IEEE Transactions on Information Theory

New Approach To Quantum Statistical Inference

Journal:	*IEEE Transactions on Information Theory*
Manuscript ID	IT-19-0550
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	06-Aug-2019
Complete List of Authors:	Shafiei Deh Abad, Ahmad; University of Tehran, School of mathematics , Statistics and Computer science Shahbazi, Mohammd; University of Tehran, School of mathematics, Statistics and Computer science
Keywords:	Quantum computing, Statistical learning, Quantum mechanics

New Approach To Quantum Statistical Inference

Ahmad Shafiei Deh Abad† and Mohammad Shahbazi⋆

†shafiei@khayam.ut.ac.ir , ⋆mshahbazi@ut.ac.ir
⋆†School of Mathematics, Statistics and Computer Science,
University of Tehran, Tehran, Iran

August 7, 2019

Abstract

This paper extends the technique of Minimum Description Length (MDL) principle introduced in classical statistics [26, 27] to the quantum regimes. Classical MDL is a kind of penalized maximum likelihood. When the observed data is limited or there is a good chance of overfitting. MDL chooses a model that trades off goodness-of-fit on the observed data with complexity or richness of the model. In contrast to traditional methods such as maximum likelihood which must be modified and extended with additional, typically ad hoc principles, MDL methods inherently protect against overfitting and can be used to estimate both the parameters and the structure of a model.

To make MDL principle fully functional for quantum statistical inference this paper provides an extensive, step-by-step introduction to Quantum MDL method, with a natural attempt to establish and formulate all important definitions and theories used in classical MDL principle.

1 Introduction

1.1 MDL Principle

The minimum description length (MDL) principle is a powerful method of inductive inference, statistical modeling, pattern recognition and machine learning. It holds that the best explanation, given a limited set of observed data, is the one that permits the greatest compression of the data. MDL methods are particularly convenient for dealing with model selection, prediction, and estimation problems in situations where the models under consideration can be arbitrarily complex, and overfitting the data is a serious concern.

The oldest and simplest form of MDL is the two-part MDL, given a probabilistic model, Shannon's coding theorems present the minimal number of bits needed to encode your data, i.e., the maximum extent to which it can be compressed. However, to complete the description, you need to specify the model as well, from among some set of alternatives. Hence you want to minimize the combined length of the description of the model, plus the description of the data under that model, it is called two-part.[10]

We describe the minimum description length principle as it has been understood since the first grand overview article by Barron, Rissanen, and Yu (1998). This form of MDL is called refined MDL. A central concept in both two-part and in refined MDL are the same. "Coding the data with the help of M" is now formalized as "using a universal model relative to M." Universal models are the core of the refined MDL. For brevity, we will simply refer to refined MDL as MDL.

We always want to find structure in the observed data $D = (x_1, \cdots, x_n)$, based on some model $\mathcal{M}$, a family of probabilistic sources (probability distributions over infinite sequences). To achieve model selection or prediction, MDL invariably starts by designing a universal model $\bar{P}$ relative to the model $\mathcal{M}$ which is a single distribution, representing the model $\mathcal{M}$ such as for any data D that we might observe, the universal model performs essentially as well as the best-fitting, maximum likelihood $P \in \mathcal{M}$ for D.

1.2 Quantum MDL

To achieve Quantum MDL (QMDL) method we have to define quantum models and quantum sources which are the candidates that the given data D is produced by one of them or a combination of them. Next we define universal quantum sources relative to model $\mathcal{M}$. To reach QMDL estimators we define quantum prediction strategies or simply quantum strategies. Moreover, Quantum strategies under certain conditions on universal models lead us to "good" estimators. Afterwards we specify a good estimator for the exponential family. Based on this estimator, we explain a quantum strategy for more general quantum models and also we define a QMDL version for classical two-part code MDL. At the end, we prove that this method is consistent, the same as the classical MDL is consistent.

We aim to present coherently all essential aspects using a homogeneous notation for both classical MDL and QMDL in order to simplify the comparison between them. In this paper we provide theorems which are more general versions of the classical theorems. Since MDL classical theorems mostly are proved either for a countable set of models or for distributions over binary sequences. On the contrary, the quantum version must be defined for arbitrary set of models and quantum sources. Regularly in this paper quantum models regarded as a Riemannian submanifold. (For more information, the reader can compare Theorem 7 and Theorem 3 with Theorem 5.1 and Theorem 15.1 of [10]).

1.3 Organization of the Paper

This paper is roughly divided into three parts and is organized as follows. The first part (Sections 2), is the preliminary section to describe briefly about classical MDL definitions and general notations. In the second part (Sections 3, 4, 5 and 6), step by step, we define all essential concepts in QMDL based on MDL principle. Section 3 studies basic definitions such as quantum models, quantum sources, Bayesian and prequential quantum sources associated with some examples. In Section 4, we define universal quantum sources which are the core of classical MDL. Theorem 1 represent a constructive way to find universal models. Next we establish quantum strategy and Theorem 2 shows its relevance to prequential quantum source. At the end of this section we describe a good quantum estimator which is based on universal prequential quantum sources and quantum strategies. In Section 5, we establish a good quantum estimator for quantum exponential families. Section 6 presents a quantum strategy for more general quantum models and a quantum version of classical two-part MDL. Finally, we will argue in the third part of the paper the consistency of our method (Section 7).

2 Preliminaries

2.1 A brief introduction to MDL Principle

Here we define some main definitions of classical MDL which are used in this paper.

Definition 1. *Let $\mathcal{X}$ be a sample space and $\mathcal{X}^n$ be the set of all possible samples of length n. A probabilistic source with outcomes in $\mathcal{X}$ is a sequence $P = (P^{(n)})_{n\in\mathbb{N}}$ where $P^{(n)} : \mathcal{X}^n \to [0,1]$ is a probability distribution such that for all $n \geq 0$, all $x^n \in \mathcal{X}^n$ we have:*

$$\sum_{z\in\mathcal{X}} P^{(n+1)}(x^n, z) = P^{(n)}(x^n) \tag{2.1}$$

Definition 2. *Let $\mathcal{M}$ be a family of probabilistic sources. A universal model relative to $\mathcal{M}$ is a sequence of distributions $\bar{P} = (\bar{P}^{(n)})_{n\in\mathbb{N}}$ on $\mathcal{X}^n$, such that,*

$$\forall P \in \mathcal{M}, \forall\epsilon > 0, \exists n_0 \in \mathbb{N}, \forall n > n_0 \Rightarrow \bar{P}^{(n)} - 2^{-n\epsilon} P(x^n) > 0 \tag{2.2}$$

Suppose that we are given some candidate model $\mathcal{M}$, some data $D \in \mathcal{X}^n$ and all elements of $\mathcal{M}$ are probabilistic sources. We look for the $P \in \mathcal{M}$ that best explains the data and allows us to make good predictions of future data. For this purpose we define a universal model for $\mathcal{M}$.

There are two major types of universal models: those that can be interpreted as sequential prediction strategies for sequences of arbitrary length horizon, called "prequential"[1], and those that cannot. MDL prediction strategy is

[1]Universal Models with a predictive-sequential interpretation for arbitrary length sequences are often simply called "prequential" (predictive-sequential), emphasizing the sequential aspect of the predictions.

using a prequential universal model to predict. Estimators based on prequential universal models are called "good" MDL estimatores because prequential models have a natural interpretation to predict sequences whose length may be unknown in advance. Universal models (especially prequential models) and strategies are the main cores of MDL prediction and similarly the main cores of this paper.

2.2 Quantum Notation

In this section, after having fixed some notations that will be used in the rest of the paper, we introduce Q-projections which is equivalent to the quantum measurements. Q-projections help us to have homogeneous notation for both classical MDL and QMDL.

Given a separable Hilbert space $\mathbb{H}$. The following sets are denoted as follows:

- $B(\mathbb{H})$: The set of all bounded operators on $\mathbb{H}$
- $B_+(\mathbb{H})$: The set of all positive operators on $\mathbb{H}$
- $D(\mathbb{H})$: The set of all density matrices on $\mathbb{H}$
- $B_T(\mathbb{H})$: The set of all trace class operators on $\mathbb{H}$
- $\mathbb{H}^{(n)} := \underbrace{\mathbb{H} \otimes \mathbb{H} \otimes \cdots \otimes \mathbb{H}}_{n}$.

