entropy

MDPI

Article

Minimax Estimation of Quantum States Based on the Latent Information Priors

Takayuki Koyama [1], Takeru Matsuda [2] and Fumiyasu Komaki [2,3,*]

1 FANUC Corporation, 3580 Furubaba Shibokusa Oshino-mura, Yamanashi 401-0597, Japan; takayuki.koyama0804@gmail.com
2 Department of Mathematical Informatics, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan; matsuda@mist.i.u-tokyo.ac.jp
3 RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
* Correspondence: komaki@mist.i.u-tokyo.ac.jp; Tel.: +81-3-5841-6941

Received: 13 September 2017; Accepted: 13 November 2017; Published: 16 November 2017

Abstract: We develop priors for Bayes estimation of quantum states that provide minimax state estimation. The relative entropy from the true density operator to a predictive density operator is adopted as a loss function. The proposed prior maximizes the conditional Holevo mutual information, and it is a quantum version of the latent information prior in classical statistics. For one qubit system, we provide a class of measurements that is optimal from the viewpoint of minimax state estimation.

Keywords: Bayes; conditional Holevo mutual information; latent information prior; predictive density operator; quantum estimation

---

1. Introduction

In quantum mechanics, the outcome of a measurement is subject to a probability distribution determined from the quantum state of the measured system and the measurement performed. The task of estimating the quantum state from the outcome of measurement is called the quantum estimation and it is a fundamental problem in quantum statistics [1–3]. Tanaka and Komaki [4] and Tanaka [5] discussed quantum estimation using the framework of statistical decision theory and showed that Bayesian methods provide better estimation than the maximum likelihood method. In Bayesian methods, we need to specify a prior distribution on the unknown parameters of the quantum states. However, the problem of prior selection has not been fully discussed for quantum estimation [6].

The quantum state estimation problem is related to the predictive density estimation problem in classical statistics [7]. This is a problem of predicting the distribution of an unobserved variable y based on an observed variable x. Suppose $(x, y) \sim p(x, y \mid \theta)$, where θ denotes an unknown parameter. Based on the observed x, we predict the distribution $p(y \mid x, \theta)$ of y using a predictive density $\hat{p}(y \mid x)$. The plug-in predictive density is defined as $\hat{p}_{\text{plug-in}}(y \mid x) = p(y \mid x, \hat{\theta}(x))$, where $\hat{\theta}(x)$ is some estimate of θ from x. The Bayesian predictive density with respect to a prior distribution $\mathrm{d}\pi(\theta)$ is defined as

$$\hat{p}_\pi(y \mid x) = \int p(y \mid x, \theta)\mathrm{d}\pi(\theta \mid x) = \frac{\int p(y \mid x, \theta)p(x \mid \theta)\mathrm{d}\pi(\theta)}{\int p(x \mid \theta)\mathrm{d}\pi(\theta)}, \tag{1}$$

where $\mathrm{d}\pi(\theta \mid x)$ is the posterior distribution. We compare predictive densities using the framework of statistical decision theory. Specifically, a loss function $L(q, p)$ is introduced that evaluates the difference between the true density q and the predictive density p. Then, the risk function $R(\theta, \hat{p})$ is defined as the average loss when the true value of the parameter is θ:

$$R(\theta, \hat{p}) = \int L(p(y \mid x, \theta), \hat{p}(y \mid x)) p(x \mid \theta) \mathrm{d}x.$$

A predictive density $\hat{p}_*$ is called minimax if it minimizes the maximum risk among all predictive densities:

$$\max_{\theta} R(\theta, \hat{p}_*) = \min_{\hat{p}} \max_{\theta} R(\theta, \hat{p}). \tag{2}$$

We adopt the Kullback–Leibler divergence

$$L(q, p) = \int q(x) \log \frac{q(x)}{p(x)} \mathrm{d}x \tag{3}$$

as a loss function, since it satisfies many desirable properties compared to other loss functions such as the Hellinger distance and the total variation distance [8]. Under this setting, Aitchison [9] proved

$$R(\pi, \hat{p}_\pi) = \min_{\hat{p}} R(\pi, \hat{p}), \tag{4}$$

where

$$R(\pi, \hat{p}) = \int R(\theta, \hat{p}) \pi(\theta) \mathrm{d}\theta$$

is called the Bayes risk. Namely, the Bayesian predictive density $\hat{p}_\pi(y \mid x)$ minimizes the Bayes risk. We provide the proof of Equation (4) in the Appendix A. Therefore, it is sufficient to consider only Bayesian predictive densities from the viewpoint of Kullback–Leibler risk, and the selection of the prior π becomes important.

For the predictive density estimation problem above, Komaki [10] developed a class of priors called the latent information priors. The latent information prior π_{LIP} is defined as a prior that maximizes the conditional mutual information $I_{\theta,y|x}(\pi)$ between the parameter θ and the unobserved variable y given the observed variable x. Namely,

$$I_{\theta,y|x}(\pi_{\mathrm{LIP}}) = \max_{\pi} I_{\theta,y|x}(\pi),$$

where

$$\begin{aligned} I_{\theta,y|x}(\pi) = & \int \textstyle\sum_{x,y} p(x, y \mid \theta) \log p(x, y \mid \theta) \mathrm{d}\pi(\theta) - \sum_{x,y} p_\pi(x, y) \log p_\pi(x, y) \\ & - \int \textstyle\sum_{x} p(x \mid \theta) \log p(x \mid \theta) \mathrm{d}\pi(\theta) + \sum_{x} p_\pi(x) \log p_\pi(x) \end{aligned} \tag{5}$$

is the conditional mutual information between y and θ given x. Here,

$$p_\pi(x, y) = \int p(x, y \mid \theta) \mathrm{d}\pi(\theta), \quad p_\pi(x) = \int p(x \mid \theta) \mathrm{d}\pi(\theta) \tag{6}$$

are marginal densities. The Bayesian predictive densities based on the latent information priors are minimax under the Kullback–Leibler risk:

$$\max_{\theta} R(\theta, \hat{p}_\pi) = \min_{\hat{p}} \max_{\theta} R(\theta, \hat{p}).$$

The latent information prior is a generalization of the reference prior [11] that is a prior maximizing the unconditional mutual information $I_{\theta,y}(\pi)$ between θ and y.

Now, we consider the problem of estimating the quantum state of a system Y based on the outcome of a measurement on a system X. Suppose the quantum state of the composed system (X, Y) be $\sigma_\theta^{\mathrm{XY}}$ where θ denotes an unknown parameter. We perform a measurement on the system X and obtain the outcome x. Based on the measurement outcome x, we estimate the state of the system Y by a predictive density operator $\rho(x)$. Similarly to the Bayesian predictive density (1), the Bayesian predictive density operator with respect to the prior $\mathrm{d}\pi(\theta)$ is defined as

$$\sigma_{\pi}^{\mathrm{Y}}(x) = \int \sigma_{\theta,x}^{\mathrm{Y}} \mathrm{d}\pi(\theta \mid x) = \frac{\int \sigma_{\theta,x}^{\mathrm{Y}} p(x \mid \theta) \mathrm{d}\pi(\theta)}{\int p(x \mid \theta) \mathrm{d}\pi(\theta)}, \tag{7}$$

where $\mathrm{d}\pi(\theta \mid x)$ is the posterior distribution. Like the predictive density estimation problem discussed above, we compare predictive density operators using the framework of statistical decision theory. There are several possibilities for the loss function $L(\sigma, \rho)$ in quantum estimation such as the fidelity and the trace norm [12]. In this paper, we adopt the quantum relative entropy

$$L(\sigma, \rho) = \operatorname{Tr} \sigma(\log \sigma - \log \rho) \tag{8}$$

as a loss function, since it is a quantum analogue of the Kullback–Leibler divergence (3). Note that the fidelity and the trace norm correspond to the Hellinger distance and the total variation distance in the classical statistics, respectively. Under this setting, Tanaka and Komaki [4] proved that the Bayesian predictive density operators minimize the Bayes risk:

$$\int R(\theta, \sigma_{\pi}^{\mathrm{Y}}) \mathrm{d}\pi(\theta) = \min_{\rho} \int R(\theta, \rho) \mathrm{d}\pi(\theta).$$

This is a quantum version of Equation (4).

From Tanaka and Komaki [4], the selection of the prior becomes important also in quantum estimation. However, this problem has not been fully discussed [6]. In this paper, we provide a quantum version of the latent information priors and prove that they provide minimax predictive density operators. Whereas the latent information prior in the classical case maximizes the conditional Shannon mutual information, the proposed prior maximizes the conditional Holevo mutual information. The Holevo mutual information, which is a quantum version of the Shannon mutual information, is a fundamental quantity in the classical-quantum communication [13]. Our result shows that the conditional Holevo mutual information also has a natural meaning in terms of quantum estimation.

Unlike the classical statistics, the measurement is not unique in quantum statistics. Therefore, selection of the measurement also becomes important. From the viewpoint of minimax state estimation, measurements that minimize the minimax risk are considered to be optimal. We provide a class of optimal measurements for one qubit system. This class includes the symmetric informationally complete measurement [14,15]. These measurements and latent information priors provide robust quantum estimation.

2. Preliminaries

2.1. Quantum States and Measurements

We briefly summarize several notations of quantum states and measurements. Let $\mathcal{H}$ be a separable Hilbert space of a quantum system. A Hermitian operator ρ on $\mathcal{H}$ is called a density operator if it satisfies

$$\operatorname{Tr} \rho = 1, \quad \rho \geq 0.$$

The state of a quantum system is described by a density operator. We denote the set of all density operators on $\mathcal{H}$ as $\mathcal{S}(\mathcal{H})$.

Denote the set of all linear operators on Hilbert space $\mathcal{H}$ by $\mathcal{L}(\mathcal{H})$ and the set of all positive linear operators by $\mathcal{L}_{+}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$. Let Ω be a measurable space of all possible outcomes of a measurement and $\mathcal{B}(\Omega)$ be a σ-algebra of Ω. A map $E : \mathcal{B}(\Omega) \to \mathcal{L}_{+}(\mathcal{H})$ is called a positive operator-valued measure (POVM) if it satisfies $E(\emptyset) = O$, $E(\Omega) = I$, and $E(\cup_i B_i) = \sum_i E(B_i)$, where $B_i \cap B_j = \emptyset$, $\forall B_i \in \mathcal{B}(\mathcal{H})$. Any quantum measurement is represented by a POVM on Ω. In this paper, we mainly assume Ω is finite. In such case, we denote $\Omega = \mathcal{X} = \{1, \ldots, N\}$ and any POVM is represented by a set of positive Hermitian operators $E = \{E_x \mid x \in \mathcal{X}\}$ such that $\sum_{x \in \mathcal{X}} E_x = I$.

