IOP Publishing J. Phys. Commun. 4 (2020) 025001 https://doi.org/10.1088/2399-6528/ab6fb3

Journal of Physics Communications

CrossMark

OPEN ACCESS

RECEIVED
15 October 2019

REVISED
11 January 2020

ACCEPTED FOR PUBLICATION
24 January 2020

PUBLISHED
4 February 2020

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Semi-Density Matrices and Quantum Statistical Inference

Ahmad Shafiei Deh Abad[1] and **Mohammad Shahbazi**

School of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, Iran

[1] Author to whom any correspondence should be addressed.

E-mail: shafiei@khayam.ut.ac.ir and mshahbazi@ut.ac.ir

Keywords: quantum statistical inference, quantum estimation, semi-density matrices, operator algebra

Abstract

In this paper, inspired by the 'Minimum Description Length Principle' in classical statistics, we introduce a new method for predicting the outcomes of performing quantum measurements and for estimating the state of quantum systems.

1. Introduction

Needless to say, nowadays nearly all our physical knowledge is based on quantum theory. So an increasingly important problem is to characterize quantum systems and to obtain information about them. In the way of solving the problem, Quantum Statistical Inference (QSI) is a unique tool. Quantum statistical inference is the quantum version of classical statistical inference. To be more precise, quantum statistical inference enables us to obtain information about quantum systems by using outcomes of performing quantum measurements. The research subject was initiated in the middle of the 1960s. The pioneers and the first researchers in the field are Holevo, Yuen, Kennedy, Belavkin, etc. Since then till now many researchers in different countries have conducted research into the subject and have extended it in different directions. Among other things, QSI contains the subject matters, quantum estimation and quantum prediction, which will be considered in this paper. To treat these problems the only tool at our disposal is performing measurements. Since quantum theory is statistical in nature, we have to perform the same quantum measurement in the same state of the quantum system many times. But, as it is well-known, after performing a measurement on a quantum system the state of the system changes drastically. To overcome the difficulty, we usually assume that there are n quantum systems described by the same Hilbert space $\mathbb{H}$ and prepared independently and identically in the same state ρ (a density matrix on $\mathbb{H}$) and we perform the same quantum measurement on each of them. In this way, we obtain a data set $D = (x_1, x_2, \ldots x_n)$. By quantum estimation we mean techniques enabling us to find an approximation of the state ρ with the help of the data set D and by prediction we mean characterizing the probability of the outcome x_{n+1} given the previous outcomes $x \in D$. An appropriate method to solve the problems is to choose a set $\mathcal{M}$ of density matrices on $\mathbb{H}$ containing ρ, called a quantum model and try to find the state ρ by methods, such as Maximum Likelihood Estimation (MLE). To be able to act in this way, we have to parameterize the set $\mathcal{M}$ in a differentiable manner. Unfortunately, ML Estimation which has been used by several authors gives rise to overfitting[2]. Moreover, in general, we do not know whether the state ρ is in the model $\mathcal{M}$ or not. Inspired by the works of J. Rissanen [1, 2], P. Grünewald [3], and others on the Minimum Description Length Principle (MDL) in classical statistics, one of our goals in this paper is to remedy this difficulty. Their works on the use of 2-part codes [3] in MDL guided us to use sets of semi-density matrices in addition to quantum models and call them generalized quantum models (for more detail see the beginning of section 3). As in classical MDL we base our work on universal sources associated with quantum models. We will show that in all interesting cases universal quantum sources exist. It will be evident that the use of universal sources automatically protects against overfitting. Moreover, we prove different versions of the consistency theorem showing that when the state ρ is in the chosen model $\mathcal{M}$, the selected universal quantum source is asymptotically equivalent to it.

The organization of the paper is as follows:

[2] The selection of an overly complex model that, while fitting observed data very well, predicts future data very badly.

© 2020 The Author(s). Published by IOP Publishing Ltd

In section 2 we introduce the notion of Q-projection which in this work will act as projective quantum measurement. At the end of section 2, it is proved that in quantum theory all results concerning prediction and estimation proved in this paper are true for general quantum measurements. In section 3 after some explanations about the MDL principle and the way we have gone through to quantize the most important notions involved in MDL, we will define fundamental concepts, such as (generalized) quantum models, universal quantum sources, which is the core concept of this work, quantum source and quantum strategy. we will also prove some important facts about them. At the end of the same section, we introduce the notion of good quantum estimator and a large class of them. Section 4 is about quantum prediction and quantum estimation. In section 5 we will introduce the notion of consistency and prove some theorems about it. In section 6, we give examples that indicates the efficiency of this method.

We emphasize that with the help of trace function, one can reduce the problems treated here to problems in the classical MDL methods and solve them classically. But in doing this the operator nature of important concepts like universal quantum source associated with quantum models, quantum strategy and conditional density matrix conditioned on density matrix will be lost. Even worse, one cannot understand that these concepts are operators. Moreover, treating the problems in the realm of operator theory are more natural and simpler. In the same vein, nearly all notations, definitions and conventions used in the paper is directly inspired by their classical counterparts in [3]. So that comparison of classical and quantum frameworks should be straightforward.

2. Q-Projection

Given a separable Hilbert space $\mathbb{H}$, in general infinite dimensional, with inner product $\langle\cdot|\cdot\rangle$, the set $\{|k\rangle | k \in \mathbb{N}\}$ will denote an orthonormal basis of $\mathbb{H}$ and its dual basis will be denoted by the set $\{\langle k| | k \in \mathbb{N}\}$. The set of all bounded operators (resp. self-adjoint bounded operators) on $\mathbb{H}$ will be denoted by $B(\mathbb{H})$ (resp. by $B_H(\mathbb{H})$) and the set of all positive operators (resp. density matrices) on $\mathbb{H}$ will be denoted by $B_+(\mathbb{H})$ (resp. by $D(\mathbb{H})$). Finally, the Hilbert space generated by trace class operators of $\mathbb{H}$ with the following inner product will be denoted by $B_T(\mathbb{H})$,

$$\langle T|S\rangle_T = Tr(T^*S) \text{ for all } T, S \in B_T(\mathbb{H})$$

with associated norm $\|T\|_T = \sqrt{\mathrm{Tr}(T^*T)}$.

A positive operator T is called a semi-density matrix if $Tr(T) \leqslant 1$ and it is called a density matrix if $Tr(T) = 1$. The mapping which sends each nonzero semi-density matrix T to its associated density matrix $\frac{T}{Tr(T)}$ will be denoted by ω.The collection of all complete sets of mutually orthogonal (minimal) projections $P = \{p_1, p_2, ...\}$ on $\mathbb{H}$, with $\sum_{n=1} p_n = 1$ (completeness), will be denoted by $\pi(\mathbb{H})$ ($\pi_0(\mathbb{H})$).

Let $P = \{p_1, p_2, ...\}$ and $Q = \{q_1, q_2, ...\}$ be elements of $\pi(\mathbb{H})$.Then, the set $\{p_i q_j | i, j \in \mathbb{N}\} - \{0\}$ will be denoted by PQ and will be called the combination of P and Q. We say that P and Q commute if $PQ = QP$. In this case clearly $PQ \in \pi(\mathbb{H})$. More generally, a subset $\mathbb{S}$ of $\pi(\mathbb{H})$ is called commutative, if any two elements of it commute. Let $\mathbb{P} = \{P_1, P_2, ..., P_k\}$ be a finite subset of $\pi(\mathbb{H})$. The combination of elements of $\mathbb{P}$ is

$$\Pi_{i=1}^{k} P_i = \{\Pi_{i=1}^{k} x_i | x_i \in P_i,\ i = 1, 2, ..., k\} - \{0\}.$$

When $\mathbb{P}$ is commutative, $\Pi_{i=1}^{k} P_i \in \pi(\mathbb{H})$.

Definition 1. Assume that $(X_j)_{j\in J}$ is a family of subsets of a nonempty set X and for each $j \in J$, there exists $I_j \subset J$ such that $X - X_j = \cup_{i\in I_j} X_i$. Then,

1. For each $I \subset J$, we say that the set $Y = \cup_{i\in I} X_i$ is a maximally connected union of the family $(X_j)_{j\in J}$ if it satisfies the following conditions:

   (a) For each proper subset K of I,

   $$\bigcup_{k\in K} X_k \cap \left(\bigcup_{i\in I-K} X_i\right) \neq \{\}.$$

   (b)

   $$Y \cap \left(\bigcup_{i\in J-I} X_i\right) = \{\}.$$

   The set of all maximally connected unions of the family $(X_j)_{j\in J}$ will be denoted by $\wedge_{j\in J} X_j$. Clearly, $\wedge_{j\in J} X_j$ is a partition of X.

2. For each $I \subset J$, the subset $Z = \cap_{i\in I} X_i$ of X will be called a minimally connected intersection of the family $(X_j)_{j\in J}$, if

$$Z \cap (\bigcup_{j \in J-I} X_j) = \{\}.$$

The set of all minimally connected intersections of the family $(X_j)_{j \in J}$ is evidently a partition of X and will be denoted by $\vee_{j \in J} X_j$.

Now assume that X is an arbitrary non-empty set. Let the set of all partitions of X be denoted by $\mathfrak{P}(X)$. Let $\underline{P}$ and $\underline{Q}$ be in $\mathfrak{P}(X)$. We say that $\underline{Q}$ is finer than $\underline{P}$ and we write $\underline{P} \preceq \underline{P}$, if each elements of $\underline{P}$ is the union of some elements of $\underline{Q}$. It is evident that the set $\mathfrak{P}(X)$ with the order relation $\underline{P} \preceq \underline{Q}$ is a partially ordered set. Assume that $\mathbb{P} = \{\underline{P}_k | k \in K\} \subseteq \mathfrak{P}(X)$ is a set of partitions of the set X. Let $\cup_{k \in K} \underline{P}_k = \{X_j | j \in J\}$. Clearly $X = \cup_{j \in J} X_j$ and the family $(X_j)_{j \in J}$ of subsets of the set X satisfies the conditions of definition 1. It is easy to see that for each partition $\underline{P}_k \in \mathbb{P}$ we have

$$\wedge_{j \in J} X_j \preceq \underline{P}_k \preceq \vee_{j \in J} X_j.$$

Let partitions $\underline{P}$ and $\underline{Q}$ of the set X be such that for all $k \in K$ we have $\underline{Q} \preceq \underline{P}_k \preceq \underline{P}$. Then, it is straightforward to see that for each $k \in K$

$$\underline{Q} \preceq \wedge_{j \in J} X_j \preceq \underline{P}_k \preceq \vee_{j \in J} X_j \preceq \underline{P}.$$

Therefore, $\wedge_{j \in J} X_j$ (resp. $\vee_{j \in J} X_j$) is the greatest lower bound (resp. the least upper bound) of the partially ordered set $\mathbb{P}$ and will be denoted by $\wedge_{k \in K} \underline{P}_k$ (resp. $\vee_{k \in K} \underline{P}_K$).

Definition 2. Let P and Q be in $\pi(\mathbb{H})$. We say that P is finer than Q, and we write $Q \preceq P$ if $PQ = P$. In this case $QP = (PQ)^\star = P$.

We say that Q and P are consistent if they have a common upper bound with respect to this order relation. More generally, a subset $A \in \pi(\mathbb{H})$ is called consistent if it has an upper bound. Then clearly any subset of A is also consistent. we say that a consistent set A is maximally consistent if there is no consistent set B such that $A \underset{\neq}{\subset} B$.

