Quantum Statistical Inference

Ahmad Shafiei Deh Abad, Mohammad Shahbazi
shafiei@khayam.ut.ac.ir, mshahbazi@ut.ac.ir

July 17, 2018

Abstract

In this paper, inspired by the “Minimum Description Length Principle” in classical Statistics, we introduce a new method for predicting the outcomes of a quantum measurement and for estimating the state of a quantum system with minimum quantum complexity, while, at the same time, avoiding overfitting.

1 Introduction

Let A be a quantum system. There are important problems about it: What is the state of the system? How can we know something about it? The only method at our disposal is measurement. Since quantum mechanics is statistical in nature we have to do measurement many times. But as we know after performing any measurement the state of the system changes drastically. To solve the problem, instead of one quantum system we consider a large number of quantum systems all prepared in the same state and perform measurement on each of them by the same measurement scheme. Other important problem is: Assume that we have n quantum systems prepared in the same state and we do measurement on each of them at the state by

*Corresponding Author
†Department of Mathematics, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran
the same measurement scheme and we obtain a sequence of outcomes. Now assume that we have another quantum system prepared in the same state, and we want to know the outcome of performing measurement on the system by the same measurement scheme without doing it.

As we have said before, quantum states are generically perturbed by measurements, we have to perform measurement on many system all prepared in the same state. Then, the problem arises of what we can infer from the outcomes of \( n \) such measurement procedures about the \( n + 1 \)-th of them. In order to solve these problems, we have to base our predictions on the information content of the first \( n \) measurement outcomes.

In the case of a binary string its information content is its Kolmogorov complexity. But we cannot use it here. Because, the Kolmogorov complexity is not computable. Moreover, since our work is in the realm of operator algebra, we prefer not to use coding theory.

In classical Statistics, to overcome the non computability of the Kolmogorov complexity, one uses a parametrized class of probability distributions and assume that the actual probability distribution belongs to this class. Then, by the maximum likelihood method or other methods, one may try to estimate the actual state. In this context, two problems appear:

1. when we do not know the actual state of the system, we do not know exactly which class of models it belongs to;

2. One can always perfectly reproduce a set of data by means of a wrong model, with hardly any predictive power, by inserting suitably many parameters, a phenomenon called overfitting.

To overcome these difficulties, J. Rissanen proposed the so called Minimum Description Length Principle in [26], further developed in [25], that has been used in many different fields: This method is based on generalized coding theory and on choosing an appropriate statistical model. In analogy with this approach, in the following we address quantum statistical inference by choosing an appropriate quantum model and an appropriate measurement scheme.

We emphasize that, even though highly inspired by the MDL principle, we do not use coding theory and we try to give a method for Quantum Statistical Inference avoiding the above mentioned difficulties. In the same vein, notations, definitions and conventions used in the quantum realm are
directly inspired by their classical counterparts in [10], so that comparison of quantum and classical frameworks should be straightforward.

We mention that the theorem 5.1 of [10] is proved for countable statistical model but our Theorem 4 generalizes it to arbitrary quantum model. Moreover, Proposition 15.1 of [10] is proved in one direction but the part one of Theorem 8 in this paper is proved in two directions.

2 A brief introduction to MDL Principle

The minimum description length (MDL) principle is a powerful method of inductive inference, the basis of 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 well-suited 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. More precisely, given a probabilistic model, Shannon’s coding theorems tell you the minimal number of bits needed to encode your data, i.e., the maximum extent to which it can be compressed. Really, however, to complete the description, you need to specify the model as well, from among some set of alternatives, and this will also require a certain number of bits. Hence you really want to minimize the combined length of the description of the model, plus the description of the data under that model. This works out to being a kind of penalized maximum likelihood: the data-given-model bit is the negative log likelihood, and the model-description term is the penalty [10].

3 Quantum preliminaries

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 of trace class operators of $\mathbb{H}$ will be denoted by $B_T(\mathbb{H})$, with
scalar product defined by
\[
\langle T|S \rangle_1 = \text{Tr}(T^*S), \quad \forall T, S \in B_T(\mathbb{H})
\]
with associated norm \(\|T\|_1 = \sqrt{\text{Tr}(T^*T)}\).