The outcome of a measurement E on a quantum system with the state $\rho \in \mathcal{S}(\mathcal{H})$ is distributed with a probability measure

$$\Pr(B) = \operatorname{Tr} E(B)\rho, \quad B \in \mathcal{B}(\Omega).$$

Let X, Y be quantum systems with Hilbert spaces $\mathcal{H}^{\mathrm{X}}$ and $\mathcal{H}^{\mathrm{Y}}$. The Hilbert space of the composed system (X, Y) is given by the tensor product $\mathcal{H}^{\mathrm{X}} \otimes \mathcal{H}^{\mathrm{Y}}$. Suppose the state of this composed system is σ^{XY}. Then, the states of two subsystems can be yielded by the partial trace:

$$\sigma^{\mathrm{X}} = \operatorname{Tr}_{\mathrm{Y}} \sigma^{\mathrm{XY}}, \quad \sigma^{\mathrm{Y}} = \operatorname{Tr}_{\mathrm{X}} \sigma^{\mathrm{XY}}.$$

If a measurement $E = \{E_x \mid x \in \mathcal{X}\}$ is performed on the system X and the measurement outcome is x, then the state of the system Y becomes

$$\sigma_x^{\mathrm{Y}} = \frac{1}{p_x} \operatorname{Tr}_{\mathrm{X}}(E_x \otimes I^{\mathrm{Y}})\sigma^{\mathrm{XY}},$$

where the normalization constant

$$p_x = \operatorname{Tr}(E_x \otimes I^{\mathrm{Y}})\sigma^{\mathrm{XY}}$$

is the probability of the outcome x. Here, I^{Y} is the identity operator on the space $\mathcal{H}^{\mathrm{Y}}$. We call the operator σ_x^{Y} the conditional density operator.

2.2. Quantum State Estimation

We formulate the quantum state estimation problem using the framework of statistical decision theory. Let X and Y be quantum systems with finite-dimensional Hilbert spaces $\mathcal{H}^{\mathrm{X}}$ and $\mathcal{H}^{\mathrm{Y}}$, where $\dim \mathcal{H}^{\mathrm{X}} = d_{\mathrm{X}}$ and $\dim \mathcal{H}^{\mathrm{Y}} = d_{\mathrm{Y}}$.

Suppose the state of the composed system (X, Y) be $\sigma_\theta^{\mathrm{XY}}$, where $\theta \in \Theta$ denotes unknown parameters. We perform a measurement $E = \{E_x \mid x \in \mathcal{X}\}$ on X, observe the outcome $x \in \mathcal{X}$, and estimate the conditional density operator $\sigma_{\theta,x}^{\mathrm{Y}}$ of Y by a predictive density operator $\rho(x)$. As discussed in the introduction (1) and (7), the Bayesian predictive density operator based on a prior $\pi(\theta)$ is defined by

$$\sigma_\pi^{\mathrm{Y}}(x) = \int \sigma_{\theta,x}^{\mathrm{Y}} \mathrm{d}\pi(\theta \mid x) = \frac{\int \sigma_{\theta,x}^{\mathrm{Y}} p(x \mid \theta) \mathrm{d}\pi(\theta)}{\int p(x \mid \theta) \mathrm{d}\pi(\theta)},$$

where $\mathrm{d}\pi(\theta \mid x)$ is the posterior distribution.

To evaluate predictive density operators, we introduce a loss function $L(\sigma, \rho)$ that evaluates the difference between the true conditional density operator σ and the predictive density operator ρ. In this paper, we adopt the quantum relative entropy (8) since it is a quantum analogue of the Kullback–Leibler divergence (3). Then, the risk function $R(\theta, \rho)$ of a predictive density operator ρ is defined by

$$R(\theta, \rho) = \sum_{x \in \mathcal{X}} p(x \mid \theta) \operatorname{Tr} \sigma_{\theta,x}^{\mathrm{Y}} (\log \sigma_{\theta,x}^{\mathrm{Y}} - \log \rho(x)),$$

where

$$p(x \mid \theta) = \operatorname{Tr}(E_x \otimes I^{\mathrm{Y}})\sigma_\theta^{\mathrm{XY}} = \operatorname{Tr} E_x \sigma_\theta^{\mathrm{X}}$$

is the probability of the outcome x. Similarly to the classical case (2), a predictive density operator ρ_* is called minimax if it minimizes the maximum risk among all predictive density operators [16,17]:

$$\max_\theta R(\theta, \rho_*) = \min_\rho \max_\theta R(\theta, \rho).$$

Tanaka and Komaki [4] showed

$$R(\pi, \sigma^{\mathrm{Y}}_\pi) = \min_\rho R(\pi, \rho), \tag{9}$$

where

$$R(\pi, \rho) = \int R(\theta, \rho) \mathrm{d}\pi(\theta)$$

is called the Bayes risk. Namely, the Bayesian predictive density operator minimizes the Bayes risk. This result is a quantum version of Equation (4). Although Tanaka and Komaki [4] considered separable models ($\sigma^{\mathrm{XY}}_\theta = \sigma^{\mathrm{X}}_\theta \otimes \sigma^{\mathrm{Y}}_\theta$), the relation (9) holds also for non-separable models as shown in the Appendix A. Therefore, it is sufficient to consider only Bayesian predictive density operators and the problem of prior selection becomes crucial.

2.3. Notations

For a quantum state family $\{\sigma^{\mathrm{XY}}_\theta \mid \theta \in \Theta\}$, we define another quantum state family

$$\mathcal{M} = \{\oplus_x p(x \mid \theta) \sigma^{\mathrm{Y}}_{\theta,x} \mid \theta \in \Theta\},$$

where

$$\oplus_x p(x \mid \theta) \sigma^{\mathrm{Y}}_{\theta,x} = \begin{pmatrix} p(1 \mid \theta)\sigma^{\mathrm{Y}}_{\theta,1} & O & \cdots & O \\ O & p(2 \mid \theta)\sigma^{\mathrm{Y}}_{\theta,2} & & \vdots \\ \vdots & & \ddots & O \\ O & \cdots & O & p(N \mid \theta)\sigma^{\mathrm{Y}}_{\theta,N} \end{pmatrix}$$

is a density operator in $\mathbb{C}^N \otimes \mathcal{H}^{\mathrm{Y}}$. Since $\dim \mathbb{C}^N \otimes \mathcal{H}^{\mathrm{Y}} = N d_{\mathrm{Y}}$, the state family $\mathcal{M}$ can be regarded as a subset of the Euclidean space $\mathbb{R}^{N^2 d_{\mathrm{Y}}^2 - 1}$. By identifying Θ with $\mathcal{M}$, the parameter space Θ is endowed with the induced topology as a subset of $\mathbb{R}^{N d_{\mathrm{Y}}^2 - 1}$.

Any measurement on the system X is represented by a projective measurement $\{e_{xx} = |x\rangle\langle x| \mid x = 1, \ldots, N\}$, where $\{|1\rangle, \cdots, |N\rangle\}$ is an orthonormal basis of $\mathbb{C}^N$. For every $x \in \mathcal{X}$, we define $S_\theta(x) \in \mathcal{L}_+(\mathcal{H}^{\mathrm{Y}})$ as

$$S_\theta(x) := \mathrm{Tr}_{\mathbb{C}^N}(e_{xx} \otimes I^{\mathrm{Y}})(\oplus_x p(x \mid \theta)\sigma^{\mathrm{Y}}_{\theta,x}) = p(x \mid \theta)\sigma_{\theta,x},$$

which is the unnormalized state of Y conditional on the measurement outcome x. We also define

$$S_\pi(x) = \int S_\theta(x) \mathrm{d}\pi(\theta), \quad p_\pi(x) = \mathrm{Tr}\, S_\pi(x), \quad \sigma_\pi(x) = \frac{S_\pi(x)}{p_\pi(x)}.$$

3. Minimax Estimation of Quantum States

In this section, we develop the latent information prior for quantum state estimation and show that this prior provides a minimax predictive density operator.

In the following, we assume the following conditions:

- Θ is compact.
- For every $x \in \mathcal{X}$, $E_x \neq O$.

- For every $x \in \mathcal{X}$, there exists $\theta \in \Theta$ such that $p(x \mid \theta) = \operatorname{Tr} E_x \sigma_\theta^{\mathrm{X}} > 0$.

The third assumption is achieved by adopting sufficiently small Hilbert space. Namely, if there exists $x \in \mathcal{X}$ such that $p(x \mid \theta) = \operatorname{Tr} E_x \sigma_\theta^{\mathrm{X}} = 0$ for every $\theta \in \Theta$, then we redefine the state space $\mathcal{H}$ as the orthogonal complement of $\operatorname{Ker} E_x$.

Let $\mathcal{P}$ be the set of all probability measures on Θ endowed with the weak convergence topology and the corresponding Borel algebra. By the Prohorov theorem [18] and the first assumption, $\mathcal{P}$ is compact.

When x is fixed, the function $\theta \in \Theta \mapsto S_\theta(x)$ is bounded and continuous. Thus, for every fixed $x \in \mathcal{X}$, the function

$$\pi \in \mathcal{P} \mapsto S_\pi(x) = \int S_\theta(x) \mathrm{d}\pi(\theta)$$

is continuous because $\mathcal{P}$ is endowed with the weak convergence topology and $\dim \mathcal{H}^{\mathrm{Y}} < \infty$. Let $\{\lambda_{x,i}\}_i, \{|\phi_{x,i}\rangle\}_i$ be the eigenvalues and the normalized eigenvectors of the predictive density operator $\rho(x)$. For every predictive density operator ρ, consider the function from $\mathcal{P}$ to $[0, \infty]$ defined by

$$\begin{aligned} D_\rho(\pi) &= \sum_x \operatorname{Tr} S_\pi(x)(\log S_\pi(x) - \log(p_\pi(x)\rho(x))) \\ &= \sum_x \operatorname{Tr} S_\pi(x)(\log S_\pi(x) - (\log p_\pi(x))I - \log \rho(x)) \\ &= \sum_x \operatorname{Tr} S_\pi(x) \log S_\pi(x) - \sum_x p_\pi(x) \log p_\pi(x) \\ &\quad + \sum_{i:\lambda_{x,i} \neq 0} -p_\pi(x)\langle \phi_{x,i} | \sigma_\pi(x) | \phi_{x,i} \rangle \log \lambda_{x,i} \\ &\quad + \sum_{i:\lambda_{x,i} = 0} -p_\pi(x)\langle \phi_{x,i} | \sigma_\pi(x) | \phi_{x,i} \rangle \log \lambda_{x,i}. \end{aligned} \tag{10}$$

The last term in (10) is lower semicontinuous under the definition $0 \log 0 = 0$ [10], since each summand takes either zero or infinity and so the set of $\pi \in \mathcal{P}$ such that this term takes zero is closed. In addition, the other terms in (10) are continuous since the von Neumann entropy is continuous [12]. Therefore, the function $D_\rho(\pi)$ in (10) is lower-semicontinuous.