Remark 1. *The collection of all complete sets of mutually orthogonal (minimal) projections $P = \{p_1, p_2, ...\}$ onto $\mathbb{H}$, with $\sum_{n=1} p_n = 1$ (completeness), will be denoted by $\pi(\mathbb{H})$ ($\pi_0(\mathbb{H})$).*

Definition 3. *Let $T \in B(\mathbb{H})$ and $Q = \{q_1, q_2, ...\} \in \pi(\mathbb{H})$. Then The element*

$$T_Q = \sum_n q_n T q_n \ , \tag{2.3}$$

will be called the Q-projection of T (see also [3]) and for each $q \in Q$

$$T_q = qTq.$$

The set of all Q-projections of elements of $B(\mathbb{H})$ will be denoted by $B_Q(\mathbb{H})$ and for each $q \in Q$, $B_q(\mathbb{H}) = \{qTq | T \in B(\mathbb{H})\}$.

3 (Generalized)Quantum Model and (Generalized)Quantum Source

In this section we define quantum models and quantum sources, particularly prequential and Bayesian quantum sources and their properties.

Let ρ_θ be a density matrix, the set $M = \{\rho_\theta; \theta \in \Theta\}$ can be regarded as a Riemannian submanifold of $B_+(\mathbb{H})$ with a coordinate system θ. In the sequel we introduce some differential geometrical structures on M which assist us in many parts of this paper especially the proof of our main theorems (e.g. Theorem 1).

Now allow us to define Riemannian volume element of M and canonical measure which are needed in the definition of quantum models.

Definition 4. *Let M be a Riemannian manifold with metric tensor g and $(U, x^1, x^2, \cdots, x^n)$ be a local coordinate system on M. Then the restriction of the* ***Riemannian volume element*** *of M to U is*

$$dvol_M = |g|^{1/2} dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n \tag{3.1}$$

where $|g|$ denotes the determinant of the tensor g.

Definition 5. *Let $\mathbb{H}$ be a separable Hilbert space, $\mathfrak{B}_{B_T(\mathbb{H})}$ be the Borel σ-field of $B_T(\mathbb{H})$, and M be a subset of $B_T(\mathbb{H})$. The* ***canonical Borel σ-field*** *of subsets of M is*

$$\Sigma = \{A \cap M | A \in \mathfrak{B}_{B_T(\mathbb{H})}\}.$$

The ***canonical measure*** *of M is the counting measure when M is countable and the measure induced by the volume element $dvol_M$ when, M is a Riemannian submanifold of $B_+(\mathbb{H})$.*

From now on, we consider M as the underlying set of a measure space. It will be a Riemannian submanifold or a countable subset of $B_+(\mathbb{H})$.

Definition 6. *Let $\mathbb{H}$ be a separable Hilbert space and let $\rho \in B_+(\mathbb{H})$. We call ρ a semi-density matrix if $\mathrm{Tr}(\rho) \leq 1$.*

From now on, while we use the word "generalized", it emphasizes that the definition or the theorem is correct for both density metrices and semi-density metrices.

Definition 7.

1. *Let M be a set of density matrices on $\mathbb{H}$ with its canonical Borel σ-field Σ and its canonical measure μ. Then, the measure space $\mathcal{M} = (M, \Sigma, \mu)$ will be called a* ***quantum model*** *on $\mathbb{H}$.*

2. *More generally, if M consists of semi-density operators, then $\underline{\mathcal{M}} = (M, \Sigma, \mu)$ will be called a* ***generalized quantum model*** *on $\mathbb{H}$, and for each $Q \in \pi(\mathbb{H})$, the set $\underline{\mathcal{M}}_Q = \{\rho_Q | \rho \in \underline{\mathcal{M}}\}$ is called a* ***Q-(generalized) quantum model****.*

Remark 2. *Let $\underline{\mathcal{M}}$ be a generalized quantum model consisting of nonzero semi-density matrices and let n be a nonzero positive integer. Then, for each $\rho \in \underline{\mathcal{M}}$ we set*

$$\rho^{(n)} := \rho \otimes \underbrace{(\frac{\rho}{\mathrm{Tr}(\rho)}) \otimes \cdots \otimes (\frac{\rho}{\mathrm{Tr}(\rho)})}_{n-1}. \tag{3.2}$$

Clearly, the set $\underline{\mathcal{M}}^{(n)} = \{\rho^{(n)} | \rho \in \underline{\mathcal{M}}\}$ is a generalized quantum model on $\mathbb{H}^{(n)}$.

Remark 3. *Let $\underline{\mathcal{M}}$ be a generalized quantum model which does not contain the trivial semi-density matrix* 0*. Then, the set*

$$\mathcal{M} = \{\frac{\rho}{\mathrm{Tr}(\rho)} | \rho \in \underline{\mathcal{M}}\}$$

is a quantum model and the mapping ω from $\underline{\mathcal{M}}$ into $\mathcal{M}$ will be defined by

$$\omega(\rho) := \frac{\rho}{\mathrm{Tr}(\rho)}. \tag{3.3}$$

Definition 8. *The generalized quantum model $\underline{\mathcal{M}}$ will be called **Bayesian** if $\int_{\mathcal{M}} \rho d\mu(\rho)$ exists and is a density matrix.*

Definition 9.

1. *For each $n \in \mathbb{N}$ let $\bar{\rho}^{(n)} \in B_+(\mathbb{H}^{(n)})$ be a (semi-)density matrix. Then the sequence $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ will be called a **(generalized) quantum source**.*

2. *It is called a **(generalized) prequential quantum source** if for all $n \in \mathbb{N}$ we have*

$$tr_n(\bar{\rho}^{(n)}) = \bar{\rho}^{(n-1)}. \tag{3.4}$$

Notation: In this paper, we will denote any (generalized) quantum source by $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ where $\bar{\rho}^{(n)} \in B_+(\mathbb{H}^{(n)})$. Let $Q = \{q_1, q_2, ...\} \in \pi_0(\mathbb{H})$. By setting $I = (i_1, i_2, ..., i_n) \in \mathbb{N}^n$ the projection $\otimes_1^n q_{i_k}$ on $\mathbb{H}^{(n)}$ will be denoted by $q_I^{(n)}$ or simply by $q^{(n)}$.

Lemma 1. *Let $\underline{\mathcal{M}}$ be a Bayesian generalized quantum model, and let $\bar{\rho}^{(n)} = \int_{\underline{\mathcal{M}}} \rho^{(n)} d\mu(\rho)$. Then, the sequence $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is a prequential quantum source.*

Proof. For each $n \in \mathbb{N}$ clearly we have $tr_{n+1}(\rho^{(n+1)}) = \rho^{(n)}$. Therefore,

$$tr_{n+1}(\bar{\rho}^{(n+1)}) = \int_{\underline{\mathcal{M}}} tr_{n+1}(\rho^{(n+1)}) d\mu(\rho) = \int_{\underline{\mathcal{M}}} \rho^{(n)} d\mu(\rho) = \bar{\rho}^{(n)}.$$

■

Lemma 2. *Let $U \in B(\mathbb{H})$ be a unitary operator and $\bar{\rho}$ be a prequential quantum source. Then $U\bar{\rho}U^\dagger = (U^{(n)}\bar{\rho}^{(n)}(U^\dagger)^{(n)})_{n\in\mathbb{N}}$ is also a prequential quantum source.*

Proof.
Obviously any element $\bar{\rho}^{(n+1)} \in B(\mathbb{H}^{(n+1)})$ can be written as
$\bar{\rho}^{(n+1)} = \sum_{i,j} R_{i,j} \otimes |i\rangle\langle j|$ where $R_{i,j} \in B(\mathbb{H}^{(n)})$. Because $\bar{\rho}^{(n+1)}$ is a prequential quantum source we have

$$tr_{n+1}(\bar{\rho}^{(n+1)}) = \sum_i R_{i,i} = \bar{\rho}^{(n)}$$

So,

$$\begin{aligned}(U\bar{\rho}U^\dagger)^{(n+1)} &= U^{(n+1)}\bar{\rho}^{(n+1)}{(U^\dagger)}^{(n+1)} \\ &= \sum_{i,j=1}^{\infty}(U^{(n)}R_{i,j}{(U^\dagger)}^{(n)})\otimes U|i\rangle\langle j|U^\dagger.\end{aligned}$$

Therefore,

$$\begin{aligned}tr_{n+1}(U\bar{\rho}U^\dagger)^{(n+1)} &= \sum_{i,j=1}^{\infty}(U^{(n)}R_{i,j}{(U^\dagger)}^{(n)})Tr(U|i\rangle\langle j|U^\dagger) \\ &= \sum_{i=1}^{\infty}(U^{(n)}R_{i,i}{(U^\dagger)}^{(n)}) \\ &= U^{(n)}(\sum_{i=1}^{\infty}R_{i,i}){(U^\dagger)}^{(n)} \\ &= U^{(n)}\bar{\rho}^{(n)}(U^\dagger)^{(n)} = (U\bar{\rho}U^\dagger)^{(n)}\end{aligned} \tag{3.5}$$

Therefore, $U\bar{\rho}U^\dagger$ is a prequential quantum source.