Lemma 1. *Let* $\mathbb{P} = \{P_k \in \pi(\mathbb{H}) | k \in K\}$. *Then*

1. *If the set is consistent it has a least upper bound and a greatest lower bound.*
2. *If the set is finite and commutative, then it is consistent.*

Proof.

1. Assume that the set $\mathbb{P}$ is consistent then it has an upper bound $R = \{r_1, r_2, \cdots\}$ which is a complete set of mutually orthogonal projections of $\mathbb{H}$. Let $Q \in \mathbb{P}$. By definition $Q \preceq R$. Let $q \in Q$ and let R_q be the sum of all elements $r \in R$ such that $q \geqslant r$. i.e. $qr = r$. Clearly $R_q^2 = R_q \neq 0$ and $qR_q = R_q q = R_q$, since $rq{=}r$ for all $r \in R_q$. Therefore $q - R_q$ is a projection and if $q - R_q = q(I_{\mathbb{H}} - R_q) \neq 0$, then there exists $r \in R$ such that $q \geqslant r$ and $rR_q = 0$ which is a contradiction. Hence, $q = R_q$. Therefore for each $Q \in \mathbb{P}$, each $q \in Q$ is the sum of some elements of R.
   Let the order preserving mapping $Q \to \underline{Q}$ from $\pi(\mathbb{H})$ into $\mathfrak{P}(R)$ be defined as follows, for each $q \in Q$, $q \to \underline{q}$, where $\underline{q}$ is the set of all summands of the projection q. Notice that q is the sum of some elements of R. Now it is clear that under this mapping we have the following bijective maps.
   $$\wedge_{k \in K} P_k \to \wedge_{k \in K} \underline{P}_k$$
   $$\vee_{k \in K} P_k \to \vee_{k \in K} \underline{P}_k$$
   we have seen above that $\vee_{k \in K} \underline{P}_k$ (resp. $\wedge_{k \in K} \underline{P}_k$) is the least upper bound (resp. the greatest lower bound) of the set $\{\underline{P}_k, k \in K\}$. Therefore, $\wedge_{k \in K} P_k$(resp.$\vee_{k \in K} P_k$) is the greatest lower bound (rep. the least upper bound) of $\mathbb{P}$.
2. Assume that the set $\mathbb{P}$ is finite and commutative. Then, $\vee_{k \in K} P_k = \Pi_{k \in K} P_k$. Therefore, $\mathbb{P}$ is consistent.

■

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

$$T_Q = \sum_n q_n T q_n,$$

will be called the Q-projection of T (see also [4]). The set of all Q-projections of elements of $B(\mathbb{H})$ will be denoted by $B_Q(\mathbb{H})$ and for each $q \in Q$, $B_q(\mathbb{H}) = \{qTq | T \in B(\mathbb{H})\}$.

The set $B_Q(\mathbb{H})$ is a complex subspace of the $C^* - algebra B(\mathbb{H})$, and the mapping $\bar{Q}$ from $B(\mathbb{H})$ into $B_Q(\mathbb{H})$ defined by $\bar{Q}(T) := T_Q$ is a projection. For T and S in $B(\mathbb{H})$ and $Q \in \pi(\mathbb{H})$ we have $(T_Q S_Q)_Q = T_Q S_Q$. Therefore $B_Q(\mathbb{H})$ is a unital C^*-subalgebra of $B(\mathbb{H})$. If $Q \in \pi_0(\mathbb{H})$ then evidently $B_Q(\mathbb{H})$ is commutative.

Let $\mathbb{H}$ and $\mathbb{H}'$ be Hilbert spaces. Let $P = \{p_1, p_2, \ldots,\}$ and $Q = \{q_1, q_2, \ldots,\}$ be complete sets of mutually orthogonal projections of the Hilbert spaces $\mathbb{H}$ and $\mathbb{H}'$. Then:

$$P \otimes Q = \{p_i \otimes q_j,\ i, j \in \mathbb{N}\}$$

is a complete set of mutually orthogonal projections on $\mathbb{H} \otimes \mathbb{H}'$. Let T (resp. S) be a bounded operator on $\mathbb{H}$ (resp. $\mathbb{H}'$). Then:

$T_P \otimes S_Q = \sum_{n,m} p_n T p_n \otimes q_m S q_m = \sum_{n,m} (p_n \otimes q_m)(T \otimes S)(p_n \otimes q_m) = (T \otimes S)_{P\otimes Q}$.

Lemma 2.

1. *The mapping $\bar{Q}$ is trace preserving.*
2. *If T is self-adjoint, then T_Q is also self-adjoint.*
3. *A necessary and sufficient condition for T to be positive is that for each $Q \in \pi(\mathbb{H})$, T_Q be positive.*
4. *Let $Q \in \pi_0(\mathbb{H})$ and $T \in B(\mathbb{H})$ be arbitrary. Then, T_Q is always normal.*

Proof.

1. $\mathrm{Tr}(T_Q) = \sum_{n=1}^{\infty} \mathrm{Tr}(q_n T q_n) = \sum_{n=1}^{\infty} \mathrm{Tr}(q_n T) = Tr(T)$, since the sets of projections $Q \in \pi(H)$ are complete.
2. If $T = T^*$, then evidently $(T_Q)^* = T_Q$.
3. Let $T \geqslant 0$; then, for each $q \in Q$, $qTq \geqslant 0$. So that for each $Q \in \pi(\mathbb{H})$, $T_Q \geqslant 0$. Vice versa, if $T_Q \geqslant 0$ for each $Q \in \pi(\mathbb{H})$, then, for each vector $|v\rangle \in \mathbb{H}$, $Tr(|v\rangle\langle v|T) = \langle v|T|v\rangle \geqslant 0$, since any such $|v\rangle\langle v|$ belongs to some $Q \in \pi(\mathbb{H})$. Therefore, $T \geqslant 0$.
4. Since in this case $B_Q(\mathbb{H})$ is a commutative algebra, the proof is clear.

■

Corollary 1. *The restriction of the mapping $\bar{Q}$ to $D(\mathbb{H})$ is a convex map from $D(\mathbb{H})$ onto $D_Q(\mathbb{H})$.*

Lemma 3.

1. *The mapping $\bar{Q} : B(\mathbb{H}) \longrightarrow B_Q(\mathbb{H})$ is continuous.*
2. *The mapping $\bar{Q} : B_T(\mathbb{H}) \longrightarrow B_Q(\mathbb{H})$ is continuous in the trace norm topology.*

Proof.

1. Let $T \in B(\mathbb{H})$ be a self-adjoint element of $B(\mathbb{H})$. Then, $\|T_Q\|$ is equal to its spectral radius r. Let $q \in Q$ and let $\|qTq\| = r$. Then

$$\|T_Q\| = \|\sum_n q_n T q_n\| = \|qTq\| \leqslant \|T\|$$

Since any $T \in B(\mathbb{H})$ can be written as a combination of two self adjoint elements $\bar{Q}$ is continuous.

2. Let $T \in B_T(\mathbb{H})$. Then, for each $q \in Q$, $qTqqT^*q \leqslant qTT^*q$. Therefore, $(T_Q)^* T_Q = ((T^*)_Q) T_Q \leqslant (T^*T)_Q$. Since $\mathrm{Tr}(T_Q) = \mathrm{Tr}(T)$,

$$\|T_Q\|_T = (Tr(T_Q{}^\star T_Q))^{1/2} \leqslant (Tr((T^\star T)_Q))^{1/2} = (Tr(T^\star T))^{1/2} = \|T\|_T.$$

■

Lemma 4. *For each element $T \in B(\mathbb{H})$ and each $Q = \{q_1, q_2 \ldots\} \in \pi(\mathbb{H})$ we have:*

1. *$T = T_Q$ if and only if for each $q \in Q$ we have $qT = Tq$.*

2. *Let $S = S_Q$ and for all $q \in Q$, $qSq = qTq$. Then, $S = T_Q$.*
3. *Let T be a normal operator and f be a continuous function defined on a neighborhood of the spectrum of T. If $T = T_Q$ then $f(T) = (f(T))_Q$.*

Proof.

1. Assume that $T = T_Q = \sum_n q_n T q_n$. Then, for each $q_n \in Q$ we have
$$q_n T = q_n T_Q = q_n T q_n = T_Q q_n = T q_n.$$
Conversely, if for each $q_n \in Q$, $T q_n = q_n T$, then, completeness of Q yields
$$T_Q = \sum_n q_n T q_n = \sum_n q_n T = T.$$

2. By hypothesis, $S = S_Q = \sum_n q_n S q_n = \sum_n q_n T q_n = T_Q$.
3. The proof is a consequence of point 1 and of functional calculus.

■

Lemma 5. *Let $T \in B(\mathbb{H})$ and $P, Q \in \pi(\mathbb{H})$. If $P \succcurlyeq Q$ then:*

1. $T_P = (T_Q)_P = (T_P)_Q$.
2. $Ker(\bar{Q}) \subset Ker(\bar{P})$

Proof. It is clear that for each element $p \in P$ there exists exactly one element $q_0 \in Q$ such that $q_0 p = p q_0 = p$ and for other elements $q \in Q$ we have $qp = pq = 0$. So
$$p(T_Q)p = p(\sum_{q \in Q} qTq)p = pq_0 T q_0 p = pTp$$

Therefore,
$$(T_Q)_P = \sum_{p \in P} pT_Q p = \sum_{p \in P} pTp = T_P$$

On the other hand for each $q \in Q$ and each $p \in P$ we have
$$\begin{aligned} qT_P &= q\sum_{p \in P} pTp = \sum_{p \in P} qpTp \\ &= \sum_{p \in P | qp \neq 0} pTp = \sum_{p \in P} pTpq \\ &= T_P q. \end{aligned} \tag{1}$$

Therefore $T_P = (T_Q)_P = (T_P)_Q$.

Since $\bar{P}(T) = T_P = (T_Q)_P = \bar{P}(\bar{Q}(T))$, the proof of the second part is clear.

Lemma 6. *Let $T = T_Q$ be an invertible element of $B(\mathbb{H})$. Then $T^{-1} = (T^{-1})_Q$.*

Proof. From lemma 3 and the fact that $qT = Tq$ implies $q = TqT^{-1}$, it follows that $T^{-1}q = qT^{-1}$. ■

Let $T = T_Q$ be a normal operator. Then T_Q is called a *pseudo-spectral decomposition* of T. Clearly, for each $q \in Q$, $q(\mathbb{H})$ is invariant under T.

Lemma 7. *Assume that* $\mathrm{T_P}$ *is a pseudo-spectral decomposition of the operator* T. *Then for each $S \in B(\mathbb{H})$, we have*
$$(ST)_P = S_P T_P \quad \text{and} \quad (TS)_P = T_P S_P \quad Tr(TS) = Tr(T_P S_P).$$

Proof. We have $(ST)_P = (ST_P)_P$. Therefore, for each $p \in P$ we have
$p(ST)p = p(ST_P)p = pSpTp = (pSp)(pTp)$. Therefore, $(ST)_P = S_P T_P$. The proof of the second equality is the same. The third equality is evident. ■

The previous lemmas lead to the following result.

Theorem 1. *Let Q be in $\pi(\mathbb{H})$. Then*

1. *$B_Q(\mathbb{H})$ is a unital C^*-algebra.*
2. *$B(\mathbb{H})$ is a left and a right $B_Q(\mathbb{H})$-module.*
3. *The mapping $\bar{Q}$ from $B(\mathbb{H})$ into $B_Q(\mathbb{H})$ is a $B_Q(\mathbb{H})$-linear form.*
4. *A necessary and sufficient condition for $B_Q(\mathbb{H})$ to be commutative is that Q be a complete set of mutually orthogonal minimal projections.*

Let S and T be in $B(\mathbb{H})$. Then, in general $ST \neq TS$. But for all $Q \in \pi_0(\mathbb{H})$, $S_Q T_Q = T_Q S_Q$. This fact motivate the following definition.