Now, we prove that the class of predictive density operators that are limits of Bayesian predictive density operators is an essentially complete class. We prepare three lemmas. Lemma 1 is useful for differentiation of quantum relative entropy (see Hiai and Petz [19]). Lemmas 2 and 3 are from Komaki [10].

Lemma 1. *Let A, B be n-dimensional self-adjoint matrices and t be a real number. Assume that $f : (\alpha, \beta) \to \mathbb{R}$ is a continuously differentiable function defined on an interval and assume that the eigenvalues of $A + tB$ are in (α, β) if t is sufficiently close to t_0. Then,*

$$\frac{\mathrm{d}}{\mathrm{d}t} \operatorname{Tr} f(A + tB) \bigg|_{t=t_0} = \operatorname{Tr}(Bf'(A + t_0 B)).$$

Lemma 2 ([10])**.** *Let μ be a probability measure on Θ. Then,*

$$\mathcal{P}_{\varepsilon\mu} = \{\varepsilon\mu + (1 - \varepsilon)\pi \mid \pi \in \mathcal{P}\}$$

is a closed subset of $\mathcal{P}$ for $0 \leq \varepsilon \leq 1$.

Lemma 3 ([10])**.** *Let $f : \mathcal{P} \to [0, \infty]$ be continuous, and let μ be a probability measure on Θ such that $p_\mu(x) := \int p(x \mid \theta) d\mu(\theta) > 0$ for every $x \in \mathcal{X}$. Then, there is a probability measure π_n in*

$$\mathcal{P}_{\mu/n} = \left\{ \frac{1}{n}\mu + \left(1 - \frac{1}{n}\right)\pi \mid \pi \in \mathcal{P} \right\}$$

for every n, such that $f(\pi_n) = \inf_{\pi \in \mathcal{P}_{\mu/n}} f(\pi)$. Furthermore, there exists a convergent subsequence $\{\pi'_m\}_{m=1}^{\infty}$ of $\{\pi_n\}_{n=1}^{\infty}$ and the equality $f(\pi'_\infty) = \inf_{\pi \in \mathcal{P}} f(\pi)$ holds, where $\pi'_\infty = \lim_{m\to\infty} \pi'_m$.

By using these results, we obtain the following theorem, which is a quantum version of Theorem 1 of Komaki [10].

Theorem 1.

(1) *Let $\rho(x)$ be a predictive density operator. If there exists a prior $\hat{\pi}^\rho \in \mathcal{P}$ such that $D_\rho(\hat{\pi}^\rho) = \inf_{\pi \in \mathcal{P}} D_\rho(\pi)$ and $p_{\hat{\pi}^\rho}(x) > 0$ for every $x \in \mathcal{X}$, then $R(\theta, \sigma_{\hat{\pi}^\rho}(x)) \le R(\theta, \rho(x))$ for every $\theta \in \Theta$.*

(2) *For every predictive density operator ρ, there exists a convergent prior sequence $\{\pi_n^\rho\}_{n=1}^{\infty}$ such that $D_\rho(\lim_{n\to\infty} \pi_n^\rho) = \inf_{\pi\in\mathcal{P}} D_\rho(\pi)$, $\lim_{n\to\infty} \sigma_{\pi_n^\rho}(x)$ exists, and $R(\theta, \lim_{n\to\infty} \sigma_{\pi_n^\rho}(x)) \le R(\theta, \rho)$ for every $\theta \in \Theta$.*

Next, we develop priors that provide minimax predictive density operators. Let x be a random variable, which represents the outcome of the measurement, i.e., $x \sim p(\cdot \mid \theta)$. Then, as a quantum analogue of the conditional mutual information (5), we define the conditional Holevo mutual information [13] between the quantum state σ_x^Y of Y and the parameter θ given the measurement outcome x as

$$\begin{aligned} I_{\theta,\sigma|x}(\pi) &= \textstyle\int \sum_x \operatorname{Tr} S_\theta(x) \log S_\theta(x) \mathrm{d}\pi(\theta) - \sum_x \operatorname{Tr} S_\pi(x) \log S_\pi(x) \\ &\quad - \textstyle\int \sum_x p(x \mid \theta) \log p(x \mid \theta) \mathrm{d}\pi(\theta) + \sum_x p_\pi(x) \log p_\pi(x) \\ &= \textstyle\int \sum_x p(x \mid \theta) \operatorname{Tr} \sigma_{\theta,x} (\log \sigma_{\theta,x} - \log \sigma_{\pi,x}) \mathrm{d}\pi(\theta), \end{aligned} \tag{11}$$

which is a function of $\pi \in \mathcal{P}$. Here, we used

$$\begin{aligned} \sum_x \operatorname{Tr} S_\theta(x) \log S_\theta(x) &= \sum_x p(x \mid \theta) \operatorname{Tr} \sigma_{\theta,x} (\log p(x \mid \theta) I + \log \sigma_{\theta,x}) \\ &= \sum_x p(x \mid \theta) \log p(x \mid \theta) + \sum_x p(x \mid \theta) \operatorname{Tr} \sigma_{\theta,x} \log \sigma_{\theta,x} \end{aligned}$$

and

$$\begin{aligned} \sum_x \operatorname{Tr} S_\pi(x) \log S_\pi(x) &= \sum_x p_\pi(x) \operatorname{Tr} \sigma_\pi(x) (\log p_\pi(x) I + \log \sigma_\pi(x)) \\ &= \sum_x p_\pi(x) \log p_\pi(x) + \sum_x p_\pi(x) \operatorname{Tr} \sigma_\pi(x) \log \sigma_\pi(x). \end{aligned}$$

The conditional Holevo mutual information provides an upper bound on the conditional mutual information as follows.

Proposition 1. *Let σ_θ^{XY} be the state of the composed system (X, Y). Suppose that a measurement is performed on X with the measurement outcome x and then another measurement is performed on Y with the measurement outcome y. Then,*

$$I_{\theta,\sigma|x}(\pi) \ge I_{\theta,y|x}(\pi). \tag{12}$$

Proof. Since any measurement is a trace-preserving completely positive map, inequality (12) follows from the monotonicity of the quantum relative entropy [13]. □

Analogous with the latent information priors [10] in classical statistics, we define latent information priors as priors that maximize the conditional Holevo mutual information. It is expected that the Bayesian predictive density operator $\sigma_{\hat{\pi},x}$ based on a latent information prior is a minimax predictive density operator. This is true from the following theorem, which is a quantum version of Theorem 2 of Komaki [10].

Theorem 2.

(1) *Let $\hat{\pi} \in \mathcal{P}$ be a prior maximizing $I_{\theta,\sigma|x}(\pi)$. If $p_{\hat{\pi}}(x) > 0$ for all $x \in \mathcal{X}$; then, $\sigma_{\hat{\pi}}(x)$ is a minimax predictive density operator.*

(2) *There exists a convergent prior sequence $\{\pi_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \sigma_{\pi_n}(x)$ is a minimax predictive density operator and the equality $I_{\theta,\sigma|x}(\pi_\infty) = \sup_{\pi\in\mathcal{P}} I_{\theta,\sigma|x}(\pi)$ holds.*

The proof of Theorems 1 and 2 are deferred to the Appendix A.

We note that the minimax risk $\inf_\rho \sup_\theta R_E(\theta, \rho)$ depends on the measurement E on X. Therefore, the measurement E with minimum minimax risk is desirable from the viewpoint of minimaxity. We define a POVM E^* to be a minimax POVM if it satisfies

$$\inf_{\rho} \sup_{\theta} R_{E^*}(\theta, \rho) = \inf_{E} \inf_{\rho} \sup_{\theta} R_E(\theta, \rho). \tag{13}$$

In the next section, we provide a class of minimax POVMs for one qubit system.

4. One Qubit System

In this section, we consider one qubit system and derive a class of minimax POVMs satisfying (13).

Qubit is a quantum system with a two-dimensional Hilbert space. It is the fundamental system in the quantum information theory. A general state of one qubit system is described by a density matrix

$$\sigma_\theta = \frac{1}{2} \begin{pmatrix} 1+\theta_z & \theta_x - i\theta_y \\ \theta_x + i\theta_y & 1-\theta_z \end{pmatrix},$$

where $\theta = (\theta_x, \theta_y, \theta_z)^\top \in \Theta = \{(\theta_x, \theta_y, \theta_z)^\top \in \mathbb{R}^3 \mid \|\theta\|^2 \leq 1\}$. The parameter space $\partial\Theta = \{(\theta_x, \theta_y, \theta_z)^\top \in \mathbb{R}^3 \mid \|\theta\|^2 = 1\}$ for pure states is called the Bloch sphere.

Let $\sigma_\theta^{XY} = \sigma_\theta \otimes \sigma_\theta$ be a separable state. We consider the estimation of $\sigma_\theta^Y = \sigma_\theta$ from the outcome of a measurement on $\sigma_\theta^X = \sigma_\theta$. Here, we assume that the state σ_θ^{XY} is separable, since the state of Y changes according to the outcome of the measurement on X and so the estimation problem is not well-defined if the state σ_θ^{XY} is not separable.

Let $\Omega := \{(x,y,z)^\top \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$ and $\mathcal{B} = \mathcal{B}(\Omega)$ be Borel sets. From Haapasalo et al. [20], it is sufficient to consider POVMs on Ω. For every probability measure μ on $(\Omega, \mathcal{B})$ that satisfies

$$\int_\Omega x d\mu(\omega) = \int_\Omega y d\mu(\omega) = \int_\Omega z d\mu(\omega) = 0,$$

we define a POVM $E : \mathcal{B} \to \mathcal{L}_+$ by

$$E(B) = \int_B \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} \mathrm{d}\mu(\omega).$$

In the following, we identify E with μ.

Let $\mathcal{E}^*_{\text{1-qubit}}$ be a class of POVMs on Ω represented by measures that satisfy the conditions

$$E_\mu[x] = E_\mu[y] = E_\mu[z] = 0,$$
$$E_\mu[xy] = E_\mu[yz] = E_\mu[zx] = 0,$$
$$E_\mu[x^2] = E_\mu[y^2] = E_\mu[z^2] = \frac{1}{3},$$

where E_μ is the expectation with respect to a measure μ. We provide two examples of POVMs in $\mathcal{E}^*_{\text{1-qubit}}$.

Proposition 2. *The POVM corresponding to*

$$\mu(\mathrm{d}\omega) = \frac{1}{4\pi}\mathrm{d}\omega, \tag{14}$$

where $d\omega$ *is surface element, is in* $\mathcal{E}^*_{\text{1-qubit}}$.

