Noninformative prior in the quantum statistical model of pure states

Fuyuhiko Tanaka*
Department of Mathematical Informatics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan
(Received 2 March 2012; published 5 June 2012)

In the present paper, we consider a suitable definition of a noninformative prior on the quantum statistical model of pure states. While the full pure-states model is invariant under unitary rotation and admits the Haar measure, restricted models, which we often see in quantum channel estimation and quantum process tomography, have less symmetry and no compelling rationale for any choice. We adopt a game-theoretic approach that is applicable to classical Bayesian statistics and yields a noninformative prior for a general class of probability distributions. We define the quantum detection game and show that there exist noninformative priors for a general class of a pure-states model. Theoretically, it gives one of the ways that we represent ignorance on the given quantum system with partial information. Practically, our method proposes a default distribution on the model in order to use the Bayesian technique in the quantum-state tomography with a small sample.

DOI: 10.1103/PhysRevA.85.062305 PACS number(s): 03.67.−a, 02.50.Le, 03.65.Wj, 89.70.Cf

I. INTRODUCTION

Several decades ago, Wootters [1] asserted that "there does not seem to be any natural measure on the set of all mixed states." Since then, possible measures over the set of all mixed states have been investigated by many authors [2–5]. While there is no compelling rationale for the choice of a probability measure for the eigenvalues of a density operator, there exists a consensus that the invariant measure (Haar measure) is natural for all pure states in a finite-dimensional Hilbert space. Some authors also agree with the Bayesian viewpoint that a natural measure is considered a prior probability over the set of density operators. Srednicki [5] argued that the nature of our ignorance about a quantum system can be represented by a prior probability. However, it seems quite difficult to represent our ignorance about a given quantum system even for a pure-states model when we make additional assumptions for the system.

Let us take an extreme example. Suppose that we have two quantum systems S_1 and S_2. Each system is described by a state vector among a set of possible pure states. For S_1, the set of possible pure states is given by $\mathcal{M}_1 := \{|e_1\rangle, |e_2\rangle, |e_3\rangle\}$, where state vectors are orthogonal, $\langle e_i|e_j\rangle = \delta_{ij}$. For S_2, we take $\mathcal{M}_2 := \{|\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle\}$, where state vectors are *nonorthogonal*, $|\langle\psi_i|\psi_j\rangle| > 0$. How should we represent our ignorance on the true state for each system? For the first system, it is natural to assign the same weight to each state; thus, our ignorance is represented as a uniform prior $\pi_1 = \pi_2 = \pi_3 = 1/3$. How about the second one? This is our starting point of the present paper.

Apart from the philosophical problem originating from randomness, the choice of natural measure is inevitable in quantum-state tomography [6,7], which is the task of inferring the state of the (partially) unknown quantum system by use of appropriate measurements. In state tomography, with a smaller sample of data, the performance of the maximum likelihood estimate [8] tends to become worse. In order to avoid such difficulties, many researchers use the Bayesian technique [9,10]. Its optimality is also shown theoretically [11]. However, as in classical Bayesian statistics, we must choose a natural measure on the parametric model of the unknown quantum state [12], which is called a *noninformative prior distribution* (or, for short, *noninformative prior*). When the dimension d of the Hilbert space gets higher, the number of parameters to be determined increases rapidly (e.g., d^4 for process tomography). In such cases, it would be better to assume a restricted model with a smaller number of parameters (called a *submodel*). However, once we adopt a submodel, then the lack of symmetry makes the choice of a noninformative prior a more serious problem.

In classical Bayesian statistics [13], there is no universal criterion on how we should determine a noninformative prior and still there are a considerable number of works on the choice of a noninformative prior like the famous Jeffreys prior [14]. In the quantum setting, for the all mixed-states models, there seems to be a lot of work [2–5]. However, these works are mainly concerned with the geometry of the whole set of density matrices. On the other hand, Hayashi [15,16] recently proposed a noninformative prior based on asymptotic minimax coding, the quantum analog of the famous result by Clark and Barron [17]. As far as the author knows, it is the only prior that has an information-theoretic meaning. However, his derivation fully uses the group symmetry of the model and is not directly applied to any submodel, in particular, to one with no symmetry. For the application to quantum tomography and other quantum Bayesian estimations, we need at least one proposal of noninformative prior which is properly defined on a wide class of submodels.

In the present paper, we focus on the pure-states submodel, including a finite set of pure states, and consider a suitable definition of a noninformative prior on the model. It is desirable that there exists a certain interpretation of the noninformative prior. We here consider statistical decision theory, a version of game theory [18], and its application to Bayesian statistics by Bernardo [19]. Our proposed definition of a noninformative prior satisfies the following: (i) In classical cases, i.e., when all distinct pure states are orthogonal, agreeing with the uniform prior; (ii) defined on any submodel of the quantum states over any (possibly infinite-dimensional) Hilbert space under some regularity conditions; and (iii) with no approximation and no asymptotics. Concerning item (iii), it is more difficult to obtain

*ftanaka@stat.t.u-tokyo.ac.jp

a general result in nonasymptotics than in asymptotics. For example, in classical statistics, an average of random variables is approximately distributed to a normal distribution in the asymptotic setting (i.e., sample size $n \to \infty$), which makes some problems more tractable. However, it is also known that there is a gap between finite-sample theory and asymptotic theory.

Our main criterion is based on a quantum detection game between Alice and Bob. While the formal classical analog of the game is trivial, it becomes nontrivial in the quantum setting. We also show that the concept of no information is beyond our intuition in some examples. Through the choice of noninformative prior, we see another new aspect of quantum theory.

In the next section, we briefly review our setting and give a nontrivial example where symmetry does not uniquely determine a noninformative prior on a pure-states submodel. We then introduce the quantum detection game and a least favorable prior. In Sec. III, we present our main result. It is shown that the minimax estimate of the unknown quantum pure states is given by the Bayes estimate for the least-favorable prior. In two-dimensional Hilbert spaces, it agrees with the maximum entropy prior of Bayesian mixture states. We also show that the least favorable prior is not uniquely determined. Examples follow in the next section. We give an example where the uniform prior yields a worse performance than our prior. Concluding remarks are given in the last section.