Definition 4. Let R be an n-ary relation on $B(\mathbb{H})$. We say that R is *weakly true* if, for each $Q \in \pi_0(\mathbb{H})$, $\bar{Q}^n(R)$ is true, where $\bar{Q}^n(R)$ is the image of R under $\bar{Q}^n(R)$, the natural extension of $\bar{Q}$: $B(\mathbb{H}) \to B_Q(\mathbb{H})$ to $\bar{Q}^n$: $(B(\mathbb{H}))^n \to (B_Q(\mathbb{H}))^n$

Remark 1. Any two elements of $B(\mathbb{H})$ always weakly commute. For some relations, being true or weakly true are equivalent. For example, if $T \geqslant S$ then clearly, this relation is weakly true.

Conversely, Assume that for each $Q \in \pi_0(\mathbb{H})$, $T_Q \geqslant S_Q$ therefore for each minimal projection q, $qTq \geqslant qSq$. Since for each vector $v \in \mathbb{H}$ the projection $|v><v|$ is contained in some $Q \in \pi_0(\mathbb{H})$ we have $\langle v|T - S|v\rangle \geqslant 0$. Therefore, $T - S \geqslant 0$.

The relation weakly equal will be denoted by $=^w$.

Let $\rho \in D(\mathbb{H})$ be a diagonal matrix. Clearly, we can consider ρ as a classical probability distribution function. But if the density matrix ρ is not diagonal we cannot interpret it in this way. The following definition serves to discriminate these two cases.

Definition 5. Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. The mapping $\nu : B(\mathbb{H}) \longrightarrow \mathbb{R}$ given by $\nu(T) = \|T - T_Q\|$ will be called Q-quantum complexity of T. When $\nu(T) = 0$, T is called Q-classical and when $T_Q = 0$, T will be called Q-maximally non-classical.

Example 1. Let $\mathbb{H}$ be a 2-dimensional Hilbert space with the standard basis $\{|0\rangle, |1\rangle\}$. Let X, Y, Z be Paoli density matrices on $\mathbb{H}$ and $Q = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. Then, it is clear that Z is Q-classical and X and Y are Q-maximally non-classical.

Important Remark. Let $Q = \{q_1, q_2, \ldots, q_n, \ldots\} \in \pi(\mathbb{H})$ and let ρ be a density matrix on $\mathbb{H}$. Assume that we perform the quantum measurement described by the set Q of measurement operators on the quantum system with state space $\mathbb{H}$ in the state ρ. Then, as it is well-known the probability of outcome associated with q_i is $Tr(q_i \rho q_i)$. Now, assume that $Q \preceq P$. Then as we have seen earlier q_i can be written as sum of some elements of P. say, $q_i = \sum_j p_j$. Then

$$Tr(q_i \rho q_i) = Tr(q_i \rho) = Tr(\sum_j p_j \rho) = Tr(\sum_j p_j \rho p_j) = \sum_j Tr(p_j \rho p_j).$$

In this work in many cases we use only $Q \in \pi_0(\mathbb{H})$. Nevertheless, interpreted in quantum theory, as is evident from the above fact, our results concerning prediction and estimation will be true for all $Q \in \pi(\mathbb{H})$. Moreover, as it is well known [5] the outcomes of a general measurement on the quantum system represented by the Hilbert space $\mathbb{H}$ can be realized by a projective measurement on the tensor product of $\mathbb{H}$ and another Hilbert space $\mathbb{H}_0$. So, our results will be true for general quantum measurement systems.

3. Quantum model, quantum source and quantum strategy

As we said in the introduction our work in this paper inspired by the Minimum Description Length Principle is based on universal quantum sources associated with quantum models. In this part, we define several versions of universal quantum sources associated with a quantum model and investigate some of their properties. In the same section, we prove the existence of universal quantum sources and give a constructive way to build it. We also define quantum strategy and treat its relation to universal quantum sources.

Before going further in this section let us give some comments on the use of semi-density matrices and on our definition of universal quantum sources.

The minimum description length principle is a powerful tool in statistical (inductive) inference. It is essentially based on two important notions:

3.1. 2-part coding

The estimation by 2-part code can be considered as a mathematical formulation of Occam's Razer which says that between different descriptions of a data set, the simpler is the better. Assume that these descriptions are encoded in such a way that they reflect their complexities. Then the description with the shortest code-length is the better.

More precisely, let $\mathcal{M}$ be a nonempty set of probability density (mass) functions on a set $\mathcal{X}$ and let $D \subset \mathcal{X}^n$ be an i.i.d data set generated by $p^\star \in \mathcal{M}$. Assume that elements of $\mathcal{M}$ are encoded. For each $p \in \mathcal{M}$, the length of its associated code-word will be denoted by $L(p)$ and $-\log_2 p(D)$ will be denoted by $L(D|p)$. Let

$$\ddot{p} = \text{argmin}_{p\in\mathcal{M}}(L(p) + L(D|p)).$$

Clearly for each $p \in \mathcal{M}$, $L(p) + L(D|p)$ is the length of an encoded description of the data set D and $\ddot{p}$ is chosen according to Occam's Razer.

3.2. universal coding

Under above assumptions on $\mathcal{M}$ and $\mathcal{X}$, assume that for each $n \in \mathbb{N}$, $\bar{p}^{(n)}$ is a probability density (mass) function on $\mathcal{X}^n$. The sequence $\bar{p} = (\bar{p}^{(n)})_{n\in\mathbb{N}}$ of probability density (mass) functions will be called universal with respect to $\mathcal{M}$, if for each $\epsilon > 0$, each $p \in \mathcal{M}$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ and all $x^{(n)} \in \mathcal{X}^n$ we have

$$-\log_2\bar{p}^{(n)}(x^{(n)}) - (-\log_2 p^{(n)}(x^{(n)})) \leqslant n\epsilon.$$

For more details see [3].

Now let us explain briefly the way we have gone through to quantize these two notions.

Let the Hilbert space $\mathbb{H}$ be the state space of a quantum system A, which is prepared in an unknown state ρ_0, a density matrix on $\mathbb{H}$, and let $Q = \{q_m | m \in O\} \in \pi_0(\mathbb{H})$ where O is the set of outcomes, be a projective quantum measurement system. Assume that $\mathcal{M}$ is a nonempty set of density matrices on $\mathbb{H}$ and $D \in O^n$ is the set of outcomes of performing the$Q$-measurement on n quantum systems identical to A and prepared in the same state ρ_0. In performing the$Q$-measurement on the quantum system A in an arbitrary state ρ the probability of outcome m is

$$\mathbb{P}(m) = Tr(q_m\rho) = Tr(q_m\rho q_m)$$

3.3. 2-part coding ⟶ semi-density matrix

Let elements of $\mathcal{M}$ be somehow encoded and for each $\rho \in \mathcal{M}$ let $L(\rho)$ be the length of the code-word associated with ρ and let $L(D|\rho) = -\log_2 Tr(\otimes^{m\in D}q_m\rho q_m)$. Then for each $\rho \in \mathcal{M}$ we have

$$\begin{aligned} L(\rho) + L(D|\rho) &= -\log_2 2^{-L(\rho)} - \log_2 Tr(\otimes^{m\in D}q_m\rho q_m) \\ &= -\log_2(2^{-L(\rho)}Tr(\otimes^{m\in D}q_m\rho q_m)) = -\log_2(Tr(2^{-L(\rho)}\otimes^{m\in D}q_m\rho q_m)) \\ &= -\log_2 Tr(\otimes^{m\in D}q_m(2^{-L(\rho)}\rho^{(n)})\otimes^{m\in D}q_m) \end{aligned}$$

But the function $\log_2$ is increasing and $Tr(\otimes^{m\in D}q_m(2^{-L(\rho)}\rho^{(n)})\otimes^{m\in D}q_m)$ is also increasing with respect to the semi density matrices $2^{-L(\rho)}\rho^{(n)}$, as in the above classical case

$$\ddot{\rho} = \text{argmin}_{\rho\in\mathcal{M}}L(\rho) + L(D|\rho) = \text{argmax}_{\rho\in\mathcal{M}}(\otimes^{m\in D}q_m)(2^{-L(\rho)}\rho^{(n)})(\otimes^{m\in D}q_m)$$

is an estimation of ρ_0 according to Occam's Razer. Notice that $(2^{-L(\rho)}\rho^{(n)})$ is a semi-density matrix.

3.4. Universal coding ⟶ universalDensity Matrix

Let $\bar{\rho}^{(n)}$ and $\bar{\rho}'^{(n)}$ be two density matrix on $\mathbb{H}^{(n)}$. Assume that as in classical case for $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$ and for all $q^{(n)} \in Q^{(n)}$ we have

$$-\log_2 Tr(q^{(n)}\bar{\rho}^{(n)}q^{(n)}) - (-\log_2 Tr(q^{(n)}\bar{\rho}'^{(n)}q^{(n)})) \leqslant n\epsilon.$$

From the above inequality we have
$\log_2 Tr(q^{(n)}\bar{\rho}^{(n)}q^{(n)}) \geqslant \log_2 2^{-n\epsilon} + \log_2 Tr(q^{(n)}\bar{\rho}'^{(n)}q^{(n)}) = \log_2 Tr(q^{(n)}(2^{-n\epsilon}\bar{\rho}'^{(n)})q^{(n)})$ But the inequality

$$\log_2 Tr(q^{(n)}\bar{\rho}^{(n)}q^{(n)}) \geqslant \log_2 Tr(q^{(n)}(2^{-n\epsilon}\bar{\rho}'^{(n)})q^{(n)}).$$

is equivalent to

$$q^{(n)}(\bar{\rho}^{(n)} - 2^{-n\epsilon}\bar{\rho}'^{(n)})q^{(n)} \geqslant 0.$$

In the following all tensor products of Hilbert spaces are topological tensor products.

The n-times tensor product of a Hilbert space $\mathbb{H}$ with itself will be denoted by $\mathbb{H}^{(n)}$ and in general, for each $T \in B(\mathbb{H})$, $T^{\otimes n} := \bigotimes^n T$. The sequence $(\mathbb{H}^{(n)})_{n\in\mathbb{N}}$ of Hilbert spaces will be denoted by $\mathbb{H}^\star$ and for $T_{(n)} \in \mathbb{H}^{(n)}$

the sequence $(T_{(n)})_{n\in\mathbb{N}}$ will be denoted by $T^{\otimes}$. In this case we say that $T^{\otimes}$ is an operator on $\mathbb{H}^{\star}$ and if for all $n\in\mathbb{N}$, $T_{(n)}$ is a (semi-)density matrix, then $T^{\otimes}$ will be called a (semi-)density matrix on $\mathbb{H}^{\star}$. A semi-density matrix $T^{\otimes}=(T_{(n)})_{n\in\mathbb{N}}$ on $\mathbb{H}^{\star}$ is called nonzero if for all $n\in\mathbb{N}$, $T_{(n)}\neq 0$. In this case the associated density matrix of $T^{\otimes}$ is $\omega(T^{\otimes})=\left(\frac{T_{(n)}}{Tr(T_{(n)})}\right)_{n\in\mathbb{N}}$. From now on semi-density matrices on $\mathbb{H}^{\star}$ will be denoted by $\bar{\rho}=(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$.The semi-density matrix $\bar{\rho}=(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ will be called

1. simple if $\bar{\rho}^{(1)}=\rho$ and for each $n\in\mathbb{N}$, $\bar{\rho}^{(n)}$ is the tensor product of ρ and $(n-1)$-times tensor product of $\omega(\rho)$.
2. a generalized quantum source if for each $1<n\in\mathbb{N}$, $\bar{\rho}^{(n-1)}=Tr_n(\bar{\rho}^{(n)})$.
3. regular if for each n, $\bar{\rho}^{(n)}$ is invertible

When for each $n\in\mathbb{N}$, $Tr(\bar{\rho}^{(n)})=1$, the generalized quantum source $\bar{\rho}=(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ will be called a quantum source.