Proof. From the symmetry of μ, $E_\mu[x] = E_\mu[y] = E_\mu[z] = E_\mu[xy] = E_\mu[yz] = E_\mu[zx] = 0$. Moreover, from $E_\mu[1] = E_\mu[x^2 + y^2 + z^2] = 1$ and the symmetry of μ, $E_\mu[x^2] = E_\mu[y^2] = E_\mu[z^2] = 1/3$. □

Proposition 3. *Suppose that* ω_i $(i = 1, 2, 3, 4) \in \Omega$ *satisfies* $|\omega_i|^2 = 1$, $\omega_i \cdot \omega_j = -1/3 (i \neq j)$. *Let* μ *be a four point discrete measure on* Ω *defined by*

$$\mu(\{\omega_1\}) = \mu(\{\omega_2\}) = \mu(\{\omega_3\}) = \mu(\{\omega_4\}) = \frac{1}{4}. \tag{15}$$

Then, the POVM corresponding to μ *belongs to* $\mathcal{E}^*_{\text{1-qubit}}$.

Proof. Let $P = (\omega_1, \omega_2, \omega_3, \omega_4) \in \mathbb{R}^{3\times 4}$ and $\mathbf{1} = (1, 1, 1, 1)^\top$. From the assumption on ω_i $(i = 1, 2, 3, 4)$,

$$P^\top P = \frac{4}{3} I_4 - \frac{1}{3} J_4, \tag{16}$$

where $I_4 \in \mathbb{R}^{4\times 4}$ is the identity matrix and $J_4 = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{4\times 4}$ is a matrix whose elements are all one. From (16), we have $\mathbf{1}^\top P^\top P \mathbf{1} = \|P\mathbf{1}\|^2 = 0$. Therefore, $P\mathbf{1} = 0$ and it implies $E_\mu[x] = E_\mu[y] = E_\mu[z] = 0$.

In addition, from (16),

$$\begin{aligned} P^\top P P^\top P &= ((4/3)I_4 - (1/3)J_4)((4/3)I_4 - (1/3)J_4) \\ &= (4/3)((4/3)I_4 - (1/3)J_4) \\ &= (4/3)P^\top P. \end{aligned}$$

Therefore, $P^\top (PP^\top - (4/3)I_3)P = 0$. Since $\operatorname{rank} P = 3$, it implies $PP^\top = (4/3)I_3$. Then, $E_\mu[xy] = E_\mu[yz] = E_\mu[zx] = 0$ and $E_\mu[x^2] = E_\mu[y^2] = E_\mu[z^2] = 1/3$. □

We note that the POVM (15) is a special case of the SIC-POVM (symmetric, informationally complete, positive operator valued measure) [14,15].

Let $\mathcal{P}^*_{\text{1-qubit}}$ be a class of priors on Θ that satisfies the conditions

$$E_\pi[\theta_x] = E_\pi[\theta_y] = E_\pi[\theta_z] = 0,$$
$$E_\pi[\theta_x\theta_y] = E_\pi[\theta_y\theta_z] = E_\pi[\theta_z\theta_x] = 0,$$
$$E_\pi[\theta_x^2] = E_\pi[\theta_y^2] = E_\pi[\theta_z^2] = \frac{1}{3},$$

where E_π is the expectation with respect to a prior π.

Proposition 4. *The uniform prior*

$$\pi(\mathrm{d}\theta) = \frac{1}{4\pi}\mathrm{d}\theta,$$

where $\mathrm{d}\theta$ *is the surface element on the Bloch sphere, belongs to* $\mathcal{P}^*_{\text{1-qubit}}$.

Proof. Same as Proposition 2. □

Proposition 5. *Suppose that* θ_i $(i = 1, 2, 3, 4) \in \Theta$ *satisfies* $|\theta_i|^2 = 1,\ \theta_i \cdot \theta_j = -1/3 (i \neq j)$. *Then, the four point discrete prior*

$$\pi(\{\theta_1\}) = \pi(\{\theta_2\}) = \pi(\{\theta_3\}) = \pi(\{\theta_4\}) = \frac{1}{4}$$

belongs to $\mathcal{P}^*_{\text{1-qubit}}$.

Proof. Same as Proposition 3. □

We obtain the following result.

Lemma 4. *Suppose* $\pi^* \in \mathcal{P}^*_{\text{1-qubit}}$. *Then, for general measurement* E, *the risk function of the Bayesian predictive density operator* σ_{π^*} *is*

$$\begin{aligned} R_E(\theta, \sigma_{\pi^*}) = &- h\left(\frac{1 + \|\theta\|}{2}\right) + \frac{1}{2}\log\frac{9}{2} - \frac{\log 2}{2}({\theta_x}^2 E_\mu[x^2] + {\theta_y}^2 E_\mu[y^2] + {\theta_z}^2 E_\mu[z^2] \\ &+ 2\theta_x\theta_y E_\mu[xy] + 2\theta_y\theta_z E_\mu[yz] + 2\theta_z\theta_x E_\mu[zx]). \end{aligned}$$

Proof. The distribution of the measurement outcome $\omega = (x, y, z)^\top$ is

$$p(B \mid \theta) = \mathrm{Tr}\,\sigma_\theta E(B) = (1 + x\theta_x + y\theta_y + z\theta_z)\mu(B).$$

Then, since $\pi^* \in \mathcal{P}^*_{\text{1-qubit}}$, the marginal distribution of the measurement outcome is

$$p(B) = \int_\Theta p(B \mid \theta)\mathrm{d}\pi^*(\theta) = \int_\Theta (1 + x\theta_x + y\theta_y + z\theta_z)\mu(B)\mathrm{d}\pi^*(\theta) = \mu(B).$$

Therefore, the posterior distribution of θ is

$$\mathrm{d}\pi^*(\theta \mid \omega) = (1 + x\theta_x + y\theta_y + z\theta_z)\mathrm{d}\pi^*(\theta).$$

The posterior mean of θ_x, θ_y and θ_z are $x/3,\ y/3$ and $z/3$, respectively.
Thus, the Bayesian predictive density operator based on prior π^* is

$$\sigma_{\pi^*}(\omega) = \int \sigma_\theta \mathrm{d}\pi^*(\theta \mid \omega) = \frac{1}{2}\begin{pmatrix} 1 + z/3 & x/3 - iy/3 \\ x/3 + iy/3 & 1 - z/3 \end{pmatrix},$$

and we have

$$\log \sigma_{\pi^*}(\omega) = \begin{pmatrix} (\log\frac{1}{3})(\frac{1-z}{2}) & (\log\frac{1}{3})(\frac{-x+iy}{2}) \\ (\log\frac{1}{3})(\frac{-x-iy}{2}) & (\log\frac{1}{3})(\frac{1+z}{2}) \end{pmatrix} + \begin{pmatrix} (\log\frac{2}{3})(\frac{1+z}{2}) & (\log\frac{2}{3})(\frac{x-iy}{2}) \\ (\log\frac{2}{3})(\frac{x+iy}{2}) & (\log\frac{2}{3})(\frac{1-z}{2}) \end{pmatrix}.$$

Therefore, the quantum relative entropy loss is

$$\begin{aligned} D(\sigma_\theta, \sigma_{\pi^*}(\omega)) &= \operatorname{Tr} \sigma_\theta (\log \sigma_\theta - \log \sigma_{\pi^*}(\omega)) \\ &= -h\left(\frac{1+\|\theta\|}{2}\right) + \frac{1}{2} \log \frac{9}{2} - \frac{x\theta_x + y\theta_y + z\theta_z}{2} \log 2. \end{aligned}$$

Hence, the risk function is

$$\begin{aligned} R_E(\theta, \sigma_{\pi^*}) &= \int_\Omega D(\sigma_\theta, \sigma_{\pi^*}(\omega)) \mathrm{d}p(\omega \mid \theta) \\ &= -h\left(\frac{1+\|\theta\|}{2}\right) + \frac{1}{2} \log \frac{9}{2} - \frac{\log 2}{2} ({\theta_x}^2 E_\mu[x^2] + {\theta_y}^2 E_\mu[y^2] + {\theta_z}^2 E_\mu[z^2] \\ &\quad + 2\theta_x\theta_y E_\mu[xy] + 2\theta_y\theta_z E_\mu[yz] + 2\theta_z\theta_x E_\mu[zx]). \end{aligned}$$

□

Theorem 3. *For a measurement $E \in \mathcal{E}^*_{\text{1-qubit}}$, every $\pi^* \in \mathcal{P}^*_{\text{1-qubit}}$ is a latent information prior:*

$$\max_\theta R(\theta, \sigma_{\pi^*}) = \min_\rho \max_\theta R(\theta, \rho).$$

In addition, the risk of the Bayesian predictive density operator based on π^* is

$$R(\theta, \sigma_{\pi^*}) = -h\left(\frac{1+\|\theta\|}{2}\right) + \frac{1}{2} \log \frac{9}{2} - \frac{\log 2}{6} \|\theta\|^2,$$

where h is the binary entropy function $h(p) = -p \log p - (1-p) \log(1-p)$.

Proof. From Lemma 4 and $E^* \in \mathcal{E}^*_{\text{1-qubit}}$,

$$R_{E^*}(\theta, \sigma_{\pi^*}) = -h\left(\frac{1+|\theta|}{2}\right) + \frac{1}{2} \log \frac{9}{2} - \frac{\log 2}{6} ({\theta_x}^2 + {\theta_y}^2 + {\theta_z}^2).$$

Therefore, the risk depends only on $r = \|\theta\|$ and we have

$$R_{E^*}(\theta, \sigma_{\pi^*}) = g(r) = -h\left(\frac{1+r}{2}\right) + \frac{1}{2} \log \frac{9}{2} - \frac{\log 2}{6} r^2. \tag{17}$$

Since

$$g'(r) = \frac{1}{2} \log \left(\frac{1+r}{1-r}\right) - \frac{\log 2}{3} r,$$

$$g''(r) = \frac{1}{1-r^2} - \frac{\log 2}{3} \geq 1 - \frac{\log 2}{3} \geq 0,$$

the function $g(r)$ is convex. In addition, we have $g(1) = \log 3 - \frac{2}{3} \log 2 > g(0) = \log 3 - \frac{3}{2} \log 2$. Therefore, $g(r)$ takes the maximum at $r = 1$.