Let \(T\) and \(S\) be in \(B(\mathbb{H})\) and \(\lambda\) be a complex number; then, the ratio \(T/S\) will be defined to be \(T/S = \lambda\), if \(T = \lambda S\), undefined otherwise.

The n-times tensor product of the Hilbert space \(\mathbb{H}\) with itself will be denoted by \(\mathbb{H}^{(n)} := \bigotimes_{j=1}^{n} \mathbb{H}\).

Let \((S_i)_{i \in I}\) be a family of subsets of \(B(\mathbb{H})\). Then by \(\prod_{i \in I} S_i\) we mean the set \(\{\prod_{i \in I} s_i | s_i \in S_i\}\).

The set of all partitions of \(X\) will be denoted by \(\mathcal{P}(X)\).

### 3.1 Q-Projection

Let \(\mathbb{H}\) be a Hilbert space. The collection of all complete sets of mutually orthogonal (minimal) projections \(P = \{p_1, p_2, \ldots\}\) onto \(\mathbb{H}\), with \(\sum_{n=1}^{\infty} p_n = 1\) (completeness), will be denoted by \(\pi(\mathbb{H})\) \((\pi_0(\mathbb{H}))\).

**Definition 1**

Let \(P\) and \(Q\) be in \(\pi(\mathbb{H})\). We say that \(P\) is **finer** than \(Q\), and we write \(P \succeq Q\) if for each two elements \(p \in P\) and \(q \in Q\), \(pq\) is equal to \(p\) or \(0\).

We say that \(Q\) and \(P\) are **consistent** if they have a common upper bound with respect to this order relation. A subset of \(\pi(\mathbb{H})\) is called consistent if any two of its elements are consistent. we say that a consistent set \(A\) is maximally consistent if there is no consistent set \(B\) such that \(A \subsetneq B\).

**Definition 2**

Assume that \((X_j)_{j \in J}\) is a family of subsets of a nonempty set \(X\) and \(X = \bigcup_{j \in J} X_j\).

1. We say that the set \(Y = \bigcup_{i \in I < J} X_i\) is a maximally connected union of the family \((X_j)_{j \in J}\) if it is a largest subset of the set \(X\) with the following property:

   For each proper subset \(K\) of \(I\),

   \[
   \bigcup_{k \in K} X_k \cap \bigcup_{i \in I - K} X_i \neq \emptyset.
   \]
The set of all maximally connected unions of the family \((X_j)_{j \in J}\) will be denoted by \(\bigwedge_{j \in J} X_j\). Clearly, \(\bigwedge_{j \in J} X_j\) is a partition of \(X\).

2. A nonempty subset \(Z\) of \(X\) will be called a maximally connected intersection of the family \((X_j)_{j \in J}\), if there exists an element \(z \in X\) such that \(Z\) is the intersection of all elements of the family containing \(z\). The set of all maximally connected intersections of the family \((X_j)_{j \in J}\) will be denoted by \(\bigvee_{j \in J} X_j\).

Now, assume that \(P = \{P_i | i \in I\}\) is a set of partitions \(P_i = \{P_{ij} | j \in J_i\}\) of the set \(X\). Then, clearly \(\bigwedge_{j \in J, i \in I} P_{ij} \leq P_i \leq \bigvee_{j \in J, i \in I} P_{ij}\).

Let partitions \(P\) and \(Q\) of \(X\) be such that for all \(i \in I\) we have \(Q \leq P_i \leq P\). Then, it is straightforward to see that

\[
\bigvee_{j \in J, i \in I} P_{ij} \leq P_i \leq \bigvee_{j \in J, i \in I} P_{ij} \leq \bigvee_{j \in J, i \in I} P_{ij}.
\]

Therefore, \(\bigwedge_{j \in J, i \in I} P_{ij}\) is the greatest lower bound and \(\bigvee_{j \in J, i \in I} P_{ij}\) is the least upper bound of the set \(P\).