Let $Q=\{q_1,q_2,\dots,q_n,\dots\}\in\pi(\mathbb{H})$ be a complete set of mutually orthogonal projections of the Hilbert space $\mathbb{H}$ and let $I=(i_1,i_2,\dots,i_n)\in\mathbb{N}^{(n)}$. Then the projection $q_{i_1}\otimes q_{i_2}\otimes\dots\otimes q_{i_n}$ will be denoted by $q_I^{(n)}$ or simply by $q^{(n)}$ if there is no ambiguity. The set $\{q_I^{(n)}|I\in\mathbb{N}^{(n)}\}$ will be denoted by $Q^{(n)}$.

Definition 6. Let $\underline{\mathcal{M}}=(M,\Sigma,\mu)$ be a measure spacewhere M is a generalized quantum model. Then, M will be called Bayesian if $\int_{\mathcal{M}}\rho d\mu(\rho)$ exists and is a density matrix.

Lemma 8. *Let $\underline{\mathcal{M}}$ be a Bayesian generalized quantum model which is a measure space 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 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 9. *Let $U\in B(\mathbb{H})$ be a unitary operator and $\bar{\rho}$ be a quantum source. Then $U\bar{\rho}U^{\dagger}=(U^{(n)}\bar{\rho}^{(n)}(U^{\dagger})^{(n)})_{n\in\mathbb{N}}$ is also a 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 $\bar{\rho}$ is a quantum source we have

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

So,

$$\begin{aligned}(U\bar{\rho}U^{\dagger})^{(n+1)}&=U^{(n+1)}\bar{\rho}^{(n+1)}(U^{\dagger})^{(n+1)}\\&=\textstyle\sum_{i,j=1}^{\infty}(U^{(n)}R_{i,j}(U^{\dagger})^{(n)})\otimes U|i\rangle\langle j|U^{\dagger}.\end{aligned}\tag{2}$$

Therefore,

$$\begin{aligned}Tr_{n+1}(U\bar{\rho}U^{\dagger})^{(n+1)}&=\textstyle\sum_{i,j=1}^{\infty}(U^{(n)}R_{i,j}(U^{\dagger})^{(n)})Tr(U|i\rangle\langle j|U^{\dagger})\\&=\textstyle\sum_{i=1}^{\infty}(U^{(n)}R_{i,i}(U^{\dagger})^{(n)})\\&=U^{(n)}(\textstyle\sum_{i=1}^{\infty}R_{i,i})(U^{\dagger})^{(n)}\\&=U^{(n)}\bar{\rho}^{(n)}(U^{\dagger})^{(n)}=(U\bar{\rho}U^{\dagger})^{(n)}\end{aligned}\tag{3}$$

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

In this work *ln* denotes natural logarithm and *log* denotes logarithm in base 2.

Definition 7. Let ρ and ρ' be density matrices. Then the quantum relative entropy of ρ and ρ' is

$$S(\rho\|\rho')=tr(\rho\,log\rho)-tr(\rho\,log\rho')$$

Definition 8. Let $\mathcal{M}$ be a quantum model and $\bar{\rho}=(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be a semi-density matrix on $\mathbb{H}^{\star}$. Let $Q\in\pi(\mathbb{H})$. We say that $\bar{\rho}$ is

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

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

2. universal in the expected sense relative to $\mathcal{M}$ if:

$$S(\rho^{(n)} \| \bar{\rho}^{(n)}) \leqslant n\epsilon.$$

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

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

4. Q-universal relative to $\mathcal{M}$ in the expected sense if

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

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

Lemma 10. *With the above notations and conventions, 1 implies 2 and 3.*

Proof. Clearly we have

$$\begin{aligned} &\bar{\rho}^{(n)} - 2^{-n\epsilon}\rho^{(n)} \geqslant 0 \\ &\quad \Rightarrow n\epsilon + log\bar{\rho}^{(n)} - log\rho^{(n)} \geqslant 0 \\ &\quad \Rightarrow n\epsilon\rho^{(n)} + (\rho^{(n)})^{1/2}(log\bar{\rho}^{(n)} - log\rho^{(n)})(\rho^{(n)})^{1/2} \geqslant 0 \\ &\quad \Rightarrow Tr(n\epsilon\rho^{(n)} + (\rho^{(n)})^{1/2}(log\bar{\rho}^{(n)} - log\rho^{(n)})(\rho^{(n)})^{1/2}) \geqslant 0 \\ &\quad \Rightarrow n\epsilon + \mathrm{Tr}\,\rho^{(n)}(log\bar{\rho}^{(n)} - log\rho^{(n)}) \geqslant 0. \\ &\quad \Rightarrow S(\rho^{(n)} \| \bar{\rho}^{(n)}) \leqslant n\epsilon. \end{aligned} \tag{4}$$

The other part is clear. ■

Example 2. 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

$$\bar{\rho}^{(n)} = \sum_{\rho \in \underline{\mathcal{M}}} \rho^{(n)} \geqslant \rho^{*(n)}.$$

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

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

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

$$\bar{\rho}^{(n)} - 2^{-n\epsilon}\rho^{(n)} \geqslant 0,$$

where $\rho = \omega(\rho^*)$. Therefore, $\bar{\rho}$ is universal relative to $\mathcal{M}$.

Example 3. Let $\mathcal{M}$ be a quantum model and let $\bar{\rho}$ be a universal density matrix relative to $\mathcal{M}$ and U be a unitary operator. Then $U\bar{\rho}U^{-1}$ is a universal density matrix relative to $U\mathcal{M}U^{-1}$ where $U\mathcal{M}U^{-1} = \{U\rho U^{-1} | \rho \in \mathcal{M}\}$

Theorem 2. *Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\mathcal{M}$ be a quantum model which is a compact Riemannian sub-manifold of $B_T(\mathbb{H})$ consisting of density matrices. Assume that*

1. $p : \mathcal{M} \longrightarrow]0, \infty[$ is a continuous function and $\int_{\mathcal{M}} p(\rho) dvol_{\mathcal{M}}(\rho) = 1$

2. There exists a positive real $c > 0$ such that

$$\min_{q \in Q} \max_{\rho \in \mathcal{M}}[Tr(q\rho q)] \geqslant c.$$

Moreover, for each $n \in \mathbb{N}$ let $\bar{\rho}^{(n)} = \int_{\mathcal{M}} p(\rho)\rho^{(n)} dvol_{\mathcal{M}}(\rho)$. Then the quantum source $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is uniformly Q-universal relative to $\mathcal{M}$.

Proof. Let $\epsilon > 0$ be given and let $\delta = c(1 - 2^{-\epsilon/2}) > 0$. Let $\rho_q = \mathrm{argmax}_{\rho\in\mathcal{M}}[Tr(q\rho q)] \geqslant c$. Then for all $\rho_q^* \in B(\rho_q, \delta)$ we have

$$\mathrm{Tr}(q\rho_q q) - \mathrm{Tr}(q\rho_q^* q) = \mathrm{Tr}(q(\rho_q - \rho_q^*)q) = \|q(\rho_q - \rho_q^*)q\|$$
$$\leqslant \|\rho_q - \rho_q^*\| \leqslant \|\rho_q - \rho_q^*\|_T \leqslant d(\rho_q, \rho_q^*) \leqslant \delta.$$

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

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

Since $\mathcal{M}$ is compact as it is proved in [6] there exists a constant $\nu > 0$ such that for all $\rho \in M$ we have

$$vol_{\mathcal{M}} B(\rho, \delta) \geqslant \nu.$$

Let $\beta = \min_{\rho\in\mathcal{M}} p(\rho)$ and let $k \in \mathbb{N}$ be such that $\beta\nu \geqslant 2^{-k\epsilon/2}$. Then we have

$$q\bar{\rho}^{(1)}q = \int_{\mathcal{M}} p(\rho)q\rho q dvol_{\mathcal{M}}(\rho) \geqslant \int_{B(\rho_q,\delta)} p(\rho)q\rho q dvol_{\mathcal{M}}(\rho)$$
$$\geqslant 2^{-\epsilon/2}\beta vol(B(\rho_q, \delta))q\rho_q q \geqslant 2^{-(k+1)\epsilon/2} q\rho_q q.$$

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

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

And for each $\rho \in \mathcal{M}$ we have

$$\bar{\rho}_Q^{(n)} = \sum_{q^{(n)}\in Q^{(n)}} (q^{(n)}\bar{\rho}^{(n)}q^{(n)}) \geqslant 2^{-n\epsilon} \sum_{q^{(n)}\in Q^{(n)}} (q^{(n)}\rho^{(n)}q^{(n)}) = 2^{-n\epsilon}\rho_Q^{(n)}.$$

Therefore, the quantum source $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is uniformly Q-universal relative to $\mathcal{M}$. ■

Theorem 3. *Let $\mathbb{H}$ be a Hilbert space and let $\mathcal{M}$ be a quantum model, which is a compact Riemannian sub-manifold of $B_T(\mathbb{H})$ consisting of density matrices. Assume that p: $\mathcal{M} \longrightarrow]0, \infty[$ is a continuous function such that $\int_{\mathcal{M}} p(\rho)dvol_{\mathcal{M}}(\rho) = 1$ and for each $Q \in \pi_0(\mathbb{H})$ there exists a positive number $c_Q > 0$ such that $\min_{q\in Q}\max_{\rho\in\mathcal{M}}[Tr(q\rho q)] \geqslant c_Q$. Then, with the above notations, the sequence $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is a universal quantum source relative to $\mathcal{M}$.*

The proof is a consequence of the above theorem and remark 1.

Corollary 2. *Let $\mathbb{H}$ be a finite dimensional Hilbert space and let $\mathcal{M}$ be a quantum model, which is a compact Riemannian sub-manifold of $B_T(\mathbb{H})$ consisting of density matrices. Assume that $p : \mathcal{M} \longrightarrow]0, \infty[$ is a continuous function such that $\int_{\mathcal{M}} p(\rho)dvol_{\mathcal{M}}(\rho) = 1$. Then, with the above notations, the sequence $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ is a universal quantum source relative to $\mathcal{M}$.*

The proof is evident.

Lemma 11. *$S_{\mathcal{M}}$ the set of all universal quantum source relative to the quantum model $\mathcal{M}$ is convex.*

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

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

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

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

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

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

■

Before going further it is better to introduce the notion of conditional density matrix.

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

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

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

Let $\mathbb{H}_1$ and $\mathbb{H}_2$ be Hilbert spaces. Let ρ be a density matrix on the Hilbert space $\mathbb{H}_1 \otimes \mathbb{H}_2$, $\rho_1 = Tr_2(\rho)$ and $\rho_{2|1} = \rho_1^{-1} \bullet \rho$. When $\mathbb{H}_1 = \mathbb{H}^{(n)}$ and $\mathbb{H}_2 = \mathbb{H}^{(m-n)}$, $\rho_1^{-1} \bullet \rho$ will be denoted by $\rho_{m|n}$.

