observables is the best measurement among $\mathcal{P}_{\Gamma_2}$. The Bayes decision POVM is obtained if we set $u(x)$ as the Bayes estimate. We set the minimum $r_{2,\min}^{(2m)} = \min_{\mathbf{M}\in\mathcal{P}_{\Gamma_2}} R_\pi(\mathbf{M})$.

4.3 4-PVMs with a prior concentrated on pure states

Instead of Γ_4, we deal with a slightly restricted implementable class

$$\Gamma_{2,2} := \left\{ \bigotimes_{j=1}^{4} \mathbf{W}(\phi_j)^{\otimes m} : \ \phi_1, \phi_3 \in [0, 2\pi); \phi_2 = \phi_1 + \pi/2, \phi_4 = \phi_3 + \pi/2 \right\}.$$

Due to symmetry, we can set $\phi_1 = 0$ and $\phi_3 \in [0, \pi/2)$. Then seeking the optimal decision POVMs in $\mathcal{P}_{\Gamma_{2,2}}$ reduces to one parameter optimization (ϕ_3),

$$\min_{\mathbf{M}\in\mathcal{P}_{\Gamma_{2,2}}} R_\pi(\mathbf{M}) = \min_{\phi_3} r_\pi(0, \pi/2, \phi_3, \phi_3 + \pi/2).$$

Numerical computation implies that

$$r_{2,\min}^{(2m)} = r_\pi(0, \pi/2, 0, \pi/2) \geq r_\pi(0, \pi/2, \phi_3, \phi_3 + \pi/2)$$

holds. Thus we define the relative risk difference $\eta(\phi_3)$ as

$$\eta(\phi_3) = \frac{r_{2,\min}^{(2m)} - r_\pi(0, \pi/2, \phi_3, \phi_3 + \pi/2)}{r_{2,\min}^{(2m)}}.$$

By definition $\eta \leq 1$. We expect $\eta(\phi_3) \geq 0, \forall \phi_3$ and have interests in $\max_{\phi_3} \eta(\phi_3)$.

From several numerical computation (but not fully analyzed), it seems that we can see the difference more clearly than that with the flat prior ($a = 0$). Thus, we set $a = 10$ and the sample size m to be between $30 \sim 100$. We directly compare measurements with $\phi_3 = 0$ and those with $\phi_3 = \pi/4$. The former is $2m$-repetition of X and Z while the latter is m-repetition of $X, Z, \frac{X+Z}{\sqrt{2}}, \frac{X-Z}{\sqrt{2}}$. The relative improvement of the estimation error reaches about $\eta(\pi/4) \sim 10\%$ when $m \sim 100$.

4.4 Discussion

Numerical results are not enough but at this point they suggest that

(i) The m-repetition of $X, Z, \frac{X+Z}{\sqrt{2}}, \frac{X-Z}{\sqrt{2}}$ is the best combination in Γ_4. In particular, *the most natural way, $2m$-repetition of two observables X, Z, is not the best any more* when we estimate the expectations of X and Z.

(ii) If experimenters do not hesitate to prepare several kinds of PVMs, measurement using many kinds of PVMs may work better than 4 kinds of PVMs.

Theoretically $\mathcal{P}_{\Gamma_4}$ is a much smaller subset of $\mathcal{P}o(\Theta)$. The latter class includes the decision POVM using the full quantum correlation (i.e., entanglement over the $4m$-systems) and it is of course better than the best decision POVM in $\mathcal{P}_{\Gamma_4}$. However, our proposed method based on the m-repetition of $X, Z, \frac{X+Z}{\sqrt{2}}$ and $\frac{X-Z}{\sqrt{2}}$ is actually implementable in a usual sense. Our proposed method indicates that *improvement in statistical estimation is possible without any entanglement.*

We also note that measurements using quantum correlation (i.e., entanglement over a composite system) may be included in Γ if the experimenter can prepare such measurements successfully. Although admitting such measurements makes the optimization problem much harder than the above example, it is possible and very

interesting to find out the best decision POVMs among $\mathcal{P}_\Gamma$. Both effective algorithms dedicated to our formulation and theoretical development of good approximation are left for future study.