■

4 Universal Quantum Source and Quantum Strategy

As we explained before, MDL invariably starts by designing a universal model $\bar{\rho}$ relative to the model $\mathcal{M}$. Here we define universal models, then in the Theorem 1, we describe a constructive way to build a Q-universal quantum source for each model $\mathcal{M}$. Next we define quantum strategy. In Theorem 2, it is demonstrated that each prequential source has an associated quantum strategy, and vice versa. At the end of this section we define “good” quantum estimators which are the QMDL prediction strategies based on the prequential universal models.

Definition 10. *Let $\underline{\mathcal{M}}$ be a (generalized) quantum model and $\bar{\rho}$ be a quantum source. Let $Q \in \pi(\mathbb{H})$, we say that $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is*

1. *Universal relative to $\underline{\mathcal{M}}$ if for each $\rho \in \underline{\mathcal{M}}$ and for each $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$ we have:*

$$\bar{\rho}^{(n)} - 2^{-n\epsilon}\rho^{(n)} \geq 0. \tag{4.1}$$

2. *Universal in the expected sense related to $\underline{\mathcal{M}}$ if:*

$$S(\rho^{(n)}\|\bar{\rho}^{(n)}) \leq n\epsilon. \tag{4.2}$$

3. *Q-Universal relative to $\underline{\mathcal{M}}$ if for each $\rho \in \underline{\mathcal{M}}$ and for each $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$ we have:*

$$\bar{\rho}_Q^{(n)} - 2^{-n\epsilon}\rho_Q^{(n)} \geq 0. \tag{4.3}$$

And Q-universal relative to $\underline{\mathcal{M}}$ in the expected sense if:

$$S({\rho_Q}^{(n)}\|\bar{\rho}_Q^{(n)}) \leq n\epsilon. \tag{4.4}$$

In the above if ϵ does not depend on ρ, $\bar{\rho}$ is called uniformly (Q-)universal.

Lemma 3. *With the above notations and conventions 1 implies 2 and 3.*

Proof. It is clear that 1 implies 3. For 1 implies 2, we have

$$\begin{array}{ll} & \bar{\rho}^{(n)} - 2^{-n\epsilon}\rho^{(n)} \geq 0 \\ \Rightarrow & n\epsilon + log\bar{\rho}^{(n)} - log\rho^{(n)} \geq 0 \\ \Rightarrow & n\epsilon + (\rho^{(n)})^{1/2}log\bar{\rho}^{(n)} - log\rho^{(n)}(\rho^{(n)})^{1/2} \geq 0 \\ \Rightarrow & n\epsilon + \mathrm{Tr}\rho^{(n)}(log\bar{\rho}^{(n)} - log\rho^{(n)}) \geq 0. \\ \Rightarrow & S(\rho^{(n)}\|\bar{\rho}^{(n)}) \leq n\epsilon. \end{array} \tag{4.5}$$

■

In the following there are two examples of universal models.

Example 1. *Let $\underline{\mathcal{M}}$ be a Bayesian countable generalized quantum model consisting of nonzero semi-density matrices and let $\mathcal{M}$ be its associated quantum model. Then for each element $\rho^* \in \underline{\mathcal{M}}$ and each $n \in \mathbb{N}$ we have*

$$\bar{\rho}^{(n)} = \sum\nolimits_{\rho\in\underline{\mathcal{M}}}\rho^{(n)} \geq \rho^{*(n)}.$$

Now let ϵ be given and let $n_0 \in \mathbb{N}$ be such that

$$\mathrm{Tr}(\rho^*) \geq 2^{-(n_0)\epsilon}.$$

Then, for each $n \geq n_0$ we have

$$\bar{\rho}^{(n)} - 2^{-n\epsilon}\rho^{(n)} \geq 0.$$

Therefore, $\bar{\rho}$ is universal for $\mathcal{M}$.

Example 2. *Let $\underline{\mathcal{M}}$ be a generalized quantum model and $\bar{\rho}$ be a universal source for $\underline{\mathcal{M}}$ and U be a unitary operator. Then $U\bar{\rho}U^{-1}$ is a universal element for $U\underline{\mathcal{M}}U^{-1}$ where $U\underline{\mathcal{M}}U^{-1} = \{U\rho U^{-1}|\rho \in \underline{\mathcal{M}}\}$*

Here we describe the Theorem 1 which is the main result of this section to find a Q-universal model for each model $\mathcal{M}$.

Theorem 1. *Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\mathcal{M} = (M, \Sigma, \mu)$ be a quantum model where M is a compact Riemannian submanifold of $B_T(\mathbb{H})$, consisting of density matrices. Assume that*

1. *$p : M \longrightarrow]0, \infty[$ is a continuous function and $\int_M p(\rho)dvol_M(\rho) = 1$*
2. *There exists a positive real $c > 0$ such that for all $q \in Q$, $q\rho_q q \geq cq$.*

Where $\rho_q \in M$ is such that we have

$$q\rho_q q = max_{\rho\in M}q\rho q.$$

Moreover, for each $n \in \mathbb{N}$ let $\bar{A}^{(n)} = \int_M p(\rho)\rho^{(n)}dvol_M(\rho)$, and $A^{(1)} = A$. Then the sequence $(\bar{A}^{(n)})_{n\in\mathbb{N}}$ is uniformly Q-universal for $\mathcal{M}$.

Proof. Let $\epsilon > 0$ be given and let $\delta = c(1 - 2^{-\epsilon/2}) > 0$.

Now for all $\rho_q^* \in B(\rho_q, \delta)$ we have

$$\begin{aligned} \mathrm{Tr}(q\rho_q q) - \mathrm{Tr}(q\rho_q^* q) &= \mathrm{Tr}(q(\rho_q - \rho_q^*)q) \\ &= \|q(\rho_q - \rho_q^*)q\| \\ &\leq \|\rho_q - \rho_q^*\| \\ &\leq \|\rho_q - \rho_q^*\|_1 \\ &\leq d(\rho_q, \rho_q') \leq \delta. \end{aligned} \tag{4.6}$$

and it is straightforward to see that for all $\rho_q^* \in B(\rho_q, \delta)$ we have

$$q\rho_q^* q \geq 2^{-\epsilon/2} q\rho_q q.$$

Now let $\beta = min_{\rho \in M} p(\rho)$. Moreover, since M is compact, as it is proved in [2], there exists a constant $v > 0$ such that for all $\rho \in M$ we have

$$vol_M B(\rho, \delta) \geq v.$$

Let $k \in \mathbb{N}$ be such that $\beta v \geq 2^{-k\epsilon/2}$ Then we have

$$\begin{aligned} qAq &= \int_M p(\rho) q\rho q dvol_M(\rho) \\ &\geq \int_{B(\rho_q,\delta)} p(\rho) q\rho q dvol_M(\rho) \\ &\geq 2^{-\epsilon/2} \beta vol(B(\rho_q, \delta)) q\rho_q q \\ &\geq 2^{-(k+1)\epsilon/2} q\rho_q q. \end{aligned} \tag{4.7}$$