Now assume that σ is a density matrix on $\mathbb{H}_1$. Then,

$$\rho(.|\sigma) = Tr_1(\sigma \bullet \rho_{2|1})$$

is clearly a positive operator on $\mathbb{H}_2$. Moreover,

$$\begin{aligned} Tr(\rho(.|\sigma)) &= Tr(Tr_1(\sigma \bullet \rho_{2|1})) = Tr(Tr_1(\sigma \bullet \rho_1^{-1} \bullet \rho)) = Tr(\sigma \bullet \rho_1^{-1} \bullet \rho) \\ &= Tr(Tr_2(\sigma \bullet \rho_1^{-1} \bullet \rho)) = Tr(\sigma^{1/2} \rho_1^{-1/2} (Tr_2(\rho)) \rho_1^{-1/2} \sigma^{1/2}) \\ &= Tr(\sigma^{1/2} \rho_1^{-1/2} \rho_1 \rho_1^{-1/2} \sigma^{1/2}) = Tr(\sigma) = 1. \end{aligned}$$

Therefore, $\rho(.|\sigma)$ is a density matrix on $\mathbb{H}_2$.

Let $\rho \in D(\mathbb{H}^{(n)} \otimes \mathbb{H})$ and $q^{(n)} \otimes q \in Q^{(n+1)}$. Then $\rho(q|q^{(n)}) = q(Tr_1(q^{(n)} \bullet \rho_{n+1|n}))q$ is called the conditional semi-density matrix of q conditioned on $q^{(n)}$ under ρ.

Definition 9. Let $\mathbb{H}$ be a separable Hilbert space and let $\hat{\rho} = (\hat{\rho}^{(n)})_{n \in \mathbb{N}}$, be a positive operator on $\mathbb{H}^{\star}$ and $\bar{\rho} = (\bar{\rho}^{(n)})_{n \in \mathbb{N}}$ where $\bar{\rho}^{(n)} = \hat{\rho}^{(1)} \bullet \hat{\rho}^{(2)} \bullet \cdots \bullet \hat{\rho}^{(n)}$, be also a positive operator on $\mathbb{H}^{\star}$. Then, the sequence $\hat{\rho} = (\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ is called a quantum strategy if the sequence $\bar{\rho} = (\bar{\rho}^{(n)})_{n \in \mathbb{N}}$ is a regular quantum source on $\mathbb{H}^{\star}$. Clearly $\hat{\rho}^{(n+1)} = (\bar{\rho}^{(n)})^{-1} \bullet \bar{\rho}^{(n+1)}$ and $\bar{\rho}^{(n+1)} = \bar{\rho}^{(n)} \bullet \hat{\rho}^{(n+1)}$

Lemma 12. *Let $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ be a quantum strategy and $(\bar{\rho}^{(n)})_{n \in \mathbb{N}}$ be its associated 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 2. For future applications we mention that because of the equality $\hat{\rho}^{(n+1)} = \bar{\rho}_{n+1|n}$, quantum strategies are also called quantum estimators. Let $(\bar{\rho}^{(n)})_{n \in \mathbb{N}}$ be a quantum source. It is straightforward to see that $({\bar{\rho}^{(n)}}_Q)_{n \in \mathbb{N}}$ is a Q-quantum source and gives rise to a Q-quantum strategy.

Definition 10. A quantum estimator $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ is called good with respect to a quantum model $\mathcal{M}$ if its associated quantum source is universalrelative to $\mathcal{M}$. Under conditions and notations of theorems 2 the quantum strategy associated with the $Q-$ universal quantum source $\bar{\rho}^{(n)} = \int_{\mathcal{M}} p(\rho)\rho^{(n)} dvol_{\mathcal{M}}(\rho)$, is good.

Example 4. Let $\mathcal{M}$ be the following quantum model.

$$\mathcal{M} = \{\rho_\theta | 0 \leqslant \theta \leqslant 1\},$$

where ρ_θ is a 2×2-density matrix defined as follows

$$\rho_\theta = \begin{pmatrix} \theta & \sqrt{c(\theta - \theta^2)} \\ \sqrt{c(\theta - \theta^2)} & 1 - \theta \end{pmatrix}$$

and $0 \leqslant c \leqslant 1$ is a real constant.

Let $Q = \{q_1, q_2\}$ where $q_1 = |0\rangle\langle 0|$ and $q_2 = |1\rangle\langle 1|$ and $\{|0\rangle, |1\rangle\}$ is the standard basis of the 2-dimensional Hilbert space $\mathbb{H} = \mathbb{C}^2$. Then

$$\rho_{\theta Q} = q_1 \rho_\theta q_1 + q_2 \rho_\theta q_2 = \begin{pmatrix} \theta & 0 \\ 0 & 1 - \theta \end{pmatrix}$$

is a diagonal matrix.

For simplicity we omit the index Q. Assume that $q^{(n)} \in Q^{(n)}$ consists of $k-$ times q_1 and $(n-k)-$ times q_2. Then for each $0 \leqslant \theta \leqslant 1$ we have

$$q^{(n)}\rho_\theta{}^{(n)}q^{(n)} = \theta^k(1-\theta)^{(n-k)}q^{(n)}.$$

It is straightforward to see that the maximum likelihood estimator for $q^{(n)}$ is $\rho_{\hat{\theta}(q^{(n)})}$ where $\hat{\theta}(q^{(n)}) = k/n$.

Clearly $\mathcal{M}$ is a Bayesian quantum model. As we have proved earlier its associated universal quantum source is $(\bar{\rho}^{(n)})_{n\in\mathbb{N}}$, where $\bar{\rho}^{(n)} = \int_0^1 \rho_\theta{}^{(n)}d\theta$ and for $q^{(n)} \in Q^{(n)}$ as above we have

$$q^{(n)}\bar{\rho}^{(n)}(q^{(n)}) = \int_0^1 q^{(n)}\rho_\theta{}^{(n)}q^{(n)}d\theta = \left(\int_0^1 \theta^k(1-\theta)^{(n-k)}d\theta\right)q^{(n)}.$$

One can compute the above integral by partial integration and see that

$$q^{(n)}\bar{\rho}^{(n)}(q^{(n)}) = \frac{1}{(n+1)\binom{n}{k}}q^{(n)}.$$

In the same way for $q^{(n+1)} = q^{(n)} \otimes q_1 \in Q^{(n+1)}$ we have

$$q^{(n+1)}\bar{\rho}^{(n+1)}(q^{(n+1)}) = \frac{1}{(n+2)\binom{n+1}{k+1}}q^{(n+1)}.$$

Therefore

$$\hat{\rho}^{(n+1)}(q_1|q^{(n)}) = \frac{q^{(n+1)}\bar{\rho}^{(n+1)}(q^{(n+1)})}{q^{(n)}\bar{\rho}^{(n)}(q^{(n)})} = \frac{(n+1)\binom{n}{k}}{(n+2)\binom{n+1}{k+1}}q_1 = \frac{k+1}{n+2}q_1.$$

The density matrix $\hat{\rho}^{(n+1)}(.|q^{(n)})$ is called modified maximum likelihood estimator for $q^{(n)}$. Evidently, for large $n \in \mathbb{N}$ it is very close to $\rho_{\hat{\theta}(q^{(n)})}$. Clearly, $\hat{\rho}^{(n+1)}$ is a good strategy.

For many smooth parametric quantum models $\mathcal{M} = \{\rho_\theta|\theta \in \Theta\}$ like the above example an associated modified maximum likelihood estimator $\hat{\rho} = (\hat{\rho}^{(n)})_{n\in\mathbb{N}}$ is a good quantum strategy. Moreover, in these cases for $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ the associated universal quantum source we have

Theorem 4. *For each $\theta \in \Theta$ we have*

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

The proof is the same as the proof in the classical case given in chapter 8 of [3] on page 246 with simple modifications.

4. Quantum prediction and quantum estimation

As we said in the introduction, quantum prediction and quantum estimation are the most important subjects of quantum statistical inference. Following the classical works in MDL principle, our method of statistical inference is in general based on universal quantum source and use of it to do quantum prediction and quantum estimation.

4.1. Quantum version of classical MDL prediction and estimation

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

$$q_I^{(n-1)} \bullet \hat{\rho}^{(n)} = argmax_{\rho\in\mathcal{M}}(q_I^{(n-1)}\rho^{(n-1)}q_I^{(n-1)}).$$

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 good. But in many cases (see the above example), a modified version of the maximum likelihood Q-quantum strategy, which is very close to the unmodified one and the difference between them tends rapidly to zero, is a good one.

This good Q-quantum strategy enables us to predict next outcome given the data $q_I^{(n-1)}$. Moreover, let the data $q_I^{(n-1)}$ be really generated by $\rho \in M$. Then as we will see in the next chapter $q_I^{(n-1)} \bullet \hat{\rho}^{(n)}$ can be considered as an estimation of ρ.

4.2. Quantum version of classical two-part code estimation

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 $I \in \mathbb{N}^{(n)}$, let $\ddot{\rho}_n$ be defined as follows

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

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. More precisely, let us suppose that $\underline{\mathcal{M}}$ is a compact Riemannian sub-manifold of the Hilbert space $(B_T(\mathbb{H}), \langle .|. \rangle_T)$ consisting of semi-density matrices where for ρ and ρ' in $B_T(\mathbb{H})$, $\langle \rho | \rho' \rangle = \mathrm{Tr}(\rho\rho')$ and $(\underline{\mathcal{M}}, \Sigma, \mu)$ be its associated canonical measure space. To obtain $\ddot{\rho}_n$, 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 $\underline{\mathcal{M}}$, and let Z' be the set of all elements $\rho \in Z$ at which the bundle map $Hessian(h)$: $T\underline{\mathcal{M}} \longrightarrow T\underline{\mathcal{M}}$ 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. (For more information about finding extremum points see [7])

In the next section we will show that given the outcome $q_I^{(n)}$, it is an estimator of the state of the system.

5. Consistency and convergence

Consistency is a very important property of different methods of statistical (inductive) inferences. Let us explain briefly what we mean by it.

Assume that $\mathbb{H}$ is a separable Hilbert space and $\mathcal{M}$ is a quantum model on $\mathbb{H}$. we say that a method of quantum statistical inference is consistent with respect to $\mathcal{M}$ if for $\rho_0 \in \mathcal{M}$ and $Q \in \pi_0(\mathbb{H})$, we perform the quantum measurement Q on the quantum system $\mathbb{H}$ in the state ρ_0 repeatedly and obtain more and more data the 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 and convergence.

5.1. Consistency based on distinguishability

Let $\mathbb{H}$ be a separable Hilbert space and let T and S be in $B_H(\mathbb{H})$ and λ be a complex number; Assume that $S \neq 0$, $T = \lambda S$ and p is the orthogonal projection onto the image of S. Then, we put $T/S = \lambda p$. Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}}$ be a quantum source on $\mathbb{H}^*$. For each $n \in \mathbb{N}$ let P_n be a unary relation on $Q^{(n)}$. Then,

$$Tr(\sum_{q^{(n)} \in Q^{(n)} | P_n(q^{(n)})} q^{(n)} \bar{\rho}^{(n)} q^{(n)})$$

will be denoted by $\bar{\rho}(P_n)$. suppose that $\bar{\rho}' = (\bar{\rho}'^{(n)})_{n\in\mathbb{N}}$ is another quantum source on $\mathbb{H}^*$. For each $n \in \mathbb{N}$, and each $\delta > 0$ let P_n^δ be the unary relation

$$\frac{q^{(n)} \bar{\rho}'^{(n)} q^{(n)}}{q^{(n)} \bar{\rho}^{(n)} q^{(n)}} > \delta$$

on $Q^{(n)}$.