Acknowledgments

This work was partially supported by the Grant-in-Aid for Young Scientists (B) (No. 24700273) and the Grant-in-Aid for Scientific Research (B) (No. 26280005). The author is also grateful to all of participants for fruitful discussions in the RIMS workshop.

REFERENCES

[1] Bogomolov, N. A.: Minimax measurements in a general statistical decision theory. Theor. Prob. Appl. **26**, (1982) 787–795.

[2] Bruß, D., Ekert, A. and Machiavello, C.: Optimal universal quantum cloning and state estimation. Phys. Rev. Lett., **81**, (1998) 2598–2601.

[3] Dunford, N. and Schwartz, J. T.: *Linear Operators Part* I: *General Theory.* (John Wiley and Sons, New York 1988), Chap. V.

[4] Fan, H., Matsumoto, K., Wang, X., and Wadati, M.: Optimal quantum detectors for unambiguous detection of mixed states. Phys. Rev. A, **65**, (2002) 012304.

[5] Ferguson, T.:*Mathematical Statistics: A Decision Theoretic Approach* (Academic Press, New York 1967), Chap.2.

[6] Guţă, M., Jenčová, A.: Local asymptotic normality in quantum statistics. Commun. Math. Phys., **276** (2007), 341–379.

[7] Hammerer, K., Wolf, M., Polzik, E. and Cirac, J.: Quantum benchmark for storage and transmission of coherent states. Phys. Rev. Lett., **94**, (2005) 150503.

[8] Häseler, H., Moroder, T. and Lütkenhaus, N.: Testing quantum devices: Practical entanglement verification in bipartite optical systems. Phys. Rev. A, **77**, (2008) 032303.

[9] Häseler, H., Lütkenhaus, N.: Quantum benchmarks for the storage or transmission of quantum light from minimal resources. Phys. Rev. A, **81**, (2010) 060306(R).

[10] Hayashi, M.: *Asymptotic Theory of Quantum Statistical Inference.* World Scientific, Singapore, 2005.

[11] Helstrom, C. W.: *Quantum Detection Theory.* Academic Press, New York, 1976.

[12] Holevo, A. S.: Statistical Decision Theory for Quantum Systems. J. Multivariate Anal., **3**, (1973) 337–394.

[13] Holevo, A. S.: Investigations in the general theory of statistical decisions. Proc. Steklov Inst. Math. **35, 124** (1976) (In Russian), AMS Transl. (1978).

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

[15] Kahn, J., Guţă, M.: Local asymptotic normality for finite dimensional quantum systems. Commun. Math. Phys., **289** (2009), 597–652.

[16] Kumagai, W., Hayashi, M.: Quantum hypothesis testing for Gaussian states: quantum analogues of χ^2, t-, and F-tests. Commun. Math. Phys. **318**, (2013) 535–574.

[17] Massar, S., Popescu, S.: Optimal extraction of information from finite quantum ensembles. Phys. Rev. Lett. **74**, (1995) 1259–1263.

[18] Namiki, R., Koashi, M. and Imoto, N.: Fidelity criterion for quantum-domain transmission and storage of coherent states beyond the unit-gain constraint. Phys. Rev. Lett., **101**, (2008) 100502.

[19] Ozawa, M.:Optimal measurements for general quantum systems. Rep. Math.

Phys. **18**, (1980) 11–28.

[20] Robert, C. P.: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* 2nd ed., Springer, New York, 2001.

[21] Strasser, H.: *Mathematical Theory of Statistics*, (De Gruyter, Berlin 1985) 240–242.

[22] Tanaka, F.: Bayesian estimation of the wave function. Phys. Lett. A, **376**, (2012) 2471–2476.

[23] Tanaka, F.: Quantum Minimax Theorem, arXiv: quant-ph/1410.3639.

[24] Yamagata, K., Fujiwara, A., and Gill, R. D.: Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.* **41** (2013), 2197–2217.

[25] Wald, A.: *Statistical Decision Functions.* John Wiley & Sons, New York, 1950.