II. PRELIMINARY

A. Basic definitions of quantum statistics

We briefly summarize some notations of quantum statistics (for quantum theory in finite-dimensional Hilbert spaces, see, e.g., Nielsen and Chuang [20]). Let $\mathcal{H}$ be a separable (possibly infinite dimensional) complex Hilbert space of a quantum system. A Hermitian operator ρ on $\mathcal{H}$ is called a *state* or *density operator* if it satisfies

$$\mathrm{Tr}\rho = 1, \quad \rho \geqslant 0.$$

We denote the set of all states on $\mathcal{H}$ as $\mathcal{S}(\mathcal{H})$. If a density operator is of rank 1, it is called a *pure state*. A normalized vector $\psi \in \mathcal{H}$ corresponds to a pure state one to one using the outer product up to the phase factor and we often identify $|\psi\rangle$ with $|\psi\rangle\langle\psi|$. When a density operator is not a pure state, it is called a *mixed state*. Note that any density operator is represented as a convex combination of pure states.

Let Ω be a space of all possible outcomes of an experiment (e.g., $\Omega = \mathbf{R}^n$) and suppose that a σ algebra $\mathcal{B} := \mathcal{B}(\Omega)$ of subsets of Ω is given. An affine map μ from $\mathcal{S}(\mathcal{H})$ into a set of probability distributions on Ω, $\mathcal{P} = \{\mu(dx)\}$ is called a *measurement*. There is a one-to-one correspondence between a measurement and a resolution of the identity. A map from $\mathcal{B}$ into the set of positive Hermitian operators

$$E : B \mapsto E(B),$$

where E satisfies

$$E(\phi) = O, \quad E(\Omega) = I,$$
$$E(\cup_i B_i) = \sum_i E(B_i), \quad B_i \cap B_j = \phi, \quad \forall \quad B_i \in \mathcal{B},$$

is called a *positive operator-valued measure* (*POVM*). Any physical measurement can be represented by a POVM. For a countable sample space $\Omega = \{x_1, x_2, \ldots,\}$, we set $\mathcal{B} = 2^\Omega$ and then we write $E_i := E(\{x_i\})$. An arbitrary POVM corresponds to a countable set of positive operators $\{E_i\}$ satisfying $\sum_i E_i = I$. Performing a measurement described by a POVM $\{E_i\}$ for an arbitrarily fixed ρ yields an outcome i with the probability $p_i := \mathrm{Tr}\rho E_i$. It is easily seen that $\{p_i\}$ is a distribution on Ω. The above holds for continuous sample spaces (for details, see, Holevo [21]).

B. Statistical model and prior selection

Our main assumption is that the unknown quantum system belongs to a finite-dimensional parametric family

$$\mathcal{M} = \{\rho_\theta \in \mathcal{S}(\mathcal{H}) : \theta \in \Theta \subseteq \mathbb{R}^k\} \subseteq \mathcal{S}(\mathcal{H}),$$

which we call a *quantum statistical model* or simly a *model*. By definition, a model could be a proper subset of $\mathcal{S}(\mathcal{H})$. When we would like to emphasize this, we call it a *submodel*. In the present paper, we deal with a pure-states model and we also write a model as

$$\mathcal{M} = \{\psi_\theta : \theta \in \Theta\},$$

where ψ_θ is a normalized vector. Since pure states are connected with each other through a unitary transformation, there is no classical counterpart. In particular, a formal mathematical analogy does not work in the pure-states model.

We now briefly explain the main idea of quantum estimation [21,22], whose purpose is to estimate the unknown parameter of physical interest or the unknown density operator itself from *finite* number of measurement data. We usually choose a suitable measurement $\{M_x\}$ and perform it for each system ρ_θ. Measurement outcome, x, which is a random variable, is distributed according to $p_\theta(x) = \mathrm{Tr}\rho_\theta M_x$. Thus, the unknown quantum system or some parameter of physical interest is to be estimated from data $x_1, \ldots, x_n$. In Bayesian statistics, the estimate of θ is constructed from the posterior distribution, which is defined by

$$\pi(\theta|x) := \frac{p_\theta(x)\pi(\theta)}{\int d\theta p_\theta(x)\pi(\theta)},$$

where $\pi(\theta)$ is a prior distribution over Θ, $\int \pi(\theta)d\theta = 1$, and $\pi(\theta) \geqslant 0$. By the above definition, it is easily seen that $\int \pi(\theta|x)d\theta = 1$ and $\pi(\theta|x) \geqslant 0$. For notational simplicity, we assume here that the prior has a density.

If we desire a certain objectivity in the Bayesian scheme and have no knowledge as to the parameter, then we seek a prior distribution that represents literally no information on the parameter, which is called a *noninformative prior*. It is also called a *vague prior* or an *objective prior*. For finite-dimensional cases, the full pure-states model admits the so-called Haar measure, which is the invariant measure on unitary group and is considered a natural measure [1–3]. However, as we shall see, there is no universally best definition of a noninformative prior in a given submodel.

Let us take the famous experiment of the Aharonov-Bohm effect as an example. For our purposes, it is enough to remember that the wave function is given by

$$\psi(x) \propto \psi_1(x) + e^{i\theta}\psi_2(x),$$

where two wave functions $\psi_1(x)$ and $\psi_2(x)$ are assumed to be known in advance. Both are normalized but not orthogonal to each other, that is, $\|\psi_i\|^2 = 1, i = 1,2$, and $0 < |\langle\psi_1|\psi_2\rangle| < 1$. The relative phase θ is proportional to the strength of the magnetic flux. We assume that the relative phase $\theta \in [0,2\pi)$ is unknown but fixed. We then obtain a one-dimensional parametric model of the unknown wave function $\psi_\theta(x) \propto \psi_1(x) + e^{i\theta}\psi_2(x), \theta \in [0,2\pi)$. When we have no information on θ, it seems natural to assume a uniform distribution

$$\pi_U(d\theta) = \frac{d\theta}{2\pi}. \tag{1}$$

On the other hand, it is possible to suggest another candidate for a noninformative prior. In order to see this, we switch to the density operator formalism. First, we rewrite a wave function as a complex vector in $\mathbb{C}^2$,

$$\psi_1 \leftrightarrow |\psi_1\rangle := \begin{pmatrix} \cos\varphi/4 \\ \sin\varphi/4 \end{pmatrix}, \quad \psi_2 \leftrightarrow |\psi_2\rangle := \begin{pmatrix} \cos\varphi/4 \\ -\sin\varphi/4 \end{pmatrix},$$

where φ is a constant defined by $|\langle\psi_1|\psi_2\rangle|^2 = \frac{1+\cos\varphi}{2}, \varphi \in (0,\pi)$. A parametric family of density matrices is given by