Definition 11. Under the above notations and conventions we say $\bar{\rho}'$ is asymptotically distinguishable from $\bar{\rho}$ if for all $\delta > 0$ we have

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

Let $\underline{\mathcal{M}}$ be a countable Bayesian set of generalized quantum sources on $\mathbb{H}^*$ and $\mathcal{M}$ be its associated set of quantum sources. For each $n \in \mathbb{N}$, let us denote $\omega(\bar{\rho}^{(n)})$ by $\rho^{(n)}$. For each $(q^{(n)}) \in Q^{(n)}$ define $\ddot{\rho}_{(n)}$ as follows:

$$(\star)\, \ddot{\rho}_{(n)} = \max_{\rho^{(n)} \in M_n} q^{(n)} \bar{\rho}^{(n)} q^{(n)}.$$

and $M_n = \{\rho^{(n)} | \rho \in \mathcal{M}\}$. Observe that $\ddot{\rho}_{(n)}$ depends on $q^{(n)}$.

Now we have the following important consistency theorem.

Theorem 5. *Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\underline{\mathcal{M}}$, $\mathcal{M}$ and $\ddot{\rho}_{(n)}$ be as above. Let $\bar{\rho}^* \in \underline{\mathcal{M}}$ and $\ddot{\mathcal{M}}$ be the subset of $\mathcal{M}$ consisting of quantum sources asymptotically distinguishable from ρ^* and $\ddot{M}_n = \{\rho^{(n)} | \rho \in \ddot{\mathcal{M}}\}$. Then*

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

Proof. Let $\ddot{\rho}_{(n)} \in \ddot{M}_n$. From the equality $(\star)$, for some $\bar{\rho}_n \in \underline{M}_n$ we have

$$q^{(n)}\bar{\rho}_n q^{(n)} \geqslant q^{(n)}\bar{\rho}^{*(n)}q^{(n)}$$

Therefore,

$$\rho^*(\ddot{\rho}_{(n)} \in \ddot{M}_n) = \rho^*\left\{for\ some\ \tau \in \ddot{M}_n,\ \frac{q^{(n)}\tau q^{(n)}}{q^{(n)}\rho^{*(n)}q^{(n)}} \geqslant \frac{Tr(\bar{\rho}^*)}{Tr(\bar{\tau})}\right\}.$$

Let us denote $\frac{Tr(\bar{\rho}^*)}{Tr(\bar{\tau})}$ by $\delta(\tau)$. Assume that $n\colon \to \rho_n$ is a bijective mapping from $\mathbb{N}$ onto $\underline{\ddot{\mathcal{M}}}$ and $m = Tr(\sum_{n=1}^{\infty}\bar{\rho}_n)$. Let $\epsilon > 0$ be given and let $\pi = m - \epsilon\ Tr(\rho^*)$. suppose that N is the least integer such that $Tr(\sum_{n=1}^{N}\bar{\rho}_n) \geqslant \pi$. Let $\underline{\ddot{\mathcal{M}}} = \{\rho_n | 1 \leqslant n \leqslant N\}$ and $\bar{\ddot{\mathcal{M}}} = \ddot{\mathcal{M}} - \underline{\ddot{\mathcal{M}}}$.

Evidently,

$$\rho^*(\ddot{\rho}_{(n)} \in \ddot{\mathcal{M}}) = \rho^*(\ddot{\rho}_{(n)} \in \underline{\ddot{\mathcal{M}}}) + \rho^*(\ddot{\rho}_{(n)} \in \bar{\ddot{\mathcal{M}}}).$$

and

$$\lim_{n\to\infty}\rho^*(\ddot{\rho}_{(n)} \in \ddot{\mathcal{M}}) = \lim_{n\to\infty}\rho^*(\ddot{\rho}_{(n)} \in \underline{\ddot{\mathcal{M}}}) + \lim_{n\to\infty}\rho^*(\ddot{\rho}_{(n)} \in \bar{\ddot{\mathcal{M}}}).$$

Assume that $\rho \in \ddot{\mathcal{M}}$ and $\delta(\rho) = Tr(\bar{\rho}^*)/Tr(\bar{\rho})$ Since $\bar{\rho}$ is asymptotically distinguishable from ρ^*, $\lim_{n\to\infty}\rho^*(P_n^{\delta(\rho)}) = 0$. Since $\underline{\ddot{\mathcal{M}}}$ is a finite set we have

$$\lim_{n\to\infty}\rho^*(\ddot{\rho}_{(n)} \in \underline{\ddot{\mathcal{M}}}) \leqslant \lim_{n\to\infty}\sum_{\bar{\rho}\in\underline{\ddot{\mathcal{M}}}}\rho^*(P_n^{\delta(\rho)}) = \sum_{\bar{\rho}\in\underline{\ddot{\mathcal{M}}}}\lim_{n\to\infty}\rho^*(P_n^{\delta(\rho)}) = 0.$$

On the other hand by the fundamental coding theorem we have

$$\rho^*(P_n^{\delta(\rho)}) \leqslant 1/\delta(\rho).$$

Hence,

$$\begin{aligned}\rho^*(\ddot{\rho}_{(n)} \in \bar{\ddot{\mathcal{M}}}) &= \textstyle\sum_{\rho\in\bar{\ddot{\mathcal{M}}}}\rho^*(P_n^{\delta(\rho)})\\ &\leqslant \textstyle\sum_{\rho\in\bar{\ddot{\mathcal{M}}}} 1/\delta(\rho)\\ &= \textstyle\sum_{\rho\in\bar{\ddot{\mathcal{M}}}} Tr(\bar{\rho})/Tr(\bar{\rho}^*)\\ &= (m-\pi)/Tr(\bar{\rho}^*)\\ &= \epsilon.\end{aligned}$$

Therefore,

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

5.2. Consistency in terms of KL risk and Cezaro average KL risk

Theorem 6. *Let $\bar{\rho}$ and ρ^* be Quantum sources and ρ^* be simple. Then*

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

Proof. Assume that Q is a complete set of mutually orthogonal minimal projections. 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:

$$\begin{aligned}S(\rho^{*(n)}\|\bar{\rho}^{(n)}) &= \mathrm{Tr}(\rho^{*(n)}\log\rho^{*(n)} - \rho^{*(n)}\log\bar{\rho}^{(n)})\\ &= \mathrm{Tr}\,\rho^{*(n)}(\log\rho^{*(n)} - \log\bar{\rho}^{(n)})\\ &= \mathrm{Tr}((\Pi_{i=1}^{n}(\rho^*_{i|(i-1)}))(\log\Pi_{i=1}^{n}\rho^*_{i|(i-1)} - \log\Pi_{i=1}^{n}\bar{\rho}_{i|(i-1)}))\\ &= \textstyle\sum_{i=1}^{n}\mathrm{Tr}(\rho^{*(i-1)}\rho^*_{i|(i-1)}(\log\rho^*_{i|(i-1)} - \log\bar{\rho}_{i|(i-1)}))\\ &= \textstyle\sum_{i=1}^{n}E_{\rho^{*(i-1)}}S(\rho^*_{i|(i-1)}\|\bar{\rho}_{i|(i-1)}).\end{aligned} \quad (5)$$

(See also [3].) ■

Definition 12. Let ρ^* and $\bar{\rho}$ be quantum sources and ρ^* be simple. and $\hat{\rho}^*$ and $\hat{\rho}$ be their associated quantum strategies. Then, the standard KL-risk of $\hat{\rho}^{*(n)}$ with respect to $\hat{\rho}^{(n)}$ is

$$RISK_n(\hat{\rho}^*, \hat{\rho}) =^w E_{\rho^{*(n-1)}}[S(\hat{\rho}^{*(n)}\|\hat{\rho}^{(n)})].$$

And the Cezaro average risk of $\hat{\rho}^{*(n)}$ with respect to $\hat{\rho}^{(n)}$ is

$$R\bar{I}SK_n(\hat{\rho}^*, \hat{\rho}) = 1/nS(\rho^{*(n)}\|\bar{\rho}^{(n)}) = 1/n\sum_{i=1}^{n} RISK_i(\hat{\rho}^*, \hat{\rho}).$$

Theorem 7 (Convergence theorem for quantum Estimators). *Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. Let $\mathcal{M}$ be a quantum model on $\mathbb{H}$ and $\bar{\rho}$ be a Q-universal quantum source with respect to $\mathcal{M}$. Then $\hat{\rho}$ the Q-quantum estimator associated with Q-universal quantum source $\bar{\rho}_Q^{(n)}$ is Cezaro consistent with respect to $\mathcal{M}$. In other words for all $\rho^* \in \mathcal{M}$ we have*

$$\lim_{n\to\infty} 1\Big/n\sum_{i=1}^{n} RISK_i(\hat{\rho}^*, \hat{\rho}) = 0.$$

The proof is a consequence of the definition of Q-universal source and theorem 6. ■

Lemma 13. *Let f and F be two increasing positive real functions defined on $\mathbb{R}^+$. If the function f/F is decreasing and $f = O(F)$, then $f(n+1) - f(n) = O(F(n+1) - F(n))$.*

Proof. Assume that there exists $c > 0$ such that for n large enouph $f(n) \leqslant cF(n)$. Let $f(n+1) = c_1F(n+1)$ and $f(n) = c_0F(n)$. Therefore,

$$f(n+1) - f(n) = c_1F(n+1) - c_0F(n).$$

Since $c_1 \leqslant c_0 \leqslant c$ we have

$$f(n+1) - f(n) = c_1F(n+1) - c_0F(n) \leqslant c_0(F(n+1) - F(n)) \leqslant c(F(n+1) - F(n)).$$

Lemma 14. *Let $f: \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ be a differentiable and integrable decreasing function. Assume that $F(x) = \int_0^x f(x)dx$. The sequence u_n is defined as follow: $u_0 = 0$ and for all $1 \leqslant n \in \mathbb{N}$, $u_n = f(n-1)$. Assume that $(a_n)_{n\in\mathbb{N}}$ is a sequence of non-negative real numbers. Then*

1) If $a_n = O(u_n)$, Then $\sum_{i=1}^n a_i = O(F(n) + 1)$. Conversely, if for n large enough the function $\sum_{i=1}^n a_i/(F(n) + 1)$ is decreasing and $\sum_{i=1}^n a_i = O(F(n) + 1)$, then $a_n = O(u_n)$.
2) If $\lim_{n\to\infty} a_n = 0$, then $\lim_{n\to\infty} \sum_{i=1}^n a_i/n = 0$. Conversely, if for n large enough the function $\sum_{i=1}^n a_i/n$ is decreasing and $\lim_{n\to\infty} \sum_{i=1}^n a_i/n = 0$, then $\lim_{n\to\infty} a_n = 0$.

Proof. 1) In approximating the integral by sum and remembering the fact that the function *f* is decreasing we have

$$F(n) = \sum_{i=1}^{n-1} u_i + O(1) = \sum_{i=1}^{n} u_i + O(1).$$

Therefore,

$$\sum_{i=1}^{n} a_i = O(\sum_{i=1}^{n} u_i) = O(F(n) + O(1)) = O(F(n) + 1).$$

Conversely, 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 $\sum_1^n a_i \leqslant c(F(n) + 1)$. By the above lemma we have

$$a_{n+1} = \sum_{1}^{n+1} a_i - \sum_{1}^{n} a_i \leqslant c(F(n+1) - F(n)) = cf(\theta_n).$$

Where, $n \leqslant \theta_n \leqslant n + 1$. Since *f* is decreasing we have $f(\theta_n) \leqslant f(n)$. Hence, $a_{n+1} \leqslant cu_{n+1}$. Therefore, $a_n = O(u_n)$.