Let us denote $\otimes^n \rho$ by $\rho^{(n)}$ and $\int_M p(\rho)\rho^{(n)} dvol_M(\rho)$ by $\bar{A}^{(n)}$. Let $n \in \mathbb{N}$ be greater than k. Then from the above it is evident that for all $\rho \in M$ we have

$$q^{(n)}\bar{A}^{(n)}q^{(n)} \geq 2^{-(k+n)\epsilon/2} q^{(n)}\rho^{(n)}q^{(n)} \geq 2^{-n\epsilon} q^{(n)}\rho^{(n)}q^{(n)}.$$

And for each $\rho \in M$ we have

$$\begin{aligned} \bar{A}_Q^{(n)} &= \sum_{q^{(n)} \in Q^{(n)}} (q^{(n)}\bar{A}^{(n)}q^{(n)}) \\ &\geq 2^{-n\epsilon} \sum_{q^{(n)} \in Q^{(n)}} (q^{(n)}\rho^{(n)}q^{(n)}) = 2^{-n\epsilon}\rho_Q^{(n)}. \end{aligned} \tag{4.8}$$

Therefore, the quantum source $(\bar{A}^{(n)})_{n \in \mathbb{N}}$ is uniformly Q-universal for $\mathcal{M}$. ∎

Corollary 1. *Let $\mathbb{H}$ be a Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\mathcal{M} = (M, \Sigma, \mu)$ be a quantum model and M be a compact Riemannian submanifold of $B_T(\mathbb{H})$ consisting of regular density matrices. Assume that $p : M \longrightarrow]0, \infty[$ is a continuous function such that $\int_M p(\rho) dvol_M(\rho) = 1$. Then, with the above notations the sequence $(\bar{A}^{(n)})_{n \in \mathbb{N}}$ is Q-universal for $\mathcal{M}$.*

Lemma 4. *$S_{\mathcal{M}}$, the set of all Q-universal (prequential) quantum source for the quantum model $\mathcal{M}$, is a convex subset of $D_Q(\mathbb{H})$.*

Proof. Let $\bar{\rho}_1$ and $\bar{\rho}_2$ be two Q-universal quantum source for the quantum model $\mathcal{M}$. Let $\rho \in \mathcal{M}$ and $\epsilon > 0$ be given. For simplicity we omit the index Q. Then there exists $n_0 \in \mathbb{N}$ such that for $k = 1, 2$ and $n \geq n_0$ we have:

$$\bar{\rho}_k^{(n)} - 2^{-n\epsilon}\rho^{(n)} \geq 0.$$

Let α and β be two positive real numbers such that $\alpha + \beta = 1$. Then

$$\alpha\bar{\rho}_1^{(n)} + \beta\bar{\rho}_2^{(n)} - 2^{-n\epsilon}\rho^{(n)} =$$

$$\alpha(\bar{\rho}_1^{(n)} - 2^{-n\epsilon}\rho^{(n)}) + \beta(\bar{\rho}_2^{(n)} - 2^{-n\epsilon}\rho^{(n)}) \geq 0.$$

Therefore $S_{\mathcal{M}}$ at each level n is convex. On the other hand,

$$(\alpha\bar{\rho}_1 + \beta\bar{\rho}_2)^{(n)} = \alpha\bar{\rho}_1^{(n)} + \beta\bar{\rho}_2^{(n)} \in (S_{\mathcal{M}})^{(n)},$$

where $(S_{\mathcal{M}})^{(n)} = \{\bar{\rho}^{(n)} | \bar{\rho} \in S_{\mathcal{M}}\}$, Therefore

$$\alpha\bar{\rho}_1 + \beta\bar{\rho}_2 \in S_{\mathcal{M}}.$$

■

Here we define Q-quantum prediction strategy or simply Q-quantum strategy and its close relationship with prequential Q-quantum source. Then according to these two definitions and the definition of universal models we define the concept of a good quantum estimators based on MDL approach.

Convention 1. *Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be Hilbert spaces, and let $T \in B(\mathbb{H}_1 \otimes \mathbb{H}_2)$ and $T_1 \in B_+(\mathbb{H}_1)$. We denote*

$$T_1 \circ T := (T_1{}^{\frac{1}{2}} \otimes I_2)T(T_1{}^{\frac{1}{2}} \otimes I_2) \tag{4.9}$$

Here I_2 is the identity mapping of $\mathbb{H}_2$.

Definition 11. *Let $Q = \{q_1, q_2, ...\}$ be a complete set of mutually orthogonal minimal projections of the Hilbert space $\mathbb{H}$ and for each $n \in \mathbb{N}$ let $\hat{\rho}^{(n)} = \hat{\rho}_Q^{(n)} \in B_+(\mathbb{H}^{(n)})$, and*

$$\forall n > 1, q_I^{(n-1)} \in Q^{(n-1)} \Rightarrow \mathrm{Tr}(q_I^{(n-1)} \circ (\hat{\rho}^{(n)})) = 1. \tag{4.10}$$

Then the sequence $\hat{\rho} = (\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ will be called a Q-quantum strategy, and for simplicity the density matrix $q_I^{(n-1)} \circ (\hat{\rho}^{(n)}) \in D(\mathbb{H})$ will be denoted by $\hat{\rho}_I^{(n)}$.

Definition 12. *(Conditional density matrices)(See also [20].) Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be Hilbert spaces and $\rho \in D(\mathbb{H}_1 \otimes \mathbb{H}_2)$. Let $\rho_1 \in D(\mathbb{H}_1)$ be defined as $\rho_1 = tr_2\rho$. Then the operator*

$$\rho_{|(1)} = \rho_1^{-1} \circ \rho \tag{4.11}$$

will be called the conditional density matrix of ρ on $\mathbb{H}_2$ conditioned on $\rho_1 \in D(\mathbb{H}_1)$. Clearly we have:

$$\rho_1 \circ \rho_{|(1)} = \rho. \tag{4.12}$$

Remark 4. *In this paper, for each operator T when we need the inverse of T, we assume that it is invertible. In the general case one can use the so-called "Moore-Penrose" inverse of T.*

Lemma 5. *Let $Q \in \pi_0(\mathbb{H}_1)$ and $P \in \pi_0(\mathbb{H}_2)$. Assume that the density matrix ρ have the following spectral decomposition*

$$\rho = \sum_{i,j=1}^{\infty} \lambda_{ij} q_i \otimes p_j.$$

Then for each $i \in \mathbb{N}$ the operator

$$\rho_{|i} := tr_1(q_i \circ \rho_{|(1)}) \tag{4.13}$$

is really a density matrix.

Proof. The operator $\rho_{|(1)}$ is positive. Therefore, $\rho_{|i}$ is clearly positive. Now we are going to prove that $\mathrm{Tr}(\rho_{|i})$ is equal to 1. Clearly, we have

$$\begin{aligned} \rho_1 = tr_2(\rho) = \textstyle\sum_{i=1}^{\infty}(\sum_{j=1}^{\infty}\lambda_{ij})q_i \\ \Longrightarrow (\rho_1)^{\frac{1}{2}} = \textstyle\sum_{i=1}^{\infty}(\sum_{j=1}^{\infty}\lambda_{ij})^{\frac{1}{2}}q_i. \end{aligned} \tag{4.14}$$

Therefore,

$$\begin{aligned} \rho_{|(1)} = & \; (\rho_1^{-1}) \circ \rho = (\rho_1{}^{-\frac{1}{2}} \otimes I)\rho(\rho_1{}^{-\frac{1}{2}} \otimes I) \\ = & \; ((\textstyle\sum_{i=1}^{\infty}(\sum_{k=1}^{\infty}\lambda_{ik})^{-\frac{1}{2}}q_i) \otimes I)\times \\ & \; (\textstyle\sum_{i,j=1}^{\infty}\lambda_{ij}(q_i \otimes p_j))\times \\ & \; ((\textstyle\sum_{i=1}^{\infty}(\sum_{k=1}^{\infty}\lambda_{ik})^{-\frac{1}{2}}q_i) \otimes I) \\ = & \; \textstyle\sum_{i,j=1}^{\infty}(\sum_{k=1}^{\infty}\lambda_{ik})^{-1}\lambda_{ij}(q_i \otimes p_j) \\ = & \; \textstyle\sum_{i=1}^{\infty}(q_i \otimes \rho_{|i}). \end{aligned} \tag{4.15}$$

Where,

$$\rho_{|i} = \sum_{j=1}^{\infty}((\sum_{k=1}^{\infty}\lambda_{ik})^{-1})\lambda_{ij}p_j.$$

Therefore $\mathrm{Tr}(\rho_{|i}) = 1$ and $\rho_{|i}$ is a density matrix.