$$\rho(\theta) = \frac{|\psi(\theta)\rangle\langle\psi(\theta)|}{\mathrm{Tr}|\psi(\theta)\rangle\langle\psi(\theta)|},$$

where $|\psi(\theta)\rangle := |\psi_1\rangle + e^{i\theta}|\psi_2\rangle$. Since $\|\psi(\theta)\|^2 = 2 + 2\cos\frac{\varphi}{2}\cos\theta$,

$$\begin{aligned}\rho(\theta) &= \frac{1}{2 + 2\cos\frac{\varphi}{2}\cos\theta} \\ &\quad \times \begin{bmatrix} \left(1+\cos\frac{\varphi}{2}\right)(1+\cos\theta) & i\sin\frac{\varphi}{2}\sin\theta \\ -i\sin\frac{\varphi}{2}\sin\theta & \left(1-\cos\frac{\varphi}{2}\right)(1-\cos\theta) \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1+z(\theta) & -iy(\theta) \\ iy(\theta) & 1-z(\theta) \end{bmatrix},\end{aligned}$$

where we set

$$y(\theta) := -\frac{\sqrt{1-a^2}\sin\theta}{1+a\cos\theta}, \quad z(\theta) := \frac{a+\cos\theta}{1+a\cos\theta},$$
$$a := \cos\frac{\varphi}{2}.$$

We easily see that $y(\theta)^2 + z(\theta)^2 = 1$. Let us define ξ by

$$-\tan\xi := \frac{y(\theta)}{z(\theta)} = \frac{\sqrt{1-a^2}\sin\theta}{a+\cos\theta}.$$

Then, since $0 < a < 1$, ξ corresponds to θ one to one. Thus, we obtain another parametrization of the model,

$$\rho(\xi) = \frac{1}{2}\begin{pmatrix} 1+\cos\xi & -i\sin\xi \\ i\sin\xi & 1-\cos\xi \end{pmatrix}, \quad \xi \in [0,2\pi).$$

Since we do not know ξ completely, we can claim that the uniform prior for ξ is given by

$$\pi'_U(\xi) = \frac{d\xi}{2\pi}. \tag{2}$$

Both uniform distributions, (1) and (2), are derived from symmetry and they do not agree with each other.

From the above example, we see that even symmetry does not necessarily determine a noninformative prior uniquely. Since our intuition of no information does not seem to work well in the quantum setting, we need a careful treatment for the choice of a noninformative prior even when a model has a certain symmetry. From practical reasons, it is also important to analyze how a noninformative prior should be chosen in a given quantum statistical submodel with little or no symmetry.

In classical Bayesian statistics, many authors investigate noninformative priors in a specific class of statistical models [13,23] and a general definition of noninformative priors arises from a basic criterion. Usually we cannot expect a universally good definition of a noninformative prior and it depends on a basic criterion which of several definitions is better. Thus, in the quantum setting, what we need to do, first, is to propose a criterion and the definition of a noninformative prior is derived from them. In the present paper, we adopt a game-theoretic approach, which gives a good interpretation of the noninformative prior in classical cases.

C. Minimax coding and prior

Before introducing our method, we mention concisely the relation between minimax code and the least-favorable prior, which is very famous in information theory and derived from game-theoretic approach [24]. Let $\mathcal{X} = \{x_1, \ldots, x_k\}$ be the set of words and $p_\theta(x)$ denote a distribution on $\mathcal{X}$, where θ is the unknown parameter. If we know the distribution, then the ideal code length for each word x_j is given by $-\ln p_\theta(x_j)$. When we do not know the distribution, but know the range of the parameter, say Θ, we use the Bayes code. Each code length is given by

$$-\ln p_\pi(x) := -\ln\left\{\int \pi(d\theta)p_\theta(x)\right\}, \quad \forall \quad x \in \mathcal{X},$$

where π is a prior distribution on Θ, which we have to choose in a suitable way. We now define the relative entropy,

$$D(p||q) := \sum_{x\in\mathcal{X}} p(x)[\ln p(x) - \ln q(x)].$$

Then, under mild regularity conditions, the following minimax theorem [18] holds

$$\begin{aligned}&\inf_{q\in\mathcal{P}(\mathcal{X})}\sup_{\pi\in\mathcal{P}(\Theta)}\int \pi(d\theta)D(p_\theta||q) \\ &\quad = \sup_{\pi\in\mathcal{P}(\Theta)}\inf_{q\in\mathcal{P}(\mathcal{X})}\int \pi(d\theta)D(p_\theta||q),\end{aligned}$$

where $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\Theta)$ denote the set of all distributions on $\mathcal{X}$ and Θ, respectively. There exists a distribution π_* that achieves the supremum and a code distribution q_* that achieves the infimum. They are called the *least favorable prior* and the *minimax code distribution*, respectively [24]. One of the most important consequences is that a minimax code is obtained as a Bayes code with respect to a least favorable prior π_*.

Historically speaking, in Bernardo [19], the famous reference prior in classical setting is defined in the above way. Thus, the extension to the quantum setting is straightforward in the mixed-states model. The formal analog of the code length is given by the operator $-\ln\rho_\theta$ and a similar result,

$$\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{\pi\in\mathcal{P}(\Theta)}\int\pi(d\theta)D(\rho_\theta||\sigma) = \sup_{\pi\in\mathcal{P}(\Theta)}\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\int\pi(d\theta)D(\rho_\theta||\sigma),$$

holds, where the quantum relative entropy [25] is defined by

$$D(\rho||\sigma) := \mathrm{Tr}[\rho(\ln\rho - \ln\sigma)].$$

However, we deal with the pure-states model and the formal analog of the code length $-\ln|\psi\rangle\langle\psi|$ has only two eigenvalues, 0 and $+\infty$. The relative entropy between two different pure states always diverges whether they are orthogonal or not. Thus, it is not so suitable to make an argument based on the interpretation of "code length" in the pure-states model.