In other words, $R_{E^*}(\theta, \sigma_{\pi^*})$ takes maximum on the Bloch sphere. In addition, since $\int (\theta_x^2 + \theta_y^2 + \theta_z^2) \mathrm{d}\pi^*(\theta) = 1/3 + 1/3 + 1/3 = 1$, the support of π^* is included in the Bloch sphere $\|\theta\|^2 = 1$. Therefore, $\int R_{E^*}(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) = \sup_\theta R_{E^*}(\theta, \sigma_{\pi^*})$ and it implies that π^* is a latent information prior. □

We note that the Bayesian predictive density operator is identical for every $\pi^* \in \mathcal{P}^*_{\text{1-qubit}}$. In fact, every $\pi^* \in \mathcal{P}^*_{\text{1-qubit}}$ also provides the minimax estimation of density operator σ^Y_θ when there is no observation system X. Figure 1 shows the risk function $g(r)$ in (17) and also the minimax risk function $g_0(r)$ when there is no observation:

$$g_0(r) = \mathrm{Tr}\left(\left(\begin{pmatrix}(1+r)/2 & 0 \\ 0 & (1-r)/2\end{pmatrix}\right)\left(\log\begin{pmatrix}(1+r)/2 & 0 \\ 0 & (1-r)/2\end{pmatrix} - \log\begin{pmatrix}1/2 & 0 \\ 0 & 1/2\end{pmatrix}\right)\right)$$
$$= -h(r) + \log 2.$$

Whereas $g(r) < g_0(r)$ around $r = 1$, we can see that $g(r) > g_0(r)$ around $r = 0$. Both risk functions take the maximum at $r = 1$ and

$$g(1) = \log 3 - (2/3)\log 2 < g_0(1) = \log 2.$$

The decrease $g_0(1) - g(1) > 0$ in the maximum risk corresponds to the gain from the observation X.

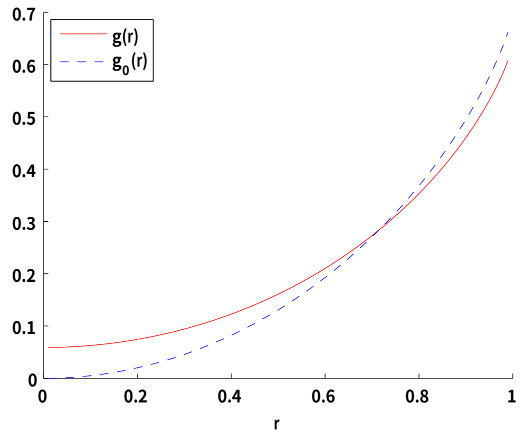

Figure 1. Risk functions of predictive density operators. solid line: $g(r)$, dashed line: $g_0(r)$.

Now, we consider the selection of the measurement E. As we discussed in the previous section, we define a POVM E^* to be a minimax POVM if it satisfies (13). We provide a sufficient condition on a POVM to be minimax. Let ρ^E be a minimax predictive density operator for the measurement E.

Lemma 5. *Suppose π^* is a latent information prior for the measurement E^*. If*

$$\int R_{E^*}(\theta, \rho^{E^*}) \mathrm{d}\pi^*(\theta) = \inf_E \int R_E(\theta, \rho^E) \mathrm{d}\pi^*(\theta),$$

then E^* is a minimax POVM.

Proof. For every (E, ρ), we have

$$\begin{aligned} \sup_\theta R_E(\theta,\rho) \geq \inf_\rho \sup_\theta R_E(\theta,\rho) &= \sup_\theta R_E(\theta, \rho^E) \\ &= \int R_E(\theta,\rho^E)\mathrm{d}\pi^*(\theta) \geq \inf_E \int R_E(\theta,\rho^E)\mathrm{d}\pi^*(\theta) \\ &= \int R_{E^*}(\theta,\rho^{E^*})\mathrm{d}\pi^*(\theta) = \sup_\theta R_{E^*}(\theta, \sigma_{\pi^*}). \end{aligned}$$

The last equality is from the minimaxity of σ_{π^*}. Therefore, E^* is a minimax POVM. □

Theorem 4. *Every $E^* \in \mathcal{E}^*_{\text{1-qubit}}$ is a minimax POVM.*

Proof. Let $\pi^* \in \mathcal{P}^*_{\text{1-qubit}}$. From Theorem 6, π^* is a latent information prior for E^*.

For general measurement E, from Lemma 4, the risk function of the Bayesian predictive density operator σ_{π^*} is

$$R_E(\theta, \sigma_{\pi^*}) = -h\left(\frac{1+\|\theta\|}{2}\right) + \frac{1}{2}\log\frac{9}{2} - \frac{\log 2}{2}({\theta_x}^2 E_\mu[x^2] + {\theta_y}^2 E_\mu[y^2] + {\theta_z}^2 E_\mu[z^2] + 2\theta_x\theta_y E_\mu[xy] + 2\theta_y\theta_z E_\mu[yz] + 2\theta_z\theta_x E_\mu[zx]).$$

Hence, the Bayes risk of σ_{π^*} with respect to π^* is

$$\int R_E(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) = \log 3 - \frac{2}{3}\log 2.$$

Now, since the Bayesian predictive density operator σ_{π^*} minimizes the Bayes risk with respect to π^* among all predictive density operators [4],

$$\int R_E(\theta, \rho^E) \mathrm{d}\pi^*(\theta) \geq \int R_E(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) = \log 3 - \frac{2}{3}\log 2$$

for every E. Therefore,

$$\inf_E \int R_E(\theta, \rho^E) \mathrm{d}\pi^*(\theta) \geq \log 3 - \frac{2}{3}\log 2.$$

On the other hand,

$$\inf_E \int R_E(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) \leq \int R_{E^*}(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) = \log 3 - \frac{2}{3}\log 2$$

is obvious.

Hence,

$$\int R_{E^*}(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) = \inf_E \int R_E(\theta, \sigma_{\pi^*}) \mathrm{d}\pi^*(\theta) = \log 3 - \frac{2}{3}\log 2.$$

From Lemma 5, E^* is minimax. □

Whereas Theorems 1 and 2 are valid even when σ_θ^{XY} is not separable, Theorems 3 and 4 assume the separability $\sigma_\theta^{XY} = \sigma_\theta^X \otimes \sigma_\theta^Y$.

From Theorem 4, the POVM (15) is a minimax POVM. Since this POVM is identical to the SIC-POVM [14,15], it is an interesting problem whether the SIC-POVM is a minimax POVM also in higher dimensions. This is a future work.

Acknowledgments: We thank the referees for many helpful comments. This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 26280005 and 14J09148.

Author Contributions: All authors contributed significantly to the study and approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Proof of (4). From the definition of $\hat{p}_\pi$ in (1),

$$\int p(x, y \mid \theta) \mathrm{d}\pi(\theta) = p_\pi(x)\hat{p}_\pi(y \mid x),$$

where

$$p_\pi(x) = \int p(x \mid \theta)\mathrm{d}\pi(\theta).$$

Therefore, for arbitrary $\hat{p}$,

$$\begin{aligned} R(\pi, \hat{p}) - R(\pi, \hat{p}_\pi) &= \int\int\int p(x, y \mid \theta)(\log \hat{p}_\pi(y \mid x) - \log \hat{p}(y \mid x))\mathrm{d}\pi(\theta)\mathrm{d}x\mathrm{d}y \\ &= \int\int p_\pi(x)\hat{p}_\pi(y \mid x)(\log \hat{p}_\pi(y \mid x) - \log \hat{p}(y \mid x))\mathrm{d}x\mathrm{d}y \\ &= \int\int p_\pi(x)L(\hat{p}_\pi(y \mid x), \hat{p}(y \mid x))\mathrm{d}x, \end{aligned}$$

which is nonnegative since the Kullback–Leibler divergence $L(q, p)$ in (3) is always nonnegative. □

Proof of (9). From the definition of $\sigma_\pi^Y(x)$ in (7),

$$\int p(x \mid \theta)\sigma_{\theta,x}^Y \mathrm{d}\pi(\theta) = p_\pi(x)\sigma_\pi^Y(x),$$

where

$$p_\pi(x) = \int p(x \mid \theta)\mathrm{d}\pi(\theta).$$

Therefore, for arbitrary $\hat{p}$,

$$\begin{aligned} R(\pi, \rho) - R(\pi, \sigma_\pi^Y) &= \int\int p(x \mid \theta) \operatorname{Tr} \sigma_{\theta,x}^Y(\log \sigma_\pi^Y(x) - \log \rho(x))\mathrm{d}\pi(\theta)\mathrm{d}x \\ &= \int p_\pi(x) \operatorname{Tr} \sigma_\pi^Y(x)(\log \sigma_\pi^Y(x) - \log \rho(x))\mathrm{d}x \\ &= \int p_\pi(x)L(\sigma_\pi^Y(x), \rho(x))\mathrm{d}x, \end{aligned}$$

which is nonnegative since the quantum relative entropy $L(\sigma, \rho)$ in (8) is always nonnegative. □

Proof of Theorem 1. (1) Let Q_x^ρ be the orthogonal projection matrix onto the eigenspace of $\rho(x)$ corresponding to eigenvalue 0, $\Theta^\rho = \{\theta \in \Theta \mid \sum_x p(x \mid \theta) \operatorname{Tr} Q_x^\rho \sigma_{\theta,x} = 0\}$ and $\mathcal{P}^\rho$ be the set of all probability measures on Θ^ρ.

If $\Theta^\rho = \emptyset$, the assertion is obvious because $R(\theta, \rho) = \infty$ for $\theta \notin \Theta^\rho$. Therefore, we assume $\Theta^\rho \neq \emptyset$ in the following. In this case, $D_\rho(\hat{\pi}^\rho) < \infty$. Since $\pi \in \mathcal{P}^\rho$ if and only if $D_\rho(\pi) < \infty$, we have $\hat{\pi}^\rho \in \mathcal{P}^\rho$.

Define

$$\tilde{\pi}_{\theta,u} := u\delta_\theta + (1-u)\hat{\pi}^\rho,$$

for $\theta \in \Theta^\rho$ and $0 \leq u \leq 1$, where δ_θ is the probability measure satisfying $\delta_\theta(\{\theta\}) = 1$. Then, $\tilde{\pi}_{\theta,u} \in \mathcal{P}^\rho$, and we have