■

Lemma 6. *Let*

$$T = \sum_{i,j=1}^{\infty}\lambda_{ij}(q_i \otimes p_j)$$

be a positive operator on $\mathbb{H}_1 \otimes \mathbb{H}_2$ such that for each $i \in \mathbb{N}$, $Tr_1(q_i \circ T)$ is a density matrix on $\mathbb{H}_2$, and

$$\rho_1 = \sum_{i=1}^{\infty}\lambda_i q_i$$

is a density matrix on $\mathbb{H}_1$. Then $\rho = \rho_1 \circ T$ is a density matrix on $\mathbb{H}_1 \otimes \mathbb{H}_2$. moreover, $T = \rho_{|(1)}$.

Proof. Clearly, ρ is a positive operator of $\mathbb{H}_1 \otimes \mathbb{H}_2$. On the other hand,

$$\mathrm{Tr}(\rho) = \mathrm{Tr}(\rho_1 \circ T) = \sum_{i,j=1}^{\infty} \lambda_i \lambda_{ij} = \sum_{i=1}^{\infty} \lambda_i \sum_{j=1}^{\infty} \lambda_{ij} = \sum_{i=1}^{\infty} \lambda_i = 1.$$

Therefore, ρ is a density matrix on $\mathbb{H}_1 \otimes \mathbb{H}_2$. Now,

$$T = \rho_1^{-1} \circ \rho = \rho_{|(1)}.$$

■

Convention 2. *Let $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be a quantum source. Then, for each $n_1, n_2 \in \mathbb{N}$ and $n_1 > n_2$, the operator $(\bar{\rho}^{(n_2)})^{-1} \circ \bar{\rho}^{(n_1)}$ will be denoted by $\bar{\rho}_{n_1|(n_2)}$.*

Theorem 2. *Let $\hat{\rho}$ be a Q-quantum strategy. Let us define the sequence $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ recursively as follows:*

1. $\bar{\rho}^{(1)} = \hat{\rho}^{(1)}$.
2. $\bar{\rho}^{(n+1)} = \bar{\rho}^{(n)} \circ \hat{\rho}^{(n+1)}$.

Then the sequence $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is a prequential Q-quantum source.
Conversely, let $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be a prequential Q-quantum source. Let

$$\hat{\rho}^{(n+1)} := (\bar{\rho}^{(n)})^{-1} \circ \bar{\rho}^{(n+1)} = \bar{\rho}_{n|(n+1)}. \tag{4.16}$$

Then the sequence $(\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ is a Q-quantum strategy.
The proof is a direct consequence of Lemma 5 and Lemma 6.

■

Lemma 7. *Let $(\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ be a Q-quantum strategy and $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be its associated Q-quantum source. Then for each $T \in B(\mathbb{H})$ and each $n \in \mathbb{N}$, $T^{(n)}\hat{\rho}^{(n)} = \hat{\rho}^{(n)}T^{(n)}$ if and only if $T^{(n)}\bar{\rho}^{(n)} = \bar{\rho}^{(n)}T^{(n)}$.*

The proof is straightforward.

■

Remark 5. *For future applications we mention that because of the equality $\hat{\rho}^{(n+1)} = \bar{\rho}_{n|(n+1)}$, Q-quantum strategies are also called Q-quantum estimators. Let $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be a prequential quantum source. It is straightforward to see that $(\bar{\rho}_Q^{(n)})_{n\in\mathbb{N}}$ is a prequential Q-quantum source and gives rise to a Q-quantum strategy.*

Definition 13. *A quantum strategy $(\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ is called good quantum estimator with respect to a quantum model $\mathcal{M}$ if its associated prequential source is universal for $\mathcal{M}$.*

5 Quantum Exponential Families

Now we define a QMDL prediction method for a set of quantum exponential family which, based on Definition 13, is a good quantum estimator.

The material in this section is extracted from [10] with necessary modifications. The steps are based on [10] to clarify the relationship between QMDL and classical MDL.

Let $\mathbb{H}$ be a separable Hilbert space, $n \in \mathbb{N}$ and $\Theta_{can} \subseteq \mathbb{R}^n$. Let

$$\psi : \Theta_{can} \longrightarrow \mathbb{R} \quad \text{and} \quad T = (T_1, T_2, ..., T_n),$$

where for $n \geq k \geq 1$, T_k is a bounded self-adjoint operator. Let ρ_0 be a regular density matrix of $\mathbb{H}$. Assume that

$$\rho_\theta = exp(\frac{\theta T - \psi(\theta)}{2})\rho_0 exp(\frac{\theta T - \psi(\theta)}{2}), \tag{5.1}$$

where for $\theta = (\theta_1, \theta_2, ..., \theta_n) \in \Theta_{can}$, $\theta T = \Sigma_{k=1}^{n} \theta_k T_k$.

The set $\{\rho_\theta | \theta \in \Theta_{can}\}$ is a quantum model. Then, the family $(\rho_\theta)_{\theta\in\Theta}$ is called an n-dimensional quantum exponential family. For more details on exponential families and quantum exponential families see [10], [21] and [23].

In this paper for simplicity we assume that Θ_{can} is an open interval of $\mathbb{R}$. Clearly ψ is smooth on a neighborhood of the closure of Θ_{can}. Now, assume that

$$Q \in \pi_0(\mathbb{H}) \text{ and } T = T_Q.$$

From the equality

$$1 = \mathrm{Tr}(\rho_{\theta Q}) = \mathrm{Tr}(exp(\theta T - \psi(\theta))\rho_{0Q}),$$

we have

$$Tr(exp(\theta T)\rho_{0Q}) = exp(\psi(\theta)).$$

And from this we have

$$\frac{d}{d\theta}\psi(\theta) = \mathrm{Tr}(T exp(\theta T - \psi(\theta))\rho_{0Q}) = E_\theta(T) \tag{5.2}$$

where $E_\theta(T)$ is the mean value of T with respect to $\rho_{\theta Q}$ and

$$\mathrm{Var}_\theta(T) = E_\theta(T^2) - (E_\theta(T))^2 = \frac{d^2}{d\theta^2}\psi(\theta) = I_\theta. \tag{5.3}$$

Where I_θ is the Fisher information of the associated Q-exponential family $\bar{Q}(\rho_\theta)$ at θ.

I_θ is strictly positive on Θ_{can}. Therefore, the mapping

$$\mu : \Theta_{can} \longrightarrow \Theta_{mean} = \mu(\Theta_{can})$$

defined by

$$\mu(\theta) = E_\theta(T)$$

is a diffeomorphism. For each $t \in \Theta_{mean}$ let us denote $\rho_{\mu^{-1}(t)}$ by ρ_t. Clearly we have $\{\rho_\theta | \theta \in \Theta_{can}\} = \{\rho_t | t \in \Theta_{mean}\}$.

Now, assume that $\alpha > 0$, and $t_0 \in \Theta_{mean}$. For each $n \in \mathbb{N}$ and for each $I \in \mathbb{N}^n$, let

$$q_I{}^{(n-1)} \circ \hat{\rho}^{(n)} := \frac{\alpha t_0 + \Sigma_{k=1}^{n} \langle q_{i_k} | T q_{i_k} \rangle_1}{\alpha + n}. \tag{5.4}$$

Then $\hat{\rho} = (\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ will be a Q-quantum strategy and this modified quantum strategy is a good quantum estimator. For more detail see [10]

Let $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be the associated Q-universal quantum source for $\mathcal{M}$. Then, as it is proved in [10] with some obvious modifications one can show that for all $\rho \in \mathcal{M}$ we have

$$S(\rho^{(n)} \| \bar{\rho}^{(n)}) = O(log(n)). \tag{5.5}$$

6 Quantum Prediction and Quantum Estimation

Quantum Prediction and Quantum Estimation are the most important subjects of Quantum Statistical Inference. Following what we have said in the introduction concerning MDL principle, our method of statistical inference is in general based on universal quantum elements and use of it to do Quantum Prediction and Quantum Estimation. In this section we suggest two quantum versions of classical MDL. The first one is based on the method that we defined in Section 5 for quantum exponential family and the second one is based on the oldest and simplest version of classical MDL, Two-Part code MDL.