D. Quantum detection game

Instead of the code length interpretation, we take another interpretation for the choice of a prior in the pure-states model. Let us explain the setting of quantum signal detection as a game between Alice (nature) and Bob (statistician). For simplicity, we only take a discrete model. Suppose that Alice is able to prepare k kinds of pure quantum states $\mathcal{M} = \{\psi_1,\ldots,\psi_k\}$, where each state is described as a unit vector in a complex Hilbert space $\mathcal{H}$. Alice first determines an arbitrary prior distribution $\{\pi_j\}_{j=1,\ldots,k}$. She then chooses one quantum state, say, $\psi_j \in \mathcal{M}$, according to the prior and sends it to Bob. The receiver Bob has exact knowledge about candidates. Bob prepares the arbitrary reference state $\phi \in \mathcal{H}$ and he tries to detect any state. It corresponds to the two-valued measurement $\{|\phi\rangle\langle\phi|, I - |\phi\rangle\langle\phi|\}$. It is possible to perform such kinds of measurements in quantum optics. For each state ψ_j, Bob's detection probability is given by physical law, $p_j := |\langle\psi_j|\phi\rangle|^2 = F(\psi_j,\phi)$, where $F(\psi,\phi)$ is called the *fidelity* [20] between ψ and ϕ.

The purpose of Alice is to obtain a smaller detection probability in an average by choosing a prior on $\Theta := \{1,\ldots,k\}$. She may have interest in the following value:

$$V_U := \inf_{\pi\in\mathcal{P}(\Theta)}\sup_{\phi,\ \|\phi\|=1}\sum_{j=1}^{k}\pi_j F(\psi_j,\phi).$$

On the other hand, the purpose of Bob is to obtain a larger detection probability on average by choosing a reference state $\phi \in \mathcal{H}$. He may have interest in the following value:

$$V_L := \sup_{\phi,\ \|\phi\|=1}\inf_{\pi\in\mathcal{P}(\Theta)}\sum_{j=1}^{k}\pi_j F(\psi_j,\phi).$$

We call a prior distribution achieving V_U a *least favorable prior with respect to the detection game* denoted as π_*. A reference state achieving V_L is denoted as ϕ_* and called a *maximin detection strategy*. Note that we adopt here a kind of success probability instead of a loss function and we have to exchange the roles of min and max. If we consider Bob's task as the estimation of the unknown state with no experimental data, ϕ_* is the minimax estimate when the performance is evaluated by $1 - F$. In a similar line to classical minimax theorem [18], we easily see $V_U \geqslant V_L$. However, generally it could happen that $V_U > V_L$ because Bob's choice is not randomized.

As is usual in game theory, Bob may take a randomized strategy. If he is able to prepare any reference state $\phi \in \mathcal{H}$ according to a probability distribution $q(d\phi)$, then its detection probability is given by

$$\int F(\psi_j,\phi)q(d\phi).$$

It can also be rewritten as

$$F(\psi_j,\sigma) := \langle\psi_j|\sigma|\psi_j\rangle,\ \sigma := \int|\phi\rangle\langle\phi|q(d\phi).$$

Thus, taking a randomized measurement is equivalent to using the two-valued measurement $\{\sigma, I - \sigma\}$, $\sigma \in \mathcal{S}(\mathcal{H})$. Bob's strategy is represented as a density operator. In practice, his choice may be restricted to the same model $\mathcal{M}$. Then, $\sigma \in \overline{\mathrm{co}\mathcal{M}}$, where $\overline{\mathrm{co}\mathcal{M}}$ is the closure of the convex hull $\mathrm{co}\mathcal{M} := \{\sum_{l=1}^{k} q_l|\psi_l\rangle\langle\psi_l| :\ \psi_l \in \mathcal{M}\}$.

When the randomization strategy is allowed, modified values are given by

$$\widetilde{V}_U := \inf_{\pi\in\mathcal{P}(\Theta)}\sup_{\sigma\in\mathcal{S}(\mathcal{H})}\sum_{j=1}^{k}\pi_j\langle\psi_j|\sigma|\psi_j\rangle,$$

$$\widetilde{V}_L := \sup_{\sigma\in\mathcal{S}(\mathcal{H})}\inf_{\pi\in\mathcal{P}(\Theta)}\sum_{j=1}^{k}\pi_j\langle\psi_j|\sigma|\psi_j\rangle,$$

and

$$\widetilde{V}_{L,B} := \sup_{\sigma\in\mathrm{co}\mathcal{M}}\inf_{\pi\in\mathcal{P}(\Theta)}\sum_{j=1}^{k}\pi_j\langle\psi_j|\sigma|\psi_j\rangle.$$

This last value represents the maximum detection probability for the worst case when Bob chooses a reference state from the same model $\mathcal{M}$. For example, both Alice and Bob have the same optical device. Clearly, $\widetilde{V}_U \geqslant \widetilde{V}_L \geqslant \widetilde{V}_{L,B}$ holds. Other obvious equalities are

$$\widetilde{V}_L = \sup_{\sigma\in\mathcal{S}(\mathcal{H})}\inf_j\langle\psi_j|\sigma|\psi_j\rangle \geqslant V_L$$

and

$$\widetilde{V}_{L,B} = \sup_{\sigma\in\mathrm{co}\mathcal{M}}\inf_j\langle\psi_j|\sigma|\psi_j\rangle.$$

Since $\sup_{\sigma\in\mathcal{S}(\mathcal{H})}\mathrm{Tr}\sigma X = \|X\|$, we obtain

$$\widetilde{V}_U = \inf_{\pi\in\mathcal{P}(\Theta)}\|\rho_\pi\| = V_U,$$

where $\rho_\pi := \sum_{j=1}^{k}\pi_j|\psi_j\rangle\langle\psi_j|$.

If each state is distinguished from the other states completely, $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, then, due to the symmetry of the model, we obtain the uniform prior $\pi_* = 1/k$ as the least favorable prior, which is very natural.

We also mention another approach by Hirota and Ikehara [26]. They deal with quantum signal identification and investigate the relation between minimaxity and Bayes' rule. Classically, their approach is close to point estimation, while ours is close to the predictive distribution in statistics.

III. MAIN RESULT

Theorem 3.1. When the above discrete model ($|\mathcal{M}| < \infty$) is assumed, then $\widetilde{V}_L = \widetilde{V}_U$. Thus, the game has a finite value. There exists a Bob's maximin strategy σ_* and a least favorable prior π_*.

Proof. It is enough to show that $\widetilde{V}_U \leqslant \widetilde{V}_L$ and it is easily shown in the same manner as the usual minimax theorem for finite cases. See Theorem 1, p. 82 in Ferguson [18]. ∎

Unfortunately, there exists the case where $\widetilde{V}_U > \widetilde{V}_{L,B}$. Thus, we seek another version for $\widetilde{V}_{L,B}$. Let us define

$$\widetilde{V}_{U,B} := \inf_{\pi\in\mathcal{P}(\Theta)} \sup_{\sigma\in\mathrm{co}\mathcal{M}} \sum_{j=1}^{k} \pi_j F(\psi_j,\sigma).$$

Obviously,

$$\widetilde{V}_{U,B} = \inf_{\pi\in\mathcal{P}(\Theta)} \sup_j \langle\psi_j|\rho_\pi|\psi_j\rangle \leqslant \inf_{\pi\in\mathcal{P}(\Theta)} \|\rho_\pi\| = \widetilde{V}_U$$

holds. Using the above value, the following minimax theorem holds.