$$
\begin{aligned}
&\frac{\partial}{\partial u} D_\rho(\tilde{\pi}_{\theta,u})\Big|_{u=0} \\
=& \frac{\partial}{\partial u} \sum_x \operatorname{Tr} S_{\tilde{\pi}_{\theta,u}}(x)\big(\log S_{\tilde{\pi}_{\theta,u}}(x) - \log(p_{\tilde{\pi}_{\theta,u}}(x)\rho(x))\big)\Big|_{u=0} \\
=& \frac{\partial}{\partial u} \sum_x \operatorname{Tr} \big(uS_\theta(x) + (1-u)S_{\tilde{\pi}_{\theta,u}}(x)\big) \\
& \times (\log(uS_\theta(x) + (1-u)S_{\hat{\pi}^\rho}(x)) - \log(up(x \mid \theta) + (1-u)p_{\hat{\pi}^\rho}(x))\rho(x)\Big|_{u=0} \\
=& \sum_x \operatorname{Tr} \left\{ \frac{\partial}{\partial u}(uS_\theta(x) + (1-u)S_{\hat{\pi}^\rho}(x))\Big|_{u=0} \right\} (\log S_{\hat{\pi}^\rho}(x) - \log(p_{\hat{\pi}^\rho}(x)\rho_x)) \\
& + \sum_x \operatorname{Tr} S_{\hat{\pi}^\rho}(x) \left\{ \frac{\partial}{\partial u} \log(uS_\theta(x) + (1-u)S_{\hat{\pi}^\rho}(x))\Big|_{u=0} \right\} \\
& - \sum_x \operatorname{Tr} S_{\hat{\pi}^\rho}(x) \left\{ \frac{\partial}{\partial u} (\log(up(x \mid \theta) + (1-u)p_{\hat{\pi}^\rho}(x))I + \log \rho_x)\Big|_{u=0} \right\} \\
=& \sum_x \operatorname{Tr} \big(S_\theta(x) - S_{\hat{\pi}^\rho}(x)\big)\big(\log S_{\hat{\pi}^\rho}(x) - \log(p_{\hat{\pi}^\rho}(x)\rho(x))\big) \\
& + \sum_x \operatorname{Tr} \big(S_\theta(x) - p_{\hat{\pi}^\rho}(x)\rho(x)\big) - \sum_x \operatorname{Tr} \left(S_{\hat{\pi}^\rho}(x) \frac{p(x \mid \theta) - p_{\hat{\pi}^\rho}(x)}{p_{\hat{\pi}^\rho}(x)}\right) \\
=& \sum_x \operatorname{Tr} S_\theta(x)\big(\log S_{\hat{\pi}^\rho}(x) - \log(p_{\hat{\pi}^\rho}(x)\rho(x))\big) \\
& - \sum_x \operatorname{Tr} S_{\hat{\pi}^\rho}(x)\big(\log S_{\hat{\pi}^\rho}(x) - \log(p_{\hat{\pi}^\rho}(x)\rho(x))\big) \geq 0.
\end{aligned}
$$

Thus, if $\theta \in \Theta^\rho$,

$$
\begin{aligned}
R(\theta, \sigma_{\hat{\pi}^\rho}(x)) =& \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\hat{\pi}^\rho}(x)) \\
\leq& \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) = R(\theta, \rho(x)) < \infty.
\end{aligned}
$$

If $\theta \notin \Theta^\rho$, $R(\theta, \rho(x)) = \infty$. Therefore, for every $\theta \in \Theta$, the inequality $R(\theta, \sigma_{\hat{\pi}^\rho}(x)) \leq R(\theta, \rho(x))$ holds.

(2) We note that Θ^ρ and $\mathcal{P}^\rho$ are compact subsets of Θ and $\mathcal{P}$, respectively.

If $\Theta^\rho = \emptyset$, the assertion is obvious, because $R(\theta, \rho_x) = \infty$ for every $\theta \notin \Theta^\rho$. Therefore, we assume $\Theta^\rho \neq \emptyset$ in the following. Let $\mathcal{X}^\rho := \{x \in \mathcal{X} \mid \exists \theta \in \Theta^\rho, p(x \mid \theta) > 0\}$ and μ^ρ be a probability measure on Θ^ρ such that $p_{\mu^\rho}(x) := \int p(x \mid \theta) d\mu^\rho(\theta) > 0$ for every $x \in \mathcal{X}^\rho$.

Because $D_\rho(\pi)$ is continuous as a function of $\pi \in \mathcal{P}^\rho$, there exists $\pi_n \in \mathcal{P}^\rho_{\mu^\rho/n} := \{(1/n)\mu^\rho + (1-1/n)\pi \mid \pi \in \mathcal{P}^\rho\}$ such that $D_\rho(\pi_n) = \inf_{\pi \in \mathcal{P}^\rho_{\mu/n}} D_\rho(\pi)$. From Lemma 3, there exists a convergent subsequence $\{\pi'_m\}_{m=1}^\infty$ of $\{\pi_n\}_{n=1}^\infty$ such that $D_\rho(\pi'_\infty) = \inf_{\pi \in \mathcal{P}^\rho} D_\rho(\pi)$, where $\lim \pi'_m \Rightarrow \pi'_\infty$.

Let n_m be the integer satisfying $\pi'_m = \pi_{n_m}$. We can make the subsequence $\{\pi'_m\}_{m=1}^\infty$ satisfy $0 < n_m/(n_{m+1} - n_m) < c$ for some positive constant c.

Since

$$
\frac{n_m}{n_{m+1}} \pi'_m + \left(1 - \frac{n_m}{n_{m+1}}\right)\delta_\theta = \frac{n_m}{n_{m+1}} \pi_{n_m} + \left(1 - \frac{n_m}{n_{m+1}}\right)\delta_\theta \in \mathcal{P}^\rho_{\mu^\rho/n_{m+1}}
$$

for every $\theta \in \Theta$, we have

$$
\tilde{\pi}_{m,\theta,u} := u\left\{\frac{n_m}{n_{m+1}} \pi'_m + \left(1 - \frac{n_m}{n_{m+1}}\right)\delta_\theta\right\} + (1-u)\pi'_{m+1} \in \mathcal{P}^\rho_{\mu^\rho/n_{m+1}}
$$

for every $\theta \in \Theta^\rho$ and $0 \le u \le 1$. Thus,

$$
\begin{aligned}
&\frac{\partial}{\partial u} D(\tilde{\pi}_{m,\theta,u})\Big|_{u=0} \\
&= \frac{\partial}{\partial u} \sum_x \operatorname{Tr} p_{\tilde{\pi}_{m,\theta,u}}(x)\big(\log S_{\tilde{\pi}_{m,\theta,u}}(x) - \log(p_{\tilde{\pi}_{m,\theta,u}}(x)\rho(x))\big)(I - Q_x^\rho)\Big|_{u=0} \\
&= \sum_x \operatorname{Tr} \left\{ \frac{\partial}{\partial u} S_{\tilde{\pi}_{m,\theta,u}}(x)\Big|_{u=0} \right\} \big(\log S_{\tilde{\pi}_{m,\theta,u}}(x) - \log(p_{\tilde{\pi}_{m,\theta,u}}(x)\rho(x))\big)(I - Q_x^\rho) \\
&= \frac{n_m}{n_{m+1}} \sum_x \operatorname{Tr} S_{\pi'_m}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
&\quad - \sum_x \operatorname{Tr} S_{\pi'_{m+1}}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
&\quad + \frac{n_{m+1} - n_m}{n_{m+1}} \sum_x \operatorname{Tr} S_\theta(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(X)\rho(x))\big)(I - Q_x^\rho) \\
&\ge 0.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\textstyle\sum_x \operatorname{Tr} S_\theta(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
\ge& \frac{n_{m+1}}{n_{m+1} - n_m} \textstyle\sum_x \operatorname{Tr} S_{\pi'_{m+1}}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
-& \frac{n_m}{n_{m+1} - n_m} \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
=& \frac{n_{m+1}}{n_{m+1} - n_m} \textstyle\sum_x \operatorname{Tr} S_{\pi'_{m+1}}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
+& \frac{n_m}{n_{m+1} - n_m} \Big\{ - \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho)(I - Q_x^{\pi'_\infty}) \\
-& \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)Q_x^{\pi'_\infty}(I - Q_x^\rho) \Big\} \\
\ge& \frac{n_{m+1}}{n_{m+1} - n_m} \textstyle\sum_x \operatorname{Tr} S_{\pi'_{m+1}}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho) \\
+& \frac{n_m}{n_{m+1} - n_m} \Big\{ - \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}(x)\rho(x))\big)(I - Q_x^\rho)(I - Q_x^{\pi'_\infty}) \\
+& \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x) \log \rho(x) Q_x^{\pi'_\infty}(I - Q_x^\rho) \Big\},
\end{aligned}
\tag{A1}
$$

where $Q_x^{\pi'_\infty}$ is the orthogonal projection matrix onto the eigenspace of $\sum_\theta \pi'_\infty(\theta) p(x \mid \theta)\sigma_{\theta,x}$ corresponding to the eigenvalue 0. Here, we have

$$
\begin{aligned}
&\lim_{m\to\infty} \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x)\big(\log S_{\pi'_{m+1}}(x) - \log(p_{\pi'_{m+1}}\rho(x))\big)(I - Q_x^\rho)(I - Q_x^{\pi'_\infty}) \\
&\quad = \textstyle\sum_x \operatorname{Tr} S_{\pi'_\infty}(x)\big(\log S_{\pi'_\infty}(x) - \log(p_{\pi'_\infty}(x)\rho(x))\big)(I - Q_x^\rho)(I - Q_x^{\pi'_\infty}),
\end{aligned}
\tag{A2}
$$

and

$$
\begin{aligned}
&\lim_{m\to\infty} \textstyle\sum_x \operatorname{Tr} S_{\pi'_m}(x) \log \rho(x) Q_x^{\pi'_\infty}(I - Q_x^\rho) = 0 \\
&\quad = -\textstyle\sum_x \operatorname{Tr} S_{\pi'_\infty}(x)\big(\log S_{\pi'_\infty}(x) - \log(p_{\pi'_\infty}(x)\rho(x))\big)Q_x^{\pi'_\infty}(I - Q_x^\rho).
\end{aligned}
\tag{A3}
$$

Therefore, from (A1)–(A3) and $0 < n_m/(n_{m+1} - n_m) < c$ for every $\theta \in \Theta^\rho$,

$$
\begin{aligned}
&\liminf_{m\to\infty} \textstyle\sum_x \operatorname{Tr} S_\theta(x)\big(\log S_{\pi'_m}(x) - \log(p_{\pi'_m}(x)\rho(x))\big)(I - Q_x^\rho) \\
&\quad \ge \textstyle\sum_x \operatorname{Tr} S_{\pi'_\infty}(x)\big(\log S_{\pi'_\infty}(x) - \log(p_{\pi'_\infty}(x)\rho(x))\big)(I - Q_x^\rho) \ge 0.
\end{aligned}
\tag{A4}
$$

By taking an appropriate subsequence $\{\pi''_k\}$ of $\{\pi'_m\}$, we can make the subsequence of density operators $\{\sigma_{\pi''_k,x}\}_{k=1}^{\infty}$ converge for all $x \in \mathcal{X}^{\rho}$ because $p_{\pi'_m}(x) > 0$ $(x \in \mathcal{X}^{\rho})$ and $0 \leq S_{\pi'_m}/p_{\pi'_m}(x) \leq I$.

Then, from (A4), if $\theta \in \Theta^{\rho}$,

$$
\begin{aligned}
R(\theta, \lim_{k\to\infty} \sigma_{\pi''_k}(x)) &= \sum_x \mathrm{Tr}\, S_\theta(x)(\log \sigma_{\theta,x} - \log \lim_{k\to\infty} \sigma_{\pi''_k}(x)) \\
&= \sum_x \mathrm{Tr}\, S_\theta(x)(\log \sigma_{\theta,x} - \log \lim_{k\to\infty} \sigma_{\pi''_k}(x))(I - Q^{\rho}_x) \\
&\leq \sum_x \mathrm{Tr}\, S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x))(I - Q^{\rho}_x) \\
&= \sum_x \mathrm{Tr}\, S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) = R(\theta, \rho(x)) < \infty.
\end{aligned}
$$

If $\theta \notin \Theta^{\rho}$, $R(\theta, \rho) = \infty$ because $-\sum_x S_\theta(x) \log \rho(x) Q^{\rho}_x = \infty$.