**Lemma 1**

Let \(\mathbb{P} = \{P_i \in \pi(\mathbb{H}), i \in I\}\). Then

1. If the set is consistent it has a least upper bound and a greatest lower bound.

2. The set is consistent if and only if for any transformation \(\sigma\) of \(I\) we have \(\Pi_{i \in I} P_{ij} = \Pi_{i \in I} P_{\sigma(i)}\).

**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 \(R \succeq Q\). Let \(0 \neq q \in Q\) and let \(R_q\) be the sum of all elements of \(r \in R\) such that \(q \geq r\). Clearly \(R_q^2 = R_q \neq 0\) and \(qR_q = R_qq = R_q\) 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 \(q \geq r\) 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 mapping $Q \to \bar{Q}$ from $\pi(\mathbb{H})$ into $\mathbb{P}(R)$ be defined as follows, for each $q \in Q$, $q \to \bar{q}$, where $\bar{q}$ is the set of all summands of the projection $q$. Notice that $\bar{q}$ is the sum of some elements of $R$. Now it is clear that under this mapping we have the following bijective maps.

\[
\land_{j \in J, i \in I} P_{ij} \rightarrow \land_{j \in J, i \in I} P_{ij}
\]
\[
\lor_{j \in J, i \in I} P_{ij} \rightarrow \lor_{j \in J, i \in I} P_{ij}
\]

As a consequence of what we said above $\land_{j \in J, i \in I} P_{ij}$ (resp. $\lor_{j \in J, i \in I} P_{ij}$) is the greatest lower (resp. least upper) bound of $\mathbb{P}$.

2. Assume that the set $\mathbb{P}$ is commutative. Then, there exists a unique complete set of mutually orthogonal minimal projections $R = \{r_1, r_2, \ldots\}$ such that for each $i \in I$ we have $P_{ij} = \sum_{k=1}^{\infty} \epsilon_{ijk} r_k$. Where $\epsilon_{ijk}$ is 0 or 1. Therefore, the set $R$ is an upper bound of the family $(P_i)_{i \in I}$ and the set $\mathbb{P}$ is consistent.

Conversely, assume that $\mathbb{P}$ is consistent then as we saw above for each $Q \in \mathbb{P}$ each $q \in Q$ is the sum of some elements of $R$. Therefore $\mathbb{P}$ is commutative.

Let $\mathcal{M}$ be a subset of $B(\mathbb{H})$ and $\Pi$ be a subset of $\pi(\mathbb{H})$. For each element $\pi$ of $\Pi$ the subset $\ker(\pi) \cap \mathcal{M}$ of $\mathcal{M}$ will be denoted by $\ker_{\mathcal{M}}(\pi)$ and the set $\{\ker_{\mathcal{M}}(\pi) | \pi \in \Pi\}$ will be denoted by $\mathcal{M}$. Sometimes we need to know if there is an element $\pi$ of $\Pi$ such that $\ker_{\mathcal{M}}(\pi)$ is a minimal element of the set $\mathcal{M}$ or not? As we will see in Lemma 5, if the partially ordered set $\Pi$ has no minimal element the answer to the above question is ”no”. This fact and the so called “Axiom of choice” motivate the definition of “chain”.

**Definition 3**

*With the order relation $\succeq$, $\pi(\mathbb{H})$ is a partially ordered set.*

Any well ordered subset of this partially ordered set will be called a chain.

We say that the chain $cc'$ is the concatenation of the chains $c$ and $c'$ if $cc'$ is equal to the union of $c$ and $c'$ and the maximal element of $c$ is less than the minimal element of $c'$.