Theorem 3.2. When the above discrete model ($|\mathcal{M}| < \infty$) is assumed, then $\widetilde{V}_{L,B} = \widetilde{V}_{U,B}$. Thus, the game has a finite value. There exists a Bob's maximin strategy $\sigma_{*,B}$ and a least-favorable prior $\pi_{*,B}$.

In practice, it is enough for Bob to prepare the physical device that generates the same sets of reference states as Alice has. The above theorem is a direct consequence of Ferguson's general result. If Bob is not allowed to randomize the detection measurement, which seems more realistic in some situations, then the following inequalities hold.

$$V_L \leqslant V_U = \inf_{\pi\in\mathcal{P}(\Theta)} \|\rho_\pi\|.$$

In the case where that the converse holds, there exists a Bob's maximin strategy ϕ_* and a least favorable prior π_*. Necessary and sufficient conditions that $V_L = V_U$ holds are unknown. The following sufficient conditions are immediate.

Theorem 3.3 When $k = 2$ or when all vectors $\psi_i \in \mathcal{M}$ are orthogonal to each other, then $V_U = V_L$. The least favorable prior is the uniform one.

Here we make some comments. When $d = \dim\mathcal{H} = 2$, the von Neumann entropy is written as

$$H(\rho_\pi) = -\|\rho_\pi\|\ln\|\rho_\pi\| - (1-\|\rho_\pi\|)\ln(1-\|\rho_\pi\|).$$

Thus, $H(\rho_\pi)$ is a decreasing function of the norm $\|\rho_\pi\|$ for $\frac{1}{2} \leqslant \|\rho_\pi\| \leqslant 1$. Maximizing the von Neumann entropy of ρ_π yields the least favorable prior π_*,

$$\begin{aligned}\pi_{\mathrm{ME}} &:= \arg\max\{H(\rho_\pi) : \pi\in\mathcal{P}(\Theta)\}\\ &= \arg\min\{\|\rho_\pi\| : \pi\in\mathcal{P}(\Theta)\} =: \pi_*.\end{aligned}$$

When $d \geqslant 3$, generally π_{ME} does not agree with π_* except for $\|\rho_\pi\| = 1/d$ or equivalently $\rho_\pi = I/d$. Apart from classical cases, the above least-favorable prior and maximum entropy prior are generally not unique. For example, two different orthonormal bases $\{e_j\}$ and $\{f_j\}$, due to completeness condition, yield the same state $I/k = \sum_j |e_j\rangle\langle e_j|/k = \sum_j |f_j\rangle\langle f_j|/k$.

Now we also mention the infinite cases ($|\Theta| = \infty$). Let Θ be a compact subset of a Euclidean space and $\theta\in\Theta\mapsto\rho_\theta\in\mathcal{S}(\mathcal{H})$ be a one-to-one continuous mapping, where the continuity of ρ_θ is defined by the continuity of each function $\theta\mapsto\langle\psi|\rho_\theta|\psi\rangle$, $\psi\in\mathcal{S}(\mathcal{H})$. Let $\mathcal{P}(\Theta)$ denote the set of all probability measures $\pi(d\theta)$ defined on the Borel subsets of Θ. It is known that $\mathcal{P}(\Theta)$ is compact in the weak topology when Θ is compact. In the sense of weak convergence, we also define

$$\rho_\pi = \int \rho_\theta \pi(d\theta).$$

We now extend the definitions and equalities in the last section to the infinite cases.

$$\begin{aligned}\widetilde{V}_L &:= \sup_{\sigma\in\mathcal{S}(\mathcal{H})} \inf_{\pi\in\mathcal{P}(\Theta)} \int F(\rho_\theta,\sigma)\pi(d\theta)\\ &= \sup_{\sigma\in\mathcal{S}(\mathcal{H})} \inf_{\theta\in\Theta} F(\rho_\theta,\sigma),\\ \widetilde{V}_U &:= \inf_{\pi\in\mathcal{P}(\Theta)} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} \int F(\rho_\theta,\sigma)\pi(d\theta)\\ &= \inf_{\theta\in\Theta} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} F(\rho_\theta,\sigma).\end{aligned}$$

Again, by definition, we easily see $\widetilde{V}_U \geqslant \widetilde{V}_L$.

Theorem 3.4. In the above condition, for a parametric model $\mathcal{M} := \{\rho_\theta : \theta\in\Theta\} \subseteq \mathcal{S}(\mathcal{H})$,

$$\widetilde{V}_L = \widetilde{V}_U \tag{3}$$

holds. There exists a Bob's maximin strategy σ_* and a least-favorable prior π_*.

Proof. First, we show that there exists a prior that achieves the infimum

$$\inf_{\pi\in\mathcal{P}(\Theta)} \|\rho_\pi\|.$$

Since $\mathcal{P}(\Theta)$ is compact, it is enough to show that $\|\rho_\pi\|$ is continuous with respect to π. We use the following lemma, whose proof is in the appendix.

Lemma 3.5. When $\pi_n \to \pi$, $\|\rho_{\pi_n} - \rho_\pi\| \to 0$.

From the lemma, we see that $\|\rho_{\pi_n}\| - \|\rho_\pi\| \to 0$ as $\pi_n \to \pi$, which implies the continuity of $\|\rho_\pi\|$.