Hence, the risk of the predictive density operator defined by

$$
\begin{cases}
\lim_{k\to\infty} \sigma_{\pi''_k}(x) \quad , & x \in \mathcal{X}^{\rho}, \\
\tau_x \quad , & x \notin \mathcal{X}^{\rho},
\end{cases}
$$

where τ_x is an arbitrary predictive density, is not greater than that of $\rho(x)$ for every $\theta \in \Theta$.

Therefore, by taking a sequence $\{\varepsilon_n \in (0,1)\}_{n=1}^{\infty}$ that converges rapidly enough to 0, we can construct a predictive density operator

$$
\lim_{k\to\infty} \sigma_{\varepsilon_k \bar{\mu} + (1-\varepsilon_k)\pi''_k}(x) = \begin{cases}
\lim_{k\to\infty} \sigma_{\pi''_k}(x), & x \in \mathcal{X}^{\rho}, \\
\sigma_{\bar{\mu}}(x), & x \notin \mathcal{X}^{\rho},
\end{cases} \tag{A5}
$$

as a limit of Bayesian predictive density operators based on priors $\{\varepsilon_k \bar{\mu} + (1-\varepsilon_k)\pi''_k\}$, where $\bar{\mu}$ is a measure on Θ such that $p_{\bar{\mu}}(x) > 0$ for every $x \in \mathcal{X}$.

Hence, the risk of the predictive density operator (A5) is not greater than that of $\rho(x)$ for every $\theta \in \Theta$. $\square$

Proof of Theorem 2. (1) Define $\tilde{\pi}_{\bar{\theta},u} := u\delta_{\bar{\theta}} + (1-u)\hat{\pi}$ for all $\theta \in \Theta$ and $u \in [0,1]$. Then,

$$
\begin{aligned}
&\frac{\partial}{\partial u} I_{\theta,\sigma|x}(\tilde{\pi}_{\bar{\theta},u}) \Big|_{u=0} \\
&= \frac{\partial}{\partial u} \Big(\int \sum_x \mathrm{Tr}\, S_\theta(x) \log S_\theta(x) d\tilde{\pi}_{\bar{\theta},u}(\theta) - \sum_x S_{\tilde{\pi}_{\bar{\theta},u}}(x) \log S_{\tilde{\pi}_{\bar{\theta},u}}(x) \\
&- \int \sum_x p(x \mid \theta) \log p(x \mid \theta) d\tilde{\pi}_{\bar{\theta},u} + \sum_x p_{\tilde{\pi}_{\bar{\theta},u}}(x) \log p_{\tilde{\pi}_{\bar{\theta},u}}(x) \Big) \Big|_{u=0} \\
&= \sum_x \mathrm{Tr}\, S_{\bar{\theta}}(x)(\log S_{\bar{\theta}}(x) - \log p_{\bar{\theta}}(x) I) - \sum_x \mathrm{Tr}\, S_{\bar{\theta}}(x)(\log S_{\hat{\pi}}(x) - \log p_{\hat{\pi}}(x) I) \\
&- \int \sum_x \mathrm{Tr}\, S_\theta(x)(\log S_\theta(x) - \log p(x \mid \theta) I) d\hat{\pi}(\theta) \\
&+ \sum_x \mathrm{Tr}\, S_{\hat{\pi}}(x)(\log S_{\hat{\pi}}(x) - \log p_{\hat{\pi}}(x) I) \leq 0.
\end{aligned}
$$

Since $p_{\hat{\pi}}(x) > 0$ for every $x \in \mathcal{X}$ and $\mathrm{Tr}\, p(x \mid \theta)\sigma_{\theta,x} \log \sigma_{\theta,x} = 0$ if $p(x \mid \theta) = 0$, we have

$$
\sum_x \mathrm{Tr}\, S_{\bar{\theta}}(x)(\log \sigma_{\bar{\theta},x} - \log \sigma_{\hat{\pi}}(x)) \leq \int \sum_x \mathrm{Tr}\, S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\hat{\pi}}(x)) d\hat{\pi}(\theta) \tag{A6}
$$

for every $\theta \in \Theta$.

On the other hand, we have

$$\begin{aligned}
&\int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\hat{\pi}}(x)) d\hat{\pi}(\theta) \\
&= \inf_\rho \int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) d\hat{\pi}(\theta) \\
&\leq \sup_{\pi \in \mathcal{P}} \inf_\rho \int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) d\hat{\pi}(\theta) \\
&\leq \inf_\rho \sup_{\pi \in \mathcal{P}} \int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) d\hat{\pi}(\theta) \\
&= \inf_\rho \sup_{\theta \in \Theta} \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) \\
&\leq \sup_{\theta \in \Theta} \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\hat{\pi}}(x)).
\end{aligned} \tag{A7}$$

Here, the first equality is from the fact [4] that the Bayes risk with respect to $\hat{\pi} \in \mathcal{P}$

$$\int R(\theta; \rho(x)) d\hat{\pi}(\theta) = \int \sum_x p(x \mid \theta) \operatorname{Tr} \sigma_{\theta,x}(\log \sigma_{\theta,x} - \log \rho(x)) d\hat{\pi}(\theta)$$

is minimized when

$$\rho(x) = \sigma_{\hat{\pi}}(x) := \frac{\int p(x \mid \theta) \sigma_{\theta,x} d\hat{\pi}(\theta)}{\int p(x \mid \theta) d\hat{\pi}(\theta)}.$$

From (A6) and (A7), we have

$$\inf_\rho \sup_{\theta \in \Theta} \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \rho(x)) = \sup_{\theta \in \Theta} \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\hat{\pi}}(x)).$$

Therefore, the predictive density operator $\sigma_{\hat{\pi}}(x)$ is minimax.

(2) Let μ be a probability measure on Θ such that $p_\mu(x) := \int p(x \mid \theta) d\mu(\theta) > 0$ for every $x \in \mathcal{X}$, and let $\pi_n \in \mathcal{P}_{\mu/n} := \{\mu/n + (1 - 1/n)\pi \mid \pi \in \mathcal{P}\}$ be a prior satisfying $I_{\theta,\sigma|x}(\pi_n) = \sup_{\pi \in \mathcal{P}_{\mu/n}} I_{\theta,\sigma|x}(\pi)$. From Lemma 3, there exists a convergent subsequence $\{\pi'_m\}$ of $\{\pi_n\}$ and $I_{\theta,\sigma|x}(\pi'_\infty) = \sup_{\pi \in \mathcal{P}} I_{\theta,\sigma|x}(\pi)$ where $\pi'_m \Rightarrow \pi'_\infty$. Let n_m be the integer satisfying $\pi'_m = \pi_{n_m}$. As in the proof of Theorem 1, we can make the subsequence $\{\pi'_m\}$ satisfy $0 < n_m/(n_{m+1} - n_m) < c$ for some positive constant c.

Then, for every $\bar{\theta} \in \Theta$,

$$\tilde{\pi}_{m,\bar{\theta},u} := u\Big\{\frac{n_m}{n_{m+1}}\pi'_m + (1 - \frac{n_m}{n_{m+1}})\delta_{\bar{\theta}}\Big\} + (1-u)\pi'_{m+1}$$

belongs to $\mathcal{P}_{\mu/n_{m+1}}$ for $0 \leq u \leq 1$ because $(n_m/n_{m+1})\pi'_m + (1 - n_m/n_{m+1})\delta_{\bar{\theta}} \in \mathcal{P}_{\mu/n_{m+1}}$ and $\pi'_{m+1} \in \mathcal{P}_{\mu/n_{m+1}}$.

Thus,

$$
\begin{aligned}
&\frac{\partial}{\partial u} I_{\theta,\rho|x}(\tilde{\pi}_{m,\bar{\theta},u})\Big|_{u=0} \\
&= \frac{\partial}{\partial u}\Big(\int \sum_x \operatorname{Tr} S_\theta(x) \log S_\theta(x) d\tilde{\pi}_{m,\bar{\theta},u}(\theta) - \sum_x \operatorname{Tr} S_{\tilde{\pi}_{m,\bar{\theta},u}}(x) \log S_{\tilde{\pi}_{m,\bar{\theta},u}}(x) \\
&- \int \sum_x p(x \mid \theta) \log p(x \mid \theta) d\tilde{\pi}_{m,\bar{\theta},u}(\theta) + \sum_x p_{\tilde{\pi}_{m,\bar{\theta},u}}(x) \log p_{\tilde{\pi}_{m,\bar{\theta},u}}\Big)\Big|_{u=0} \\
&= \frac{n_m}{n_{m+1}} \int \sum_x \operatorname{Tr} S_\theta(x) \log S_\theta(x) d\pi'_m(\theta) + \Big(1 - \frac{n_m}{n_{m+1}}\Big) \sum_x \operatorname{Tr} S_{\bar{\theta}}(x) \log S_{\bar{\theta}}(x) \\
&- \int \sum_x \operatorname{Tr} S_\theta(x) \log S_\theta(x) d\pi'_{m+1}(\theta) - \sum_x \operatorname{Tr} \frac{\partial}{\partial u} S_{\tilde{\pi}_{m,\bar{\theta},u}}(x)\Big|_{u=0} \log S_{\pi'_{m+1}}(x) \\
&- \frac{n_m}{n_{m+1}} \int \sum_x p(x \mid \theta) \log p(x \mid \theta) d\pi'_{m+1}(\theta) - \Big(1 - \frac{n_m}{n_{m+1}}\Big) \sum_x p_{\bar{\theta}}(x) \log p_{\bar{\theta}}(x) \\
&+ \int \sum_x p(x \mid \theta) \log p(x \mid \theta) d\pi'_{m+1}(\theta) + \sum_x \frac{\partial}{\partial u} p_{\tilde{\pi}_{m,\bar{\theta},u}}(x)\Big|_{u=0} \log p_{\pi'_{m+1}}(x) \\
&= \Big(1 - \frac{n_m}{n_{m+1}}\Big) \sum_x \operatorname{Tr} S_{\bar{\theta}}(x)(\log S_{\bar{\theta}}(x) - \log p(x \mid \bar{\theta}) I) \\
&- \Big(1 - \frac{n_m}{n_{m+1}}\Big) \sum_x \operatorname{Tr} S_{\bar{\theta}}(x)(\log S_{\pi'_{m+1}}(x) - \log p_{\pi'_{m+1}}(x) I) \\
&+ \frac{n_m}{n_{m+1}} \int \sum_x \operatorname{Tr} S_\theta(x)(\log S_\theta(x) - \log p(x \mid \theta) I) d\pi'_m(\theta) \\
&- \int \sum_x \operatorname{Tr} S_\theta(x)(\log S_\theta(x) - \log p(x \mid \theta)) d\pi'_{m+1}(\theta) \\
&- \frac{n_m}{n_{m+1}} \sum_x \operatorname{Tr} S_{\pi'_m}(x)(\log S_{\pi'_{m+1}}(x) - \log p_{\pi'_{m+1}}(x) I) \\
&+ \sum_x \operatorname{Tr} S_{\pi'_{m+1}}(x)(\log S_{\pi'_{m+1}}(x) - \log p_{\pi'_{m+1}}(x) I) \\
&\leq 0.
\end{aligned}
$$