We say that the chain $c_1$ is a subchain of the chain $c_2$, if the set $c_1$ is a subset of the set $c_2$. A chain $c$ will be called complete if there is no chain $c' \neq c$ such that $c$ is a subchain of $c'$.
Example 1
Let $\mathbb{H}$ be the state space of $n$ qubits and let $Q = Q_0 = \{q_1, q_2, \ldots, q_{2^n}\}$ be a complete mutually orthogonal minimal projections of $\mathbb{H}$. Then the set $\{Q_k|0 \leq k \leq n-1\}$, where $Q_k = \{q_{k,l}|0 \leq l \leq 2^{n-1}-1\}$ and $q_{k,l} = \sum_{j=1}^{2^{n-k}} q_{j+l2^{n-k}}$ is a chain, which is generally not complete.

Example 2
Let $\mathbb{H}$ be a Hilbert space and $Q = Q_0 = \{q_1, q_2, \ldots, \} \in \pi(\mathbb{H})$ be a complete set of mutually orthogonal minimal projections of $\mathbb{H}$. For $k \in \mathbb{N}$ let $Q_k$ be the complete set of mutually orthogonal projections defined as follows,

1. $q_{k,l} = \sum_{l=0}^{k} q_l$ for $0 \leq l < k$
2. $q_{k,l} = q_{k+1,l}$ for $l > k$.

It is not difficult to show that $\{Q_k|l \in \mathbb{N}\} \cup \mathbb{I}_\mathbb{H}$ is a complete chain.

If a chain is complete, its minimum element is the identity mapping of $\mathbb{H}$ and its maximum element is a complete set of mutually orthogonal minimal projections of $\mathbb{H}$.

Definition 4
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 [3]) and for each $q \in Q$

$$T_q = q T q$$

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

The set $B_Q(\mathbb{H})$ is a complex subspace of the $C^*$-algebra $B(\mathbb{H})$, and the mapping $Q$ from $B(\mathbb{H})$ into $B_Q(\mathbb{H})$ defined by $Q(T) := T_Q$ 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 a complete set 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. \( \text{Tr}(T_Q) = \sum_{n=1}^{\infty} \text{Tr}(q_n T q_n) = \sum_{n=1}^{\infty} \text{Tr}(q_n T) = \text{Tr}(T) \), since the sets \( Q \in \pi(H) \) are chosen to be complete (see Definition 1).

2. If \( T = T^* \), then \( (T_Q)^* = T_Q \).

3. Let \( T \geq 0 \); then, \( q T q \geq 0 \) for each \( q \in Q \) so that \( T_Q \geq 0 \) for each \( Q \in \pi(\mathbb{H}) \). Vice versa, if \( T_Q \geq 0 \) for each \( Q \in \pi(\mathbb{H}) \), then, for each vector \( |v\rangle \in \mathbb{H} \), \( \text{Tr}(|v\rangle \langle v| T) = \langle v| T |v\rangle \geq 0 \), since any such \( |v\rangle \langle v| \) belongs to some \( Q \in \pi(\mathbb{H}) \). Therefore, \( T \geq 0 \).

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 trace-norm preserving convex map from \( D(\mathbb{H}) \) onto \( D_Q(\mathbb{H}) \).

**Lemma 3**
1. The mapping \( Q : B(H) \rightarrow B_Q(H) \) is continuous.

2. The mapping \( Q : B_T(H) \rightarrow B_Q(H) \) is continuous in the trace norm topology.

**Proof.**

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

\[
\| T Q \| = \left\| \sum_n q_n T q_n \right\| = \| q T q \| \leq \| T \| .
\]

Since any \( T \in B(H) \) can be written as the sum of two self adjoint elements \( Q \) is continuous.

2. Let \( T \in B_T(H) \). Then, for each \( q \in Q \), \( q T q T^* q \leq q T T^* q \). Therefore, \( (T_Q)^* T_Q = ((T^*)_Q) T_Q \leq (T^* T)_Q \). Since \( \text{Tr}(T_Q) = \text{Tr}(T) \),

\[
\| T_Q \|_1 = \sqrt{\text{Tr} \left( (T_Q)^* T_Q \right)} \leq \sqrt{\text{Tr} \left( (T^* T)_Q \right)} = \| T \|_1 .
\]

**Lemma 4**

For each element \( T \in B(H) \) and each \( Q = \{ q_1, q_2, \ldots \} \in \pi(H) \) we have:

1. \( T = T_Q \) if and only if for each \( q_n \in Q \) we have \( q_n T = T q_n \).

2. Let \( S = S_Q \) and for all \( q_n \in Q \), \( q_n S q_n = q_n T q_n \). 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. ■

Remark: Let \( Q = \{ q_1, q_2, \ldots \} \) and \( P = \{ p_1, p_2, \ldots \} \) be two complete sets of mutually orthogonal projections of the Hilbert space \( \mathbb{H} \) such that for each \( i \), the mapping \( q_i : p_i \mathbb{H} \to q_i \mathbb{H} \) is an isomorphism of complex vector spaces, and for \( i \neq j \) we have \( \lambda_i \neq \lambda_j \). If 
\[
\sum_{i=1}^{\infty} \lambda_i q_i = T = \sum_{i=1}^{\infty} \lambda_i p_i,
\]
then for each \( i \) we have \( p_i = q_i \).

This is a consequence of the theorem of spectral decomposition.

Lemma 5

Let \( T \in B(\mathbb{H}) \) and \( P, Q \in \pi(\mathbb{H}) \). If \( P \succeq Q \) then:

1) \( T_P = (T_Q)_P = (T_P)_Q \).

2) \( \text{Ker}(Q) \subset \text{Ker}(P) \).

Proof. It is clear that for each element \( p \in P \) there exists exactly one element \( q_0 \in Q \) such that \( q_0 p = pq_0 = p \) 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, 
\[
(T_Q)_P = \sum_{p \in P} pTQp = \sum_{p \in P} pTp = T_P
\]

On the other hand for each \( q \in Q \) and each \( p \in P \) we have 
\[
qT_P = q\sum_{p \in P} pTp = \sum_{p \in P} qppTp
= \sum_{p \in P \mid qp \neq 0} pTp = \sum_{p \in P} pTpq
= T_P q.
\]

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

Since \( \bar{P}(T) = T_P = (T_Q)_P = \bar{P}(Q(T)) \), 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 \( T_P \) is a pseudo-spectral decomposition of the operator \( T \). Then for each \( S \in B(\mathbb{H}) \), we have
\[
(ST)_P = S_P T_P \quad \text{and} \quad (TS)_P = T_P S_P.
\]

Proof. We have \((ST)_P = (ST_P)_P\). Therefore, for each \( p \in P \) we have
\[
p(ST)p = p(ST_P)p = pS_PT_P = (pS_P)(pT_P). \quad \text{Therefore,} \quad (ST)_P = S_P T_P. \quad \text{The proof of the other equality is the same.}
\]

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 unitary \( C^* \)-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.
5. A necessary and sufficient condition for \( B_Q(\mathbb{H}) \) to be commutative is that for all \( i \in \mathbb{N} \) the algebra \( B_{\bar{q}_i}(\mathbb{H}) \) be commutative.

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

Definition 5
Let \( R \) be an n-ary relation on \( B(\mathbb{H}) \) and let \( T_1, T_2, \ldots, T_n \in B(\mathbb{H}) \), \((n \text{ may be infinity})\). We say that \( R(T_1, T_2, \ldots, T_n) \) is weakly true if, for each minimal projection \( q \), \( R(T_{1q}, T_{2q}, \ldots, T_{nq}) \) is true.
For example, any two elements of $B(\mathbb{H})$ always weakly commute. For some relations, being true or weakly true are equivalent. For example, if $T \geq S$ then clearly, this relation is weakly true. Conversely let for each minimal projection $q$, $T_q \geq S_q$. Then for each vector $v \in \mathbb{H}$ we have $\langle v|T - S|v \rangle \geq 0$. Therefore, $T - S \geq 0$. The relation weakly equal will be denoted by $=^w$.

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

**Definition 6**

Let $\mathbb{H}$ be a separable Hilbert space and $Q \in \pi_0(\mathbb{H})$. The mapping $\nu : B(\mathbb{H}) \rightarrow \mathbb{R}$ 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 nonclassical.

**Example 3**