Next, we show the equality (3). It is enough to show $\widetilde{V}_L \geqslant \widetilde{V}_U$. Let ϵ be a fixed positive number. Since Θ is compact, there exists a finite set

$$\Theta_\epsilon := \{\theta_1,\ldots,\theta_k\} \subseteq \Theta$$

such that

$$\forall\quad \theta\in\Theta,\quad \exists\quad \theta_j \quad \text{s.t.} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} |F(\rho_\theta,\sigma) - F(\rho_{\theta_j},\sigma)| < \epsilon.$$

Now we set $F_\epsilon(\sigma) := \inf_{\theta\in\Theta_\epsilon} F(\rho_\theta,\sigma) = \min_j F(\rho_{\theta_j},\sigma)$. Then for arbitrary $\sigma\in\mathcal{S}(\mathcal{H})$ and $\theta\in\Theta$, we choose θ_j such that $F(\rho_\theta,\sigma) > F(\rho_{\theta_j},\sigma) - \epsilon$. Hence,

$$F(\rho_\theta,\sigma) > F(\rho_{\theta_j},\sigma) - \epsilon \geqslant \min_j F(\rho_{\theta_j},\sigma) - \epsilon = F_\epsilon(\sigma) - \epsilon$$

and, taking the infimum with respect to θ,

$$\inf_{\theta\in\Theta} F(\rho_\theta,\sigma) \geqslant F_\epsilon(\sigma) - \epsilon \tag{4}$$

holds for arbitrary $\sigma\in\mathcal{S}(\mathcal{H})$. By Theorem 3.1,

$$\begin{aligned}\sup_{\sigma\in\mathcal{S}(\mathcal{H})} F_\epsilon(\sigma) &= \inf_{\theta\in\Theta_\epsilon} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} F(\rho_\theta,\sigma) = \inf_{\pi\in\mathcal{P}(\Theta_\epsilon)} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} F(\rho_\theta,\sigma)\\ &\geqslant \inf_{\pi\in\mathcal{P}(\Theta)} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} F(\rho_\theta,\sigma) = \widetilde{V}_U.\end{aligned}$$

Finally, from Eq. (4) we have

$$\widetilde{V}_L = \sup_{\sigma\in\mathcal{S}(\mathcal{H})} \inf_{\theta\in\Theta} F(\rho_\theta,\sigma) > \sup_{\sigma\in\mathcal{S}(\mathcal{H})} F_\epsilon(\sigma) - \epsilon \geqslant \widetilde{V}_U - \epsilon.$$

Since ϵ is arbitrary, we obtain $\widetilde{V}_L \geqslant \widetilde{V}_U$.

Bob's maximin strategy is given by the first eigenvector ϕ_* of the Bayesian mixture

$$\rho_* := \int \rho_\theta \pi_*(d\theta),$$

where π_* is the least favorable prior, which achieves $\widetilde{V}_U$. ■

In our setting, the extension of Theorem 3.2 to infinite cases is also shown. While the above theorem includes noncompact set $\mathcal{S}(\mathcal{H})$, the restricted class $\overline{\mathrm{co}\mathcal{M}}$ is shown to be compact, which makes the proof easier than Theorem 3.4 and is, thus, omitted.

IV. EXAMPLES

A. Two nonorthogonal states

In order to show how our method is applicable to a pure-states model, we illustrate a simple example first. Suppose that we have two known nonorthogonal pure states $\mathcal{M} = \{\psi_1,\psi_2\}$. The Bayesian mixture then is given by $\rho_a(\psi_1,\psi_2) := a|\psi_1\rangle\langle\psi_1| + (1-a)|\psi_2\rangle\langle\psi_2|$, $0 \leqslant a \leqslant 1$. Since $\|\rho_a(\psi_2,\psi_1)\| = \|\rho_a(\psi_1,\psi_2)\|$, we obtain the following inequality:

$$\begin{aligned}\left\|\rho_{\frac{1}{2}}(\psi_1,\psi_2)\right\| &\leqslant \tfrac{1}{2}\|\rho_a(\psi_1,\psi_2)\| + \tfrac{1}{2}\|\rho_a(\psi_2,\psi_1)\| \\ &= \|\rho_a(\psi_1,\psi_2)\|.\end{aligned}$$

Thus,

$$\left\|\rho_{\frac{1}{2}}(\psi_1,\psi_2)\right\| \leqslant \min_a \|\rho_a(\phi,\psi)\|.$$

and $a = \frac{1}{2}$ achieves the minimum. A least-favorable prior for a two-pure-states model is always uniform, $\pi(\psi_1) = \pi(\psi_2) = \frac{1}{2}$, which agrees with our intuition.

If we have more than two pure states, then there are nontrivial cases. The above setting seems too simple, but still we often see such models in quantum control [27].

B. Continuous model with one outlier

Let us consider the following model:

$$\mathcal{M} := \{|\psi(\theta)\rangle \in \mathbb{C}^2 : \theta \in \Theta := [-1/6\pi, 1/6\pi]\} \cup \{|\psi_O\rangle\},$$

where $|\psi(\theta)\rangle := (\cos\theta/2\ \sin\theta/2)^\top$ and $|\psi_O\rangle := |\psi(2/3\pi)\rangle = (1/2\ \sqrt{3}/2)^\top$ is called an *outlier*. In the above model, physical intuition no longer works. If the equally probable criterion is applied, then we assign zero to the outlier ψ_O and obtain the uniform distribution $\pi_U(\theta)$ over Θ.

Very interestingly, the least-favorable prior assigns the weight only to two states and is given by

$$\pi(2/3\pi) = \pi(-1/6\pi) = \tfrac{1}{2}.$$

The maximin strategy is given by $\psi_* = \psi(1/4\pi)$. We obtain the infimum of the norm $\|\rho_\pi\|$,

$$\rho_* = \frac{1}{2}\begin{pmatrix} 1 + \frac{\sqrt{3}-1}{4} & \frac{\sqrt{3}-1}{4} \\ \frac{\sqrt{3}-1}{4} & 1 - \frac{\sqrt{3}-1}{4}\end{pmatrix}$$

and

$$\|\rho_*\| = \frac{1}{2} + \frac{\sqrt{6}-\sqrt{2}}{8}.$$

The uniform strategy is worse than the above. Indeed, the uniform prior over Θ yields

$$\rho_{\pi_U} = \frac{1}{2}\begin{pmatrix} 1 + \frac{3}{\pi} & 0 \\ 0 & 1 - \frac{3}{\pi}\end{pmatrix}$$

and clearly $\|\rho_{\pi_U}\| = \frac{1}{2} + \frac{3}{2\pi} > \|\rho_*\|$.

C. Case where randomization is strictly better

Let us consider the following model:

$$\begin{aligned}\mathcal{M} :&= \{|\psi(\theta,\xi)\rangle \in \mathbb{C}^2\}, \\ \psi(\theta,\xi) :&= (e^{-i\xi/2}\cos\theta/2 e^{i\xi/2}\sin\theta/2)^\top, \\ \theta &\in [1/4\pi, 3/4\pi], \quad \xi \in [0,2\pi).\end{aligned}$$

In the above model, the maximin strategy is given by either $\psi_{*,N} = (1\ 0)^\top$ or $\psi_{*,S} = (0\ 1)^\top$. Then

$$V_L = \frac{1}{2} - \frac{\sqrt{2}}{4} < \widetilde{V}_L = \frac{1}{2}(= \widetilde{V}_U)$$

holds.