1. Let $\mathbb{H}$ be a separable Hilbert space and let $Q \in \pi_0(\mathbb{H})$. Let $\mathcal{M}$ be a quantum model and let $\hat{\rho}^{(n)} \in B_+(\mathbb{H})$ be such that for $I \in Q^{(n-1)}$ we have

   $$q_I^{(n-1)} \circ \hat{\rho}^{(n)} := argmax_{\rho \in \mathcal{M}} (q_I^{(n-1)} \rho^{(n-1)} q_I^{(n-1)}). \tag{6.1}$$

   Clearly, $\hat{\rho} = (\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ is the maximum likelihood Q-quantum strategy associated with $\mathcal{M}$. Unfortunately, $\hat{\rho}$ is not a good QMDL estimator as we defined in Definition 13. But in many cases (for example, quantum exponential families in Section 5), a modified version of the maximum likelihood Q-quantum strategy is very close to the unmodified one and the difference between them tends rapidly to zero.

   This Q-quantum strategy enables us to predict next outcome given the data $\mathcal{D} = \{q_{i_1}, q_{i_2}, ..., q_{i_{n-1}}\}$. Moreover, it enables us to estimate the state of the system given the data set $\mathcal{D}$.

2. Let $\mathbb{H}$ be a separable Hilbert space and let $Q \in \pi_0(\mathbb{H})$. Assume that $\underline{\mathcal{M}}$ is a generalized quantum model. For $I \in \mathbb{N}$, let $\ddot{\rho}_n$ be defined as follows

$$\ddot{\rho}_n := \omega(argmax_{\rho\in\underline{\mathcal{M}}} q_I^n \rho^{(n)} q_I^n). \tag{6.2}$$

If the maximum is achieved by more than one ρ we choose the one with the maximum trace. And if there is still more than one ρ there is no further preference.

According to the above explanation, given the outcome $\mathcal{D} = \{q_{i_1}, q_{i_2}, ..., q_{i_n}\}$, $\ddot{\rho}_n$ is a estimator of the state of the system.

To obtain $\ddot{\rho}_n$, it is useful to suppose that M is a compact Riemannian submanifold of the Hilbert space $(B_T(\mathbb{H}), \langle.|.\rangle)$ consisting of semi-density matrices where for ρ and ρ' in $B_T(\mathbb{H})$, $\langle\rho|\rho'\rangle = \mathrm{Tr}(\rho\rho')$. Let the generalized quantum model $\underline{\mathcal{M}} = (M, \Sigma, \mu)$ be its associated canonical measure space. Now let Z be the set of all extremum points of the smooth function $h : \rho \longrightarrow \mathrm{Tr}(q_I^n \rho^{(n)} q_I^n)$ on M, and let Z' be the set of all elements $\rho \in Z$ at which the bundle map $Hessian(h) : TM \longrightarrow TM$ is negative. Clearly, Z' is the set of all maximum points of h. Now, let ρ_0 be the element of Z' with least trace. Then, $\ddot{\rho}_n = \omega(\rho_0)$. If there are more than one ρ_0 in Z' we do not have any further preference among them.

7 Consistency and Convergence

Consistency is a very important property of different methods of statistical (inductive) inferences. Like classical MDL we need to proof the consistency of our method.

Assume that $\mathbb{H}$ is a separable Hilbert space and $\mathcal{M}$ is a quantum model on $\mathbb{H}$. Let $\rho_0 \in \mathcal{M}$ and $Q \in \pi_0(\mathbb{H})$, we say that a method of quantum statistical inference is consistent with respect to $\mathcal{M}$ if by repeatedly performing the same measurement on more and more quantum systems all prepared in the same state ρ_0, the estimated state yielded by the method is more and more close to the state ρ_0 in some sense.

In this section we investigate different approaches to consistency.

7.1 Consistency in the Sense of Probability Convergence

Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let ρ be a density matrix on $\mathbb{H}$ and for each $n \in \mathbb{N}$ let P_n be a unary relation on $Q^{(n)}$. Then the semi-density matrix $\sum_{q^{(n)}\in Q^{(n)}|P_n(q^{(n)})} q^{(n)}\rho^{(n)}q^{(n)}$ will be denoted by $\rho(P_n)$. Now let ρ' be also a density matrices on $\mathbb{H}$. For each $n \in \mathbb{N}$, and $\delta > 0$ let ${P_n}^{\delta}$ be the unary relation

$$q^{(n)}\rho' q^{(n)}/q^{(n)}\rho q^{(n)} > \delta \tag{7.1}$$

on $Q^{(n)}$.

Definition 14. *Under the above notations and conventions we say ρ' is asymptotically distinguishable from ρ if for all $\delta > 0$ we have*

$$lim_{n\to\infty}\rho(P_n^{\delta}) = 0.$$

Let $\underline{\mathcal{M}}$ be a generalized quantum model on $\mathbb{H}$ and $\ddot{\rho}_n$ be defined as follows

$$\ddot{\rho}_{(n)}(q^{(n)}) = \omega(argmax_{\rho\in\underline{\mathcal{M}}}q^{(n)}\rho^{(n)}q^{(n)}).$$

Now assume that $\mathcal{M}$ is a quantum model. Then the unary relation $\ddot{\rho}_{(n)}(q^{(n)}) \in \mathcal{M}$ on $Q^{(n)}$ will be denoted by $\ddot{\rho}_{(n)} \in \mathcal{M}$.

Now we have the following important consistency theorem.

Theorem 3. *Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\underline{\mathcal{M}}$ be a generalized quantum model. Assume that for all $n \in N$, $\ddot{\rho}_{(n)}$ is defined as above. Let $\mathcal{M}$ be the quantum model associated with $\underline{\mathcal{M}}$ and $\rho' \in \underline{\mathcal{M}}$ and $\rho^* = \omega(\rho') \in \mathcal{M}$. Let $\ddot{\mathcal{M}}$ be the subset of $\mathcal{M}$ consisting of density matrices asymptotically distinguishable from ρ^*. Then*

$$lim_{n\to\infty}\rho^*(\ddot{\rho}_{(n)} \in \ddot{\mathcal{M}}) = 0. \quad (7.2)$$

Proof. Let $D \subset Q^*$ be defined as follows:
$q^{(n)} \in Q^*$, is in D if and only if there exists an element $\rho \in \underline{\mathcal{M}}$ different from ρ' such that

$$q^{(n)}{\rho'}^{(n)}q^{(n)} \leq q^{(n)}\rho^{(n)}q^{(n)}.$$

Let the mapping $\eta : D \longrightarrow \underline{\mathcal{M}}$ be a function with the following property
For each $q^{(n)} \in D$, if there exists $\rho \in \underline{\mathcal{M}}$ satisfying the above inequality and $\omega(\rho) \in \ddot{\mathcal{M}}$ then $\eta(q^{(n)}) = \rho$. Otherwise, $\eta(q^{(n)}) = \rho$ is an element of $\underline{\mathcal{M}}$ with the above property. Let $\mathcal{M}_1$ be the image of D under $\omega \circ \eta$, and $\mathcal{M}_2 = \mathcal{M}_1 \cup \{\rho^*\}$. Clearly, $\mathcal{M}_2$ is a countable set and for each $n \in \mathbb{N}$ we have

$$\rho^*(\ddot{\rho}_{(n)} \in \ddot{\mathcal{M}}) \leq \rho^*(D^{(n)}).$$

Let $\mathcal{M}^* = \ddot{\mathcal{M}} \cap \mathcal{M}_2$. Now, in the countable quantum model $\mathcal{M}_2$, the set of all elements asymptotically distinguishable from ρ^* is $\mathcal{M}^*$. The rest of the proof is the same as the proof of Theorem 5.1 of [10].