2) From the equality $\lim_{n\to\infty} a_n = 0$ it follows that for each $\epsilon > 0$ there exists $n_1 \in \mathbb{N}$ such that for all $n > n_1$, we have $a_n \leqslant \epsilon$. Suppose that for $k \in \mathcal{N}$, $\frac{\sum_{i\leqslant n_1} a_i}{k} \leqslant \epsilon$. Let $n \in \mathcal{N}$ be greater than $k - 2n_1$. Then $2(n + n_1) > k + n$. So,

$$\sum_{i=1}^{n+n_1} a_i = \sum_{i=1}^{n_1} a_i + \sum_{i=n_1}^{n+n_1} a_i < (k+n)\epsilon < 2(n+n_1)\epsilon.$$

Hence, $\frac{\sum_{i=1}^{n_0} a_i}{n_0} < 2\epsilon$, where $n_0 = n + n_1$. It is clear that for all $n \geqslant n_0$ we have

$$\frac{\sum_{i=1}^{n} a_i}{n} < 2\epsilon.$$

Therefore,

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n}a_i}{n}=0.$$

Conversely, since for n large enough the sequence $\frac{\sum_{i=1}^{n}a_i}{n}$ is decreasing we have $\frac{\sum_{i=1}^{n+1}a_i}{n+1}\leqslant\frac{\sum_{i=1}^{n}a_i}{n}$. Then,

$$\sum_{i=1}^{n+1}a_i\leqslant\frac{(n+1)}{n}\sum_{i=1}^{n}a_i.$$

Therefore, $a_{n+1}\leqslant\frac{\sum_{i=1}^{n}a_i}{n}$. But the sequence $\frac{\sum_{i=1}^{n}a_i}{n}$ is convergent. Therefore, the sequence a_n is also convergent.

Theorem 8. *Let $\rho^*=(\rho^{*(n)})_{n\in\mathbb{N}}$ and $\bar{\rho}=(\rho^{(n)})_{n\in\mathbb{N}}$ be $Q-$ quantum sources on the Hilbert space $\mathbb{H}^*$. Then*

1) if $\lim_{n\to\infty}RISK_n(\hat{\rho}^*,\hat{\rho})=^w 0$ then $\lim_{n\to\infty}\frac{1}{n}S(\rho^{*(n)}\|\bar{\rho}^{(n)})=^w 0$. Conversely, if $\frac{1}{n}S(\rho^{*(n)}\|\bar{\rho}^{(n)})$ is decreasing and $\lim_{n\to\infty}\frac{1}{n}S(\rho^{*(n)}\|\bar{\rho}^{(n)})=^w 0$ then $\lim_{n\to\infty}RISK_n(\hat{\rho}^*,\hat{\rho})=^w 0$

2) Let $f:\mathbb{R}^+\longrightarrow\mathbb{R}^+$ be a differentiable and integrable decreasing function and let $F(x)=\int_0^x f(t)dt$. Then, if $RISK_n(\hat{\rho}^*,\hat{\rho})=^w O(f(n))$ then

$$S(\rho^{*(n)}\|\bar{\rho}^{(n)})=O((F(n)+1)).$$

Conversely, if $\frac{S(\rho^{*(n)}\|\bar{\rho}^{(n)})}{(F(n)+1)}$ is decreasing and $S(\rho^{*(n)}\|\bar{\rho}^{(n)})=O(F(n)+1)$, then $RISK_{n+1}(\hat{\rho}^*,\hat{\rho})=^w O(f(n))$

Proof. The proof is a consequence of the definitions and lemma 14. See also [3]. ■

5.3. Consistency in terms of Renyi divergences and Hellinger distance

Let $\mathbb{H}$ be a Hilbert space. Let ρ_1 and ρ_2 be density matrices. Then

1) The quantum relative entropy of ρ_1 to ρ_2 is

$$S_{nat}(\rho_1\|\rho_2)=Tr(\rho_1\,ln\rho_1)-Tr(\rho_1\,ln\rho_2).$$

2) The Helinger distance of ρ_1 and ρ_2 is

$$He^2(\rho_1\|\rho_2)=\|\rho_1^{1/2}-\rho_2^{1/2}\|_T^2$$

3) Let $\lambda>0$ be a real number. The Renyi divergence of order λ of ρ_1 and ρ_2 is defined as follows:

$$\bar{d}_\lambda(\rho_1\|\rho_2)=-\frac{1}{1-\lambda}ln(<\rho_1^\lambda|\rho_2{}^{1-\lambda}>_T).$$

Observe that

$$\begin{aligned}He^2(\rho_1\|\rho_2)&=\|\rho_1^{1/2}-\rho_2^{1/2}\|_T^2=Tr[(\rho_1^{1/2}-\rho_2^{1/2})^2]=Tr(\rho_1+\rho_2-2\rho_1{}^{1/2}\rho_2{}^{1/2})\\&=2(1-Tr(\rho_1{}^{1/2}\rho_2{}^{1/2}))\leqslant[-2\,ln<\rho_1^{1/2}|\rho_2^{1/2}>_T]=\bar{d}_{1/2}(\rho_1\|\rho_2).\end{aligned}$$

Let $\mathcal{M}=\{\tau_n|n\in\mathbb{N}\}$ be a countable quantum model and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of nonzero real numbers such that $\sum n\in\mathbb{N}u_n=1$. The set consisting of all elements of the form $u_n\tau_n$ will be denoted by $\underline{\mathcal{M}}$. Let $Q\in\pi_0(\mathbb{H})$. For $\alpha\geqslant 1$, let $\underline{\mathcal{M}}_\alpha=\{\rho_\alpha|\rho\in\underline{\mathcal{M}}\}$. Where, $\rho_\alpha=[Tr(\rho)]^{\alpha-1}\rho$. Let $\bar{\rho_\alpha}^{(n)}$ be defined as follows:

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

Assume that $(\bar{\rho_\alpha}^{(n)})_{n\in\mathbb{N}}$ is a universal semi-density matrix for $\mathcal{M}$. Suppose $\ddot{\rho}_n$ is defined as follows:

For $q^{(n)}\in Q^{(n)}$,

$$\ddot{\rho}_n=\text{argmax}_{\rho\in\mathcal{M}_n}q^{(n)}\rho^{(n)}q^{(n)}.$$

Observe that $\ddot{\rho}_n$ depends on $q^{(n)}$. Let $\rho_k=u_k\tau_k=argmax_{\rho\in\underline{\mathcal{M}}}q^{(n)}\rho^{(n)}q^{(n)}$. Then evidently

$$q^{(n)}\bar{\rho_\alpha}^{(n)}q^{(n)}=u_k{}^\alpha q^{(n)}\ddot{\rho}_n^{(n)}q^{(n)}.$$

In the following, we write $\ddot{\rho}_n$ instead of $\ddot{\rho}_n^{(n)}$.

Theorem 9. *Let ρ^* be the state of the system. Under the above notations and conventions for all $\alpha>1$ and $0<\lambda=1-1/\alpha$ we have*

$$E_{\rho_Q^{*(n)}}(\bar{d}_\lambda(\rho_Q^{*(n)}\|\ddot{\rho}_{nQ}^{(n)}))\leqslant\frac{1}{n}S_{nat}(\rho_Q^{*(n)}\|\bar{\rho}_{\alpha Q}^{(n)}).$$

And for $\alpha = 2$ we have

$$E_{\rho_Q^{*(n)}}(He^2(\rho_Q^{*(n)}\|\ddot{\rho}_{nQ}^{(n)})) \leqslant \frac{1}{n}S_{nat}(\rho_Q^{*(n)}\|\bar{\rho}_{\alpha Q}^{(n)}).$$

Proof. (The proof is a modified version of the proof of theorem 15.3 of [3]). For simplicity we omit the index Q. Since $\lambda = 1 - \alpha^{-1}$. we have $\alpha = 1/1 - \lambda$. Let $A(\rho^*\|\ddot{\rho}) = Tr(\rho^{*\lambda}\ddot{\rho}^{1-\lambda})$. For each $q^{(n)} \in Q^{(n)}$ we have

$$\begin{aligned}
q^{(n)}\bar{d}_\lambda(\rho^{*(n)}\|\ddot{\rho}_n) &= (-1/1-\lambda)q^{(n)}\ln A(\rho^{*(n)}\|\ddot{\rho}_n)\\
&= \frac{1}{n}q^{(n)}\ln\frac{u_k^\alpha q^{(n)}\ddot{\rho}_n q^{(n)}}{q^{(n)}\bar{\rho}_\alpha^{(n)}q^{(n)}} + \frac{\alpha}{n}q^{(n)}\ln\frac{1}{A^{(n)}(\rho^*\|\ddot{\rho})}\\
&= \frac{1}{n}q^{(n)}\ln\frac{q^{(n)}\rho^{*(n)}q^{(n)}}{q^{(n)}\bar{\rho}_\alpha^{(n)}q^{(n)}} + \frac{\alpha}{n}q^{(n)}\ln\frac{\left(\frac{q^{(n)}\ddot{\rho}_n q^{(n)}}{q^{(n)}\rho^{*(n)}q^{(n)}}\right)^{1/\alpha}u_k}{A^{(n)}(\rho^*\|\ddot{\rho})}\\
&= \frac{1}{n}q^{(n)}\ln\frac{q^{(n)}\rho^{*(n)}q^{(n)}}{q^{(n)}\bar{\rho}_\alpha^{(n)}q^{(n)}} + \frac{\alpha}{n}q^{(n)}\ln\frac{\left(\frac{q^{(n)}\ddot{\rho}_n q^{(n)}}{q^{(n)}\rho^{*(n)}q^{(n)}}\right)^{1-\lambda}u_k}{A^{(n)}(\rho^*\|\ddot{\rho})}\\
&\leqslant \frac{1}{n}q^{(n)}\ln\frac{q^{(n)}\rho^{*(n)}q^{(n)}}{q^{(n)}\bar{\rho}_\alpha^{(n)}q^{(n)}} + \frac{\alpha}{n}q^{(n)}\ln\sum_{m\in\mathbb{N}}\frac{\left(\frac{q^{(n)}\ddot{\rho}_n q^{(n)}}{q^{(n)}\rho^{*(n)}q^{(n)}}\right)^{1-\lambda}u_m}{A^{(n)}(\rho^*\|\rho_m)}
\end{aligned}$$

$$\begin{aligned}
\text{Therefore,}\quad E_{\rho^{*(n)}}[\bar{d}_\lambda(\rho^{*(n)}\|\ddot{\rho}_n)] &= \left(\frac{1}{n}\right)S_{nat}(\rho^{*(n)}\|\bar{\rho}_\alpha^{(n)})\\
&+\left(\frac{\alpha}{n}\right)Tr(E_{\rho^{*(n)}}\ln\sum_{m\in\mathbb{N}}\frac{\left(\frac{\boldsymbol{q}^{(n)}\ddot{\rho}_n\boldsymbol{q}^{(n)}}{\boldsymbol{q}^{(n)}\rho^{*(n)}\boldsymbol{q}^{(n)}}\right)^{1-\lambda}u_m}{A^{(n)}(\rho^*\|\rho_m)})
\end{aligned}$$

where $\boldsymbol{q}^{(n)}$ is a random projection under the density matrix ρ^* with values in $Q^{(n)}$.