Since $p_{\hat{\pi}_m}(x) > 0$ for every m and $p(x \mid \theta)\sigma_{\theta,x} \log \sigma_{\theta,x} = 0$ if $p(x \mid \theta) = 0$, we have

$$
\begin{aligned}
&\left(1 - \frac{n_m}{n_{m+1}}\right) \sum_x \operatorname{Tr} S_{\bar{\theta}}(x)(\log \sigma_{\bar{\theta},x} - \log \sigma_{\pi'_{m+1}}(x)) \\
&+ \frac{n_m}{n_{m+1}} \int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x)) d\pi'_m(\theta) \\
&- \int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x)) d\pi'_{m+1}(\theta) \leq 0.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\textstyle\sum_x \operatorname{Tr} S_{\bar\theta}(x)(\log \sigma_{\bar\theta}(x) - \log \sigma_{\pi'_{m+1}}(x)) \\
\le & -\frac{n_m}{n_{m+1}-n_m}\Big\{ \textstyle\int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x))(1-Q_x^{\pi'_\infty}) d\pi'_m(\theta) \\
&+ \textstyle\int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x)) Q_x^{\pi'_\infty} d\pi'_m(\theta) \Big\} \\
&+ \frac{n_{m+1}}{n_{m+1}-n_m} \textstyle\int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x)) d\pi'_{m+1}(\theta) \\
\le & -\frac{n_m}{n_{m+1}-n_m}\Big\{ \textstyle\int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x))(1-Q_x^{\pi'_\infty}) d\pi'_m(\theta) \\
&+ \textstyle\int \sum_x \operatorname{Tr} S_\theta(x) \log \sigma_{\theta,x} Q_x^{\pi'_\infty} d\pi'_m(\theta) \Big\} \\
&+ \frac{n_{m+1}}{n_{m+1}-n_m} \textstyle\int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x)) d\pi'_{m+1}(\theta),
\end{aligned}
\tag{A8}
$$

where $Q_x^{\pi'_\infty}$ is the orthogonal projection matrix onto the eigenspace of $S_{\pi'_\infty}(x)$ corresponding to the eigenvalue 0. Here, we used two equalities

$$
\begin{aligned}
&\textstyle\lim_{m\to\infty} \int \sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \sigma_{\pi'_{m+1}}(x))(1-Q_x^{\pi'_\infty}) d\pi'_m(\theta) \\
= & \textstyle\int \sum_x \operatorname{Tr} S_\theta(x)(\log(p_{\pi'_\infty}(x)\sigma_{\theta,x}) - \log S_{\pi'_\infty}(x)) d\pi'_\infty(\theta)
\end{aligned}
\tag{A9}
$$

and

$$
\begin{aligned}
&\lim_{m\to\infty} \int \sum_x \operatorname{Tr} S_\theta(x) \log \sigma_{\theta,x} Q_x^{\pi'_\infty} d\pi'_m(\theta) \\
= & \int \sum_x \operatorname{Tr} S_\theta(x) \log \sigma_{\theta,x} Q_x^{\pi'_\infty} d\pi'_\infty(\theta) \\
= & \int \sum_x \operatorname{Tr} S_\theta(x)(\log(p_{\pi'_\infty}(x))\sigma_{\theta,x}) - \log S_{\pi'_\infty,x}) Q_x^{\pi'_\infty} d\pi'_\infty(\theta) = 0,
\end{aligned}
\tag{A10}
$$

since $\operatorname{Tr} S_\theta(x) \log \sigma_{\theta,x}$ is a bounded continuous function of θ.

From (A8)–(A11), and $0 < n_m/(n_{m+1}-n_m) < c$, we have, for every $\bar\theta \in \Theta$,

$$
\begin{aligned}
&\limsup_{m\to\infty} \sum_x \operatorname{Tr} S_{\bar\theta}(x)(\log \sigma_{\bar\theta}(x) - \log \sigma_{\pi'_m}(x)) \\
\le & \int \sum_x \operatorname{Tr} S_\theta(x)(\log(p_{\pi'_\infty}(x)\sigma_\theta(x)) - \log S_{\pi'_\infty}(x)) d\pi'_\infty(\theta).
\end{aligned}
$$

By taking an appropriate subsequence $\{\pi''_k\}$ of $\{\pi'_m\}$, we can make $\{\sigma_{\pi''_k}(x)\}_{k=1}^\infty$ converge for every x. Then, for every $\bar\theta \in \Theta$,

$$
\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\bar\theta,x} - \log \lim_{k\to\infty} \sigma_{\pi''_k}(x)) \tag{A11}
$$

$$
\le \int \sum_x S_\theta(x)(\log(\sigma_{\theta,x} - \log \lim_{k\to\infty} \sigma_{\pi''_k}(x)) d\pi''_\infty(\theta), \tag{A12}
$$

since $\lim_{k\to\infty} \sigma_{\pi''_k}(x) = \sigma_{\pi''_\infty}(x)$ for x with $p_{\pi''_\infty}(x) > 0$.

On the other hand, we have

$$\begin{aligned}
&\int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \lim_{k\to\infty}\sigma_{\pi_k''}(x))d\pi_\infty''(\theta) \\
=\ &\inf_\rho \int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x))d\pi_\infty''(\theta) \\
\le\ &\sup_{\pi\in\mathcal{P}} \inf_\rho \int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x))d\pi(\theta) \\
\le\ &\inf_\rho \sup_{\pi\in\mathcal{P}} \int \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x))d\pi(\theta) \\
=\ &\inf_\rho \sup_{\theta\in\Theta} \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \rho(x)) \\
\le\ &\sup_{\theta\in\Theta} \textstyle\sum_x \operatorname{Tr} S_\theta(x)(\log \sigma_{\theta,x} - \log \lim_{k\to\infty} \sigma_{\pi_k''}(x)).
\end{aligned} \tag{A13}$$

Here, the first equality is from the fact [4] that the Bayes risk

$$\int R(\theta;\rho)d\pi_\infty''(\theta) = \int \sum_x \operatorname{Tr} p(x \mid \theta)\sigma_{\theta,x}(\log \sigma_{\theta,x} - \log \rho(x))d\pi_\infty''(\theta)$$

is minimized when $\rho(x) = \sigma_{\pi_\infty''}(x)$. Although $p_{\pi_\infty''}(x)$ is not uniquely determined for x with $p_{\pi_\infty''}(x) = 0$, the Bayes risk does not depend on the choice of $\sigma_{\pi_\infty''}(x)$ for such x.

From (A12) and (A13),

$$\begin{aligned}
&\inf_\rho \sup_{\theta\in\Theta} \sum_x \operatorname{Tr} p(x \mid \theta)\sigma_{\theta,x}(\log \sigma_{\theta,x} - \rho(x)) \\
=&\sup_{\theta\in\Theta} \sum_x \operatorname{Tr} p(x \mid \theta)\sigma_{\theta,x}(\log \sigma_{\theta,x} - \log \lim_{k\to\infty} \sigma_{\pi_k''}(x)).
\end{aligned}$$

Therefore, the predictive density operator $\lim_{k\to\infty} \sigma_{\pi_k''}(x)$ is minimax. □

References

1. Barndorff-Nielsen, O.E.; Gill, R.D.; Jupp, P.E. On quantum statistical inference. *J. R. Stat. Soc. B* **2003**, *65*, 775–804.
2. Holevo, A.S. *Probabilistic and Statistical Aspects of Quantum Theory*; Elsevier: Amsterdam, The Netherlands, 1982.
3. Paris, M.; Rehacek, J. *Quantum State Estimation*; Springer: Berlin, Germany, 2004.
4. Tanaka, F.; Komaki, F. Bayesian predictive density operators for exchangeable quantum-statistical models. *Phys. Rev. A* **2005**, *71*, 052323.
5. Tanaka, F. Bayesian estimation of the wave function. *Phys. Lett. A* **2012**, *376*, 2471–2476.
6. Tanaka, F. Noninformative prior in the quantum statistical model of pure states. *Phys. Rev. A* **2012**, *85*, 062305.
7. Geisser, S. *Predictive Inference: An Introduction*; Chapman & Hall: London, UK, 1993.
8. Csiszar, I. Axiomatic characterizations of information measures. *Entropy* **2008**, *10*, 261–273.
9. Aitchison, J. Goodness of prediction fit. *Biometrika* **1975**, *62*, 547–554.
10. Komaki, F. Bayesian predictive densities based on latent information priors. *J. Stat. Plan. Inference* **2011**, *141*, 3705–3715.
11. Bernardo, J.M. Reference posterior distributions for Bayesian inference. *J. R. Stat. Soc. B* **1979**, *41*, 113–147.
12. Petz, D. *Quantum Information and Quantum Statistics*; Springer: New York, NY, USA, 2008.
13. Holevo, A.S. *Quantum Systems, Channels, Information: A Mathematical Introduction*; Walter de Gruyter: Berlin, Germany, 2013.
14. Appleby, D.M. SIC-POVMs and the extended Clifford group. *J. Math. Phys.* **2004**, *46*, 547–554.
15. Renes, J.M.; Blume-Kohout, R.; Scott, A.J.; Caves, C.M. Symmetric informationally complete quantum measurements. *J. Math. Phys.* **2004**, *45*, 2171–2180.
16. Ferrie, C.; Blume-Kohout, R. Minimax quantum tomography: Estimators and relative entropy bounds. *Phys. Rev. Lett.* **2016**, *116*, 090407.
17. Tanaka, F. Quantum minimax theorem. *arXiv* **2014**, arXiv1410:3639.

18. Billingsley, P. *Convergence of Probability Measures*; Wiley: New York, NY, USA, 1999.
19. Hiai, F.; Petz, D. *Introduction to Matrix Analysis and Applications*; Springer: New York, NY, USA, 2014.
20. Haapasalo, E.; Heinosaari, T.; Pellonpää, J.P. Quantum measurements on finite dimensional systems: Relabeling and mixing. *Quantum Inf. Process.* **2012**, *11*, 1751–1763.

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).