D. Symmetric pure-states model

Next let us consider the following model, $\mathcal{M} = \{\rho_1,\ldots,\rho_k\}$, where $\rho_j = |\psi_j\rangle\langle\psi_j|$ and we assume that there exists a unitary U that satisfies

$$\rho_{l+1} = U\rho_l U^*, \quad l = 1,2,\ldots,k,$$

where $\rho_{k+1} := \rho_1$. The least-favorable prior then is given by

$$\pi_* = (1/k,\ldots,1/k).$$

We set

$$\rho_* := \frac{1}{k}\sum_l \rho_l.$$

The maximin strategy is given by the first eigenvector $|\varphi_\rho\rangle$ of ρ_*. Indeed, from $[\rho_*,U] = 0$, we obtain $[|\varphi_\rho\rangle\langle\varphi_\rho|,U] = 0$. Thus, $|\langle\varphi_\rho|\psi_j\rangle|^2 = V$ is a constant and

$$V \leqslant V_L \leqslant \widetilde{V}_L.$$

On the other hand,

$$V_U \leqslant \|\rho_*\| = \langle\varphi_*|\rho_*|\varphi_*\rangle = \sum_j \frac{1}{k}|\langle\varphi_\rho|\psi_j\rangle|^2 = V.$$

Thus, $V_L = \widetilde{V}_L = \widetilde{V}_U = V_U$ holds.

We often see such examples in Helstrom [22] and Eldar *et al.* [28,29], where they mainly discuss quantum signal identification. They might be solved by using the group covariant measurement theoretically [21]. Previous works focus only on the best measurement and not on the choice of a noninformative prior. In the above model, we agree with using the uniform prior. However, even only a slight modification of the model could spoil the justification on the usage of the uniform prior. On the other hand, our result applies to models that have no group symmetry. Even if we remove some states, $\psi_{l_1},\ldots,\psi_{l_m}$ such that

$$\rho_* \in \mathrm{co}\big\{\mathcal{M} \setminus \{\psi_{l_1},\ldots,\psi_{l_m}\}\big\},$$

the same equality still holds since $[\rho_*,U]=0$. The maximin strategy remains the same, but the least-favorable prior changes and it is not the uniform prior. Although our result is based on a rule that differs from that of previous authors, it suggests that the uniform prior should be changed for less-symmetric models.

E. Rotational model

Straightforward generalization of the above result is as follows. Let J be a Hermitian operator and set

$$|\psi(\theta)\rangle := e^{-i\theta J}|\psi_0\rangle.$$

The periodic condition

$$|\psi(\theta+2\pi)\rangle = |\psi(\theta)\rangle$$

is also imposed. It corresponds to rotation around a certain axis. We then obtain the one-parameter model

$$\mathcal{M}([0,2\pi)) := \{\rho_\theta = |\psi(\theta)\rangle\langle\psi(\theta)| \;:\; \theta \in [0,2\pi)\}.$$

Again, $V_L = \widetilde{V}_L = \widetilde{V}_U = V_U$ holds. Indeed, we introduce the invariant integral

$$\rho_* := \int \frac{d\theta}{2\pi} e^{-i\theta J}|\psi_0\rangle\langle\psi_0|e^{i\theta J}$$

and the minimax strategy is given by the first eigenvector of ρ_*. Even if we restrict the parameter θ to a certain region $\Theta \subset [0,2\pi)$ such that $\rho_* \in \mathrm{co}\{\mathcal{M}(\Theta)\}$, the above result holds.

For a compact group G, its (projective) unitary representation V_g and finite invariant measure $\mu(dg)$, it is possible to construct a more general result using the above technique.

V. CONCLUDING REMARKS

In the present paper, we define a quantum detection game and show that a least favorable prior with respect to the game is one candidate for a noninformative prior. Unfortunately, it is not uniquely determined, which is mainly due to the quantumness. Our result is mainly of theoretical interest but it suggests that a certain minimax estimation is deeply related to the choice of noninformative prior even in pure-state models. In more specific models, we need to investigate the performance of a quantum pure-state estimation based on our prior distribution compared with others, which is left for future study.

From geometrical viewpoints, our result is also stimulating because noninformative priors are often related to certain metrics on parametric models. But again we have no classical counterpart in the pure-states model and, thus, need further investigation in order to find a relevant metric.

ACKNOWLEDGMENTS

The author was supported by Kakenhi for Young Researchers (B) (No. 20700250) and JST PRESTO program.

APPENDIX

We present a complete proof of Lemma 3.5. First, from the assumption of ρ_θ, $f_\varphi(\theta) := \langle\varphi|\rho_\theta|\varphi\rangle$ is a bounded continuous function with respect to θ for an arbitrary $\varphi \in \mathcal{H}$. Therefore, when $\pi_n \to \pi$,

$$\begin{aligned}\int f_\varphi(\theta)\pi_n(d\theta) - \int f_\varphi(\theta)\pi(d\theta) \\ = \langle\varphi|\left\{\int \rho_\theta\pi_n(d\theta) - \int \rho_\theta\pi(d\theta)\right\}|\varphi\rangle \\ = \mathrm{Tr}\big(\rho_{\pi_n} - \rho_\pi\big)|\varphi\rangle\langle\varphi| \to 0,\ \forall\, \varphi \in \mathcal{H}\end{aligned}$$

holds. We may replace $|\varphi\rangle\langle\varphi|$ with a finite-rank density operator in the above. Now we show that the weak convergence $\rho_{\pi_n} \to \rho_\pi$ implies the strong convergence $\|\rho_{\pi_n} - \rho_\pi\| \to 0$.

Since $\mathrm{Tr}\rho_\pi = 1$, for an arbitrary positive number ϵ there exists a finite-dimensional projection P_ϵ such that

$$\mathrm{Tr}\rho_\pi P_\epsilon \geqslant 1-\epsilon.$$

Since $\mathrm{Tr}(\rho_{\pi_n} - \rho_\pi)P_\epsilon \to 0$, we may take a sufficiently large N satisfying

$$n \geqslant N \Rightarrow \big|\mathrm{Tr}\rho_{\pi_n}P_\epsilon - \mathrm{Tr}\rho_\pi P_\epsilon\big| < \epsilon.$$

It follows that

$$0 \leqslant \mathrm{Tr}\rho_\pi Q_\epsilon \leqslant \epsilon, \quad 0 \leqslant \mathrm{Tr}\rho_{\pi_n}Q_\epsilon < 2\epsilon, \quad \forall\, n \geqslant N,$$

where we set $Q_\epsilon := I - P_\epsilon$.