Since $Q \in \pi_0(\mathbb{H})$ and

$$Tr(q^{(n)}\rho_k^{(n)}q^{(n)}\ln q^{(n)}\rho_l^{(n)}q^{(n)}) = Tr(q^{(n)}\rho_k^{(n)}q^{(n)})Tr(\ln q^{(n)}\rho_l^{(n)}q^{(n)}),$$

Now by Jensen's inequality we have

$$\begin{aligned}
&\left(\frac{\alpha}{n}\right)Tr(E_{\rho^{*(n)}}\ln\sum_{m\in\mathbb{N}}\frac{\left(\frac{\boldsymbol{q}^{(n)}\ddot{\rho}_n\boldsymbol{q}^{(n)}}{\boldsymbol{q}^{(n)}\rho^{*(n)}\boldsymbol{q}^{(n)}}\right)^{1-\lambda}u_m}{A^{(n)}(\rho^*\|\rho_m)})\\
&\leqslant\left(\frac{\alpha}{n}\right)Tr\ln E_{\rho^{*(n)}}\sum_{m\in\mathbb{N}}\frac{\left(\frac{\boldsymbol{q}^{(n)}\ddot{\rho}_n\boldsymbol{q}^{(n)}}{\boldsymbol{q}^{(n)}\rho^{*(n)}\boldsymbol{q}^{(n)}}\right)^{1-\lambda}u_m}{A^{(n)}(\rho^*\|\rho_m)})\\
&\leqslant\left(\frac{\alpha}{n}\right)\ln\sum_{m\in\mathbb{N}}\left[\left(\frac{u_m}{A^{(n)}(\rho^*\|\rho_m)}\right)Tr(\sum_{q^{(n)}\in Q^{(n)}}(q^{(n)}\rho^{*(n)}q^{(n)})^\lambda(q^{(n)}\rho_m^{(n)}q^{(n)})^{1-\lambda})\right]\\
&=\left(\frac{\alpha}{n}\right)\ln\sum_{m\in\mathbb{N}}\left[\left(\frac{u_m}{A^{(n)}(\rho^*\|\rho_m)}\right)Tr(\Pi^n(\rho^{*\lambda}\rho_m^{1-\lambda}))\right]\\
&=\left(\frac{\alpha}{n}\right)\ln\sum_{m\in\mathbb{N}}\left[\left(\frac{u_m}{A^{(n)}(\rho^*\|\rho_m)}\right)\Pi^n Tr(\rho^{*\lambda}\rho_m^{1-\lambda})\right]\\
&=\left(\frac{\alpha}{n}\right)\ln\sum_{m\in\mathbb{N}}\left[\left(\frac{u_m}{A^{(n)}(\rho^*\|\rho_m)}\right)A^{(n)}(\rho^*\|\rho_m)\right]\\
&=\left(\frac{\alpha}{n}\right)\ln\sum_{m\in\mathbb{N}}u_m.
\end{aligned}$$

But $\sum_{m\in\mathbb{N}}u_m = 1$. Therefore, $E_{\rho^*}[\bar{d}_\lambda(\rho^{(n)}\|\ddot{\rho}_\alpha^{(n)})] \leqslant (\frac{1}{n})S_{not}(\rho^{*(n)}\|\bar{\rho}_\alpha^{(n)})$.

Corollary 3. *From the above theorem, theorem 6 and the relation between Renyi divergences and Hellinger distance explained above we have:*

1. $\lim_{n\to\infty}E_{\rho_Q^{*(n)}}(He^2(\rho_Q^{*(n)}\|\ddot{\rho}_{nQ}^{(n)})) = 0.$

2. Let $\alpha > 1$ and $0 < \lambda = 1 - 1/\alpha$. Then,

$$\lim_{n\to\infty} E_{\rho_Q^{*(n)}}(\bar{d}_\lambda(\rho_Q^{*(n)} \| \ddot{\rho}_{nQ}^{(n)})) = 0.$$

6. Applications

As we described before, estimation and prediction are the most important purposes of quantum statistical inference and particularly this paper. In order to show the advantages of our method, in this section we explain the usage of this method by two examples. The first example that we choose is selecting a density matrix among three ones which are originally considered in [8]. For multiple ions quantum tomography, two famous traditional methods, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are used for estimation. For more information please see [8].

In this case, the quantum model consists of three one-ion states of different degrees of purity: a pure state, one with eigenvalues (0.95, 0.05), and the other with eigenvalues (0.72, 0.28). For each state, they simulated datasets with varying numbers of repetitions n $=$ 10, 50, 100, 250, 500. Table 1, shows the number of times (out of 1000 samples) that BIC and AIC chose correctly, [8].

Now we choose among these states with the quantum version of classical two-part code estimation, semi-density matrices.

Let

$$\rho = \begin{bmatrix} a & b \\ b & 1-a \end{bmatrix}, \; Q = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}.$$

Let $q_{i_j} = |0\rangle\langle 0|$ then $q_{i_j}\rho q_{i_j} = a q_{i_j}$ and if $q_{i_j} = |1\rangle\langle 1|$ then $q_{i_j}\rho q_{i_j} = (1-a)q_{i_j}$.

$$q_I^{(n)}(2^{-L(\rho)}\rho^{(n)})q_I^{(n)} = 2^{-L(\rho)}(q_{i_1}\rho q_{i_1}) \otimes (q_{i_2}\rho q_{i_2}) \otimes \cdots \otimes (q_{i_n}\rho q_{i_n}).$$

Assume that $q_I^{(n)} \in Q^{(n)}$ consists of $k-$ times $|0\rangle\langle 0|$ and $(n-k)-$ times $|1\rangle\langle 1|$. Then:

$$q_I^{(n)}(2^{-L(\rho)}\rho^{(n)})q_I^{(n)} = 2^{-L(\rho)}a^k(1-a)^{(n-k)}q_I^{(n)}$$

Now let us calculate this for the states considered in [8]. In the following, when there is no ambiguity for we omit q's.

Example 5. For the states in [8], we define the following quantum generalized model

$$\mathcal{M} = \left\{\rho_1 = \frac{1}{3}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \rho_2 = \frac{4}{9}\begin{bmatrix} 0.95 & 0 \\ 0 & 0.05 \end{bmatrix}, \rho_3 = \frac{2}{9}\begin{bmatrix} 0.72 & 0 \\ 0 & 0.28 \end{bmatrix}\right\}$$

$$q_I^{(n)}(2^{-L(\rho_1)}\rho_1^{(n)})q_I^{(n)} = \begin{cases} \frac{1}{3} & k = n \\ 0 & \text{Otherwise} \end{cases}$$

$$q_I^{(n)}(2^{-L(\rho_2)}\rho_2^{(n)})q_I^{(n)} = \frac{4}{9}(0.95)^k(0.05)^{(n-k)}$$

$$q_I^{(n)}(2^{-L(\rho_3)}\rho_3^{(n)})q_I^{(n)} = \frac{2}{9}(0.75)^k(0.25)^{(n-k)}$$

For each state, we simulated datasets with varying numbers of repetitions n $=$ 10, 50, 100, 250, 500. Table 2, shows the number of times (out of 1000 samples) that the quantum version of classical two-part code estimation chose correctly.

As expected, for small sample sizes, n, the quantum version of classical two-part code estimation may select the wrong model because it has a built-in preference for 'simple' models. But for all large n, it will select the correct model. Yet for the small n, it is far better than classical methods, like AIC and BIC. In the case of the pure state because of the appropriate choice of weight, it never missed and always chose correctly. On the other hand, it avoids overfitting and it did well for the mixed states too. AIC and BIC have mistakes even for the large number of n. The comparison between tables 1 and 2 will show the difference between using semi-density matrices and common traditional models.

In the next example, we will show a concrete example of calculating a universal quantum source and predicting the $n+1$-th outcome by a quantum strategy.

Table 1. AIC and BIC Model Selection.

		Measurement Repetition				
		10	50	100	250	500
State 1	BIC	987	990	994	992	996
	AIC	945	944	919	927	930
State 2	BIC	25	83	183	394	706
	AIC	77	312	502	802	942
State 3	BIC	384	973	998	997	988
	AIC	594	992	998	997	998

Performance of BIC and AIC model selection for 3 states: pure (state 1), almost pure (state 2), and mixed (state 3). This table is based on the results in [8].

Table 2. The quantum version of classical two-part code estimation.

	Measurement Repetition				
	10	50	100	250	500
State 1	1000	1000	1000	1000	1000
State 2	336	926	995	1000	1000
State 3	747	980	998	1000	1000

Performance of the quantum version of classical two-part code estimation for 3 states: ρ_1 (state 1), ρ_2 (state 2), and ρ_3 (state 3).

Example 6. Let $\mathcal{M}$ be the following quantum generalized model

$$\mathcal{M} = \left\{ \rho_1 = \frac{1}{12}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \rho_2 = \frac{4}{12}\begin{bmatrix} \frac{2}{3} & \frac{2}{9} \\ \frac{2}{9} & \frac{1}{3} \end{bmatrix}, \rho_3 = \frac{7}{12}\begin{bmatrix} \frac{1}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} \end{bmatrix} \right\},$$

and we want to predict the $n + 1$-th outcome, after observing n measurements. Based on what we said in the previous sections the $Q-$ universal quantum source is as follows:

$$q_I^{(n)} \bar{\rho}^{(n)} q_I^{(n)} = \frac{1}{12}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{(n-k)} + \frac{4}{12}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{(n-k)} + \frac{7}{12}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{(n-k)}.$$

The quantum strategy associated with the above universal model is

$$\begin{aligned} &\hat{\rho}^{n+1}(|0\rangle\langle 0||q_I^{(n)}) \\ &= \frac{\frac{1}{12}\left(\frac{1}{2}\right)^{k+1}\left(\frac{1}{2}\right)^{(n-k)} + \frac{4}{12}\left(\frac{2}{3}\right)^{k+1}\left(\frac{1}{3}\right)^{(n-k)} + \frac{7}{12}\left(\frac{1}{4}\right)^{k+1}\left(\frac{3}{4}\right)^{(n-k)}}{\frac{1}{12}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{(n-k)} + \frac{4}{12}\left(\frac{2}{3}\right)^{k}\left(\frac{1}{3}\right)^{(n-k)} + \frac{7}{12}\left(\frac{1}{4}\right)^{k}\left(\frac{3}{4}\right)^{(n-k)}} \\ &= \frac{1}{12} \times \frac{6^{n+1} + 2^{2n+k+5} + 7 \times 3^{2n-k-1}}{6^n + 2^{2n+k+2} + 7 \times 3^{2n-k}}. \end{aligned}$$

Acknowledgments

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.

ORCID iDs

Ahmad Shafiei Deh Abad https://orcid.org/0000-0002-6844-8931

References

[1] Rissanen J 2007 *Information and Complexity in Statistical Modeling* (Berlin: Springer)
[2] Rissanen J 1986 *Stochastic Complexity and Modeling, Annals Statistics* **14** 1080–1100
[3] Grunewald P D 2007 *The Minimum Description Length Principle* (Massachusetts: The MIT Press)
[4] Benatti F 2009 *Dynamics, Information and complexity in Quantum Systems* (Berlin: Springer)
[5] Nielsen M and Chuang I 2010 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[6] Athanasios C 2012 Sobolev Inequalities on Complete Riemannian Manifolds *Master Thesis* University of Crete
[7] Ivancevich V G and Ivancevich T T 2007 *Applied Differential Geometry, A modern Introduction* (Singapore: World Scientific)
[8] Guta M, Kypraios T and Dryden I 2012 Rank-based model selection for multiple ions quantum tomography *New Journal of Physics* vol 14 (London: IOP Publishing and Deutsche Physikalische Gesellschaft) (https://doi.org/10.1088/1367-2630/14/10/105002)