■

7.2 Consistency in the sense of Cezaro, Hellinger and Re'nyi divergence

Definition 15. *Let $\rho, \rho' \in D(\mathbb{H})$, then*
1) The Hellinger distance of ρ and ρ' is defined as follows

$$He^2(\rho\|\rho') = \|\rho^{1/2} - {\rho'}^{1/2}\|_1^2.$$

2) Let $\lambda > 0$ be a real number, the Renyi divergence of order λ of ρ and ρ' is defined as follows

$$\bar{d}_\lambda(\rho\|\rho') = -\frac{1}{1-\lambda} ln(\langle \rho^\lambda | {\rho'}^{1-\lambda}\rangle_1).$$

Remark 6. *The above lemma is also true when $f : I \longrightarrow [0, \infty[$ is a bounded continuous function with finite number of extremum point, and the proof is like above with some modifications.*

Theorem 4. *Let $\bar{\rho}$ and ρ^* be Quantum sources. Then*

$$S({\rho^*}^{(n)}\|\bar{\rho}^{(n)}) =^w \sum_{i=1}^{n} E_{\rho^{*(i-1)}} S(\rho^*_{i|(i-1)}\|\bar{\rho}_{i|(i-1)}). \tag{7.3}$$

Proof. Assume that Q is a complete set of mutually orthogonal minimal projections. For the moment we assume that $\rho^{*(n)}_{Q^{(n)}}$ and $\bar{\rho}^{(n)}_{Q^{(n)}}$ are invertible. For simplicity we omit the subscript $Q^{(k)}$. By definition, and previous lemmas and theorems we have:

$$\begin{aligned}
S({\rho^*}^{(n)}\|\bar{\rho}^{(n)}) &= \mathrm{Tr}{\rho^*}^{(n)} \log {\rho^*}^{(n)} - {\rho^*}^{(n)} \log \bar{\rho}^{(n)}) \\
&= \mathrm{Tr}{\rho^*}^{(n)}(\log {\rho^*}^{(n)} - \log \bar{\rho}^{(n)}) \\
&= \mathrm{Tr}\Pi_{i=1}^{n}(\rho^*_{i|(i-1)})(\log \Pi_{i=1}^{n}\rho^*_{i|(i-1)} - \log \Pi_{i=1}^{n}\bar{\rho}_{i|(i-1)})) \\
&= \textstyle\sum_{i=1}^{n} \mathrm{Tr}({\rho^*}^{(i-1)})\rho^*_{i|(i-1)}(\log \rho^*_{i|(i-1)} - \log \bar{\rho}_{i|(i-1)})) \\
&= \textstyle\sum_{i=1}^{n} E_{\rho^{*(i-1)}} S(\rho^*_{i|(i-1)}\|\bar{\rho}_{i|(i-1)}).
\end{aligned}$$

(See also [10].)

■

Theorem 5. *(Convergence Theorem for quantum Estimators) (Barron 1998) Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\mathcal{M}$ be a set of quantum sources on $\mathbb{H}$, and $\bar{\rho}$ be a prequential Q-universal quantum source with respect to $\mathcal{M}$. Then $\hat{\rho}$, the Q-quantum estimator associated with Q-universal quantum source $\bar{\rho}^{(n)}_Q$, is Cesaro consistent with respect to $\mathcal{M}$. In other words*

$$\lim_{n\to\infty} \frac{1}{n} S(\bar{\rho}^{(n)}_Q\|\rho^{(n)}_Q) = 0. \tag{7.4}$$

The proof is a consequence of the definition of Q-universal element and lemma 7.

■

Let $\underline{\mathcal{M}}$ be a generalized quantum model such that there exist a countable and dense Bayesian subset $\underline{\mathcal{M}}'$ of $\underline{\mathcal{M}}$.

For $\alpha \geq 1$, let $\underline{\mathcal{M}}_\alpha = \{\rho_\alpha | \rho \in \underline{\mathcal{M}}\}$. Where, $\rho_\alpha = [tr(\rho)]^{\alpha-1}\rho$. Let

$$\ddot{\rho}_{\alpha n} = \omega(argmax_{\rho\in\underline{\mathcal{M}}} q_I^n {\rho_\alpha}^{(n)} q_I^n).$$

And let $\bar{\rho_\alpha}^{(n)}$ be defined as follows

$$q_I^n \bar{\rho_\alpha}^{(n)} q_I^n = max_{\rho \in \underline{\mathcal{M}}}(q_I^n \rho_\alpha{}^{(n)} q_I^n).$$

In many cases for example when Q is finite and M is a compact Riemannian manifold $\bar{\rho_\alpha} = (\bar{\rho_\alpha}^{(n)})_{n\in\mathcal{N}}$ is a Q-quantum universal source for $\underline{\mathcal{M}}$.

Now let $\bar{\rho_\alpha} = (\bar{\rho_\alpha}^{(n)})_{n\in\mathcal{N}}$ be a Q-quantum universal source for $\mathcal{M} = \omega(\underline{\mathcal{M}})$. Then we have the following theorem.

Theorem 6. *Under the above notations and conventions for $\alpha > 1$ and all $0 < \lambda \leq 1 - 1/\alpha$ we have*

$$E_{\rho_Q^{(n)}}(\bar{d}_\lambda(\rho_Q^{(n)} \| \ddot{\rho}_{\alpha n Q})) \leq \frac{1}{n} S(\rho_Q^{(n)} \| \bar{\rho}_{\alpha Q}^{(n)}). \tag{7.5}$$

And for $\alpha = 2$ we have

$$E_{\rho_Q^{(n)}}(He^2(\rho_Q^{(n)} \| \ddot{\rho}_{\alpha n Q})) \leq \frac{1}{n} S(\rho_Q^{(n)} \| \bar{\rho}_{\alpha Q}^{(n)}). \tag{7.6}$$

Proof. For $\underline{\mathcal{M}}$ a Countable Bayesian model the proof is an special case of the proof of Theorem 15.3 of [10] with obvious modifications and when $\underline{\mathcal{M}}$ is a compact Riemannian manifold the proof is a consequence of the fact that $\underline{\mathcal{M}}'$ is a countable dense Bayesian subset of $\underline{\mathcal{M}}$.

■

Definition 16. *Let R be an n-ary relation on $B(\mathbb{H})$ and let $T_1, T_2, ..., T_n \in B(\mathbb{H})$, ($n$ may be infinity). We say that $R(T_1, T_2, ..., T_n)$ is weakly true if, for each $Q \in \pi_0(\mathbb{H})$, $R(T_{1Q}, T_{2Q}, ..., T_{nQ})$ is true. The "weakly equal" relation will be denoted by $=^w$.*

Definition 17. *Let ρ^* and $\bar{\rho}$ be quantum sources. Then,*

1. *The standard KL-risk of ρ^* and $\bar{\rho}$ is*

$$Risk_n(\rho^*, \bar{\rho}) =^w E_{\rho^*}[S(\rho^*_{n|n-1} \| \bar{\rho}_{n|n-1})]$$

2. *The Cesaro risk of ρ^* and $\bar{\rho}$ is*

$$\overline{Risk}_n(\rho^*, \bar{\rho}) =^w \frac{1}{n} S(\rho^{*(n)} \| \bar{\rho}^{(n)})$$

The following equality is a consequence of Lemma 8 and the definition of KL-risk.

$$\overline{Risk}_n(\rho^*, \bar{\rho}) =^w \frac{1}{n} \Sigma_{j=1}^n Risk_j(\rho^*, \bar{\rho})$$

See also [10].

Remark 7. *Let $\epsilon < 1$ be a nonzero positive real number and let $t : [\epsilon, \infty[\longrightarrow [1, \infty[$ be defined as follows*

$$t(x) = (x + 1 - 2\epsilon)/(1 - \epsilon) \quad if \quad x \leq 1$$

and

$$t(x) = x + 1 \quad if \quad x \geq 1.$$

Clearly, the function t is a homeomorphism from $[\epsilon, \infty[$ onto $[1, \infty[$, with inverse

$$t^{-1}(y) = (2 - y)\epsilon + (y - 1) \quad if \quad 1 \leq y \leq 2$$

and

$$t^{-1}(y) = y - 1 \quad if \quad y \geq 2.$$

The function t defined in the above remark will be used in the following lemma.

Lemma 8. *Let $\epsilon < 1$ be a nonzero positive real number and let $I = [\epsilon, \infty[$. Let $F \in C^1(I, \mathbb{R}^+)$ be increasing and its derivative f be decreasing. Assume that $(a_n)_{n\in\mathbb{N}}$ is a sequence of nonnegative real numbers. Moreover, let $g : [1, \infty[\longrightarrow [0, \infty[$ be a continuous function such that for all $x \geq 1$ we have $g(x) = f(t^{-1}(x))$. Then*

1. *A necessary and sufficient condition for $\sum_{i=1}^{n} a_i = O(F(n) + 1)$ is that $a_n = O(g(n))$.*
2. *$lim_{n\to\infty} \sum_{i=1}^{n} a_i/n = 0$ if and only if $lim_{n\to\infty} a_n = 0$.*

Proof.

1. Assume that $\sum_1^n a_i = O(F(n) + 1)$. Then there exists a constant $c \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ greater than some n_0 we have $\sum_1^n a_i \leq c(F(n) + 1)$. Then,

$$a_{n+1} = \sum\nolimits_1^{n+1} a_i - \sum\nolimits_1^{n} a_i \leq c(F(n + 1) - F(n)) = cf(\theta_n).$$

Where, $n \leq \theta_n \leq n + 1$. Since the continuous function f is decreasing, $f(\theta_n) \leq f(n)$. Hence, $a_{n+1} \leq cf(n)$. Therefore, $a_n = O(g(n))$.

Conversely, In approximating the integral by sum we have

$$F(n) = f(\epsilon) + \sum\nolimits_{i=1}^{n-1} f(i) + O(1) = \sum\nolimits_{i=1}^{n} g(i) + O(1).$$

Therefore,

$$\begin{aligned} \textstyle\sum_{i=1}^{n} a_i &= O(\textstyle\sum_{i=1}^{n} g(i)) \\ &= O(F(n) + O(1)) = O(F(n) + 1) \end{aligned}$$

2. From $lim_{n\to\infty} a_n = 0$, it follows that for each $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, we have $a_n \leq \epsilon$. Therefore, for all large n we have $\sum_{i=1}^{n} a_i/n \leq 2\epsilon$. Since ϵ is arbitrary we have $lim_{n\to\infty}\sum_{i=1}^{n} a_n/n = 0$.

   Conversely, assume that $lim_{n\to\infty}\sum_{i=1}^{n} a_n/n = 0$. Therefore, for each $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have

$$\sum_{i=1}^{n} a_i/n \leq \epsilon \quad or \quad \sum_{i=1}^{n} a_i \leq n\epsilon.$$

   Hence,

$$a_n = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n-1} a_i \leq (n - (n-1))\epsilon = \epsilon.$$

   Therefore, $lim_{n\to\infty} a_n = 0$.

■

Theorem 7. *Let $\rho^* = ({\rho^*}^{(n)})_{n\in\mathbb{N}}$ and $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be Q-quantum sources over the Hilbert space $\mathbb{H}$. Then*

1. $lim_{n\to\infty} Risk_n(\rho^*, \bar{\rho}) =^w 0$ *if and only if* $lim_{n\to\infty}\overline{Risk}_n(\rho^*, \bar{\rho}) =^w 0$.

2. *Let ϵ be a nonzero positive real number and let $I = [\epsilon, \infty[$. Let $F \in C^1(I, \mathbb{R}^+)$ be increasing and its derivative f be decreasing.*

   Then, $Risk_n(\rho^*, \bar{\rho}) =^w O(g(n))$ if and only if

$$\overline{Risk}_n(\rho^*, \bar{\rho}) =^w O((F(n) + 1)/n). \tag{7.7}$$

   Where, the function g is defined in the above lemma.

Proof. The proof is a consequence of the definitions and the above lemma. See also [10].

■

Acknowledgment

The authors would like to express their very great appreciation to Prof. Fabio Benatti for his valuable and constructive suggestions. His willingness to give his time so generously has been very much appreciated. We would also like to thank Prof. Peter D. Grunewald for his very interesting book, “the MDL principle”, from which we learned the classical MDL principle.

References

[1] P. Alquier, K. Meziani and G. Peyre, Adaptive estimation of the density matrix in quantum homodyne tomography with noisy data, iopscience journal.

[2] C. Athanasios, Sobolev Inequalities on Complete Riemannian Manifolds, Master Thesis, University of Crete, 2012

[3] F. Benatti, Dynamics, Information and complexity in Quantum Systems, Springer 2009.

[4] M. Berger, P. Gauduchon, E. Mazet, Le Spectre d'une Varie'te' Riemanniane, Lecture notes in Mathematics 194, Spinger-Verlag, 1971

[5] Z. Brazil, J. Rehacek, J. Fiurasek, M. Jezek, Maximum likelihood Methods in Quantum Mechanics, in Quantum State Estimation, Lecture Notes in Physics, Vol. 649, Springer 2004.

[6] H.B. Coldenstrodt-Ronge, J.S. Lundeen, K.L. Pregnell, A. Feito, B.J. Smith,W. Mauerer, Ch. Silberhorn, J. Eisert, M.B. Plenio, and I.A. Walmsley1, A proposed testbed for detector tomography in Journal of Modern Optics 56(2-3), 2009

[7] G. M. D'Ariano, M. G. A. Paris, M. F. Sacchi, Quantum Tomographic Methods, in Quantum State Estimation, Lecture Notes in Physics, Vol. 649, Springer 2004.

[8] J. Dieudonne', E'le'ments d'Analyse, tome II and III, Editions Jacques Gabay, Paris 1983

[9] P. D. Grunewald, I. J. Myung, and M. A. Pitt (Eds.), Advances in minimum Description Length: Theory and Applications, MIT Press 2005.

[10] P. D. Grunewald, the Minimum Description Length Principle, The MIT Press 2007.

[11] Z. Hradil, Quantum state estimation, Physical review, 1998

[12] M. Hayashi, Asymptotic Theory of Quantum Statistical Inference, World Scientific 2005.

[13] Masahito Hayashi, Quantum Information, An Introduction, Springer-Verlag 2006.

[14] C. W. Helstrom, Quantum Detection and Estimation Theory, Academic Press, 1976.

[15] A. S. Holevo, Probabilistic and Statistical Aspects of Quantum Theory, North Holland, 1982.

[16] A. S. Holevo, Quantum Systems, Channels, Information, a Mathematical Introduction, De Gruyter, 2012.

[17] A. S. Holevo, Statistical Structure of Quantum Theory, Lecture Notes in Physics 67, Springer 2001.

[18] V. G. Ivancevich and T. T. Ivancevich, Applied Differential Geometry, A modern Introduction, World Scientific Co. 2007.

[19] D. Korkinof and Y. Demiris, Online Quantum Mixture Regression for Trajectory Learning by Demonstration International Conference on Intelligent Robots and Systems (2013), pp: 3222-3229 DOI: 10.1109/IROS.2013.6696814

[20] M. S. Leifer, Conditional Density Operators in Quantum Information, Banff 2007 and quant-ph/0611233.

[21] F. Liese, K. Miescke, Statistical Decision theory, Springer 2008.

[22] M. Lobino, D. Korystov, C. Kupchak, E. Figueroa, B. C. Sanders and A. I. Lvovsky, Complete Characterization of Quantum-Optical Processes, Science journal, 2008

[23] H. Nagaoka, On Fisher Information of Quantum Statistical Models, in Asymptotic Theory of Quantum Statistical Inference, World Scientific 2005.

[24] M. G. A. Paris, Quantum estimation and quantum technology, International Journal of Quantum Information, 2009

[25] J. Rissanen, Information and Complexity in Statistical Modeling, Springer-Verlag 2007.

[26] J. Rissanen, Model Selection and Testing by the MDL Principle, in Information Theory and Statistical Learning, 2009.

[27] J. Rissanen, Stochastic Complexity and Modeling, Annals Statistics 14, 1986.

[28] L. Schwartz, Analyse, tome III and IV, Hermann, Paris, 1993.