Now we prove that $n \geqslant N \Rightarrow \|\rho_{\pi_n} - \rho_\pi\| < C\epsilon$, where $C > 0$ is a positive constant independent of n. First, we decompose

$$\begin{aligned}\|\rho_{\pi_n} - \rho_\pi\| &= \|(P_\epsilon + Q_\epsilon)\big(\rho_{\pi_n} - \rho_\pi\big)(P_\epsilon + Q_\epsilon)\| \\ &= \|P_\epsilon\big(\rho_{\pi_n} - \rho_\pi\big)P_\epsilon\| + \|Q_\epsilon\big(\rho_{\pi_n} - \rho_\pi\big)Q_\epsilon\| \\ &\quad + \|Q_\epsilon\big(\rho_{\pi_n} - \rho_\pi\big)P_\epsilon\| + \|P_\epsilon\big(\rho_{\pi_n} - \rho_\pi\big)Q_\epsilon\|\end{aligned} \tag{A1}$$

The first term in the last equality (A1) goes to zero as n increases.

Now we omit ϵ and set $X_n := \rho_{\pi_n} - \rho_\pi$. In order to evaluate the third term in the decomposition (A1), we need some inequalities,

$$\|X\| \leqslant \|X\|_2 := \{\mathrm{Tr}X^*X\}^{\frac{1}{2}}, \tag{A2}$$

$$\sigma^2 \leqslant \sigma, \quad \sigma \in \mathcal{S}(\mathcal{H}), \tag{A3}$$

$$2\rho^2 + 2\sigma^2 \geqslant (\rho - \sigma)^2, \quad \rho,\sigma \in \mathcal{S}(\mathcal{H}) \tag{A4}$$

Now we evaluate the term $\|Q(\rho_{\pi_n} - \rho_\pi)P\| = \|QX_nP\|$.

First, by the inequality (A2),

$$\|QX_nP\| \leqslant \{\mathrm{Tr}QX_nP^2X_nQ\}^{\frac{1}{2}} \leqslant \{\mathrm{Tr}QX_n^2Q\}^{\frac{1}{2}}$$

holds. We then use the inequalities (A4) and (A3)

$$\begin{aligned}\mathrm{Tr}QX_n^2Q &= \mathrm{Tr}Q(\rho_{\pi_n} - \rho_\pi)^2Q \leqslant 2\{\mathrm{Tr}Q\rho_{\pi_n}^2Q + \mathrm{Tr}Q\rho_\pi^2Q\}\\ &\leqslant 2\{\mathrm{Tr}Q\rho_{\pi_n}Q + \mathrm{Tr}Q\rho_\pi Q\}\\ &= 2\{\mathrm{Tr}Q\rho_{\pi_n} + \mathrm{Tr}Q\rho_\pi\} < 6\epsilon.\end{aligned}$$

The term $\|QX_nP\|$ is also evaluated in the same way. Thus, the third term is evaluated as

$$\|QX_nP\| + \|PX_nQ\| \leqslant 2\sqrt{6\epsilon}.$$

Finally, we evaluate the second term in the decomposition (A1). We use the inequalities with respect to the trace norm $\|X\|_1$.

$$\|X\| \leqslant \|X\|_1 := \mathrm{Tr}|X|, \tag{A5}$$

$$\|X - Y\|_1 \leqslant \|X\|_1 + \|Y\|_1, \tag{A6}$$

By the inequalities (A5) and (A6),

$$\begin{aligned}\|QX_nQ\| \leqslant \|QX_nQ\|_1 &\leqslant \|Q\rho_{\pi_n}Q\|_1 + \|Q\rho_\pi Q\|_1\\ &= \mathrm{Tr}Q\rho_{\pi_n} + \mathrm{Tr}Q\rho_\pi < 4\epsilon\end{aligned}$$

holds. Thus, putting together and choosing a sufficiently large N, we obtain

$$n \geqslant N \quad \Rightarrow \|\rho_{\pi_n} - \rho_\pi\| < C\epsilon.$$

■

[1] W. K. Wootters, Found. Phys. **20**, 1365 (1990).
[2] M. J. W. Hall, Phys. Lett. A **242**, 123 (1998).
[3] P. B. Slater, Phys. Lett. A **247**, 1 (1998).
[4] P. B. Slater, arXiv:quant-ph/0507014.
[5] M. Srednicki, Phys. Rev. A **71**, 052107 (2005).
[6] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[7] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
[8] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[9] V. Bužek, R. Derka, G. Adam, and P. L. Knight, Ann. Phys. (NY) **266**, 454 (1998).
[10] R. Derka, V. Bužek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).
[11] F. Tanaka and F. Komaki, Phys. Rev. A **71**, 052323 (2005).
[12] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).
[13] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (Springer, New York, 2001).
[14] H. Jeffreys, *Theory of Probability* (Oxford University Press, Oxford, 1961).
[15] M. Hayashi, J. Phys.: Conf. Ser. **143**, 012017 (2009).
[16] M. Hayashi, Commun. Math. Phys. **293**, 171 (2010).
[17] B. S. Clarke and A. R. Barron, J. Statist. Plan. Infer. **41**, 37 (1994).
[18] T. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach* (Academic Press, New York, 1967).
[19] J. M. Bernardo, J. R. Statist. Soc. B **41**, 113 (1979).
[20] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[21] S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[22] C. W. Helstrom, *Quantum Detection Theory* (Academic Press, New York, 1976).
[23] J. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer, New York, 1985).
[24] L. Davisson and A. Leon-Garcia, IEEE Trans. Inform. Theory **26**, 166 (1980).
[25] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer, Berlin, 2004).
[26] O. Hirota and S. Ikehara, IEEE Trans. IECE Jpn. **65**, 627 (1982).
[27] A. M. Brańczyk, P. E. M. F. Mendonça, A. Gilchrist, A. C. Doherty, and S. D. Bartlett, Phys. Rev. A **75**, 012329 (2007).
[28] Y. C. Eldar, A. Megretski, and G. C. Verghese, IEEE Trans. Inform. Theory **49**, 1007 (2003).
[29] Y. C. Eldar, A. Megretski, and G. C. Verghese, IEEE Trans. Inform. Theory **50**, 1198 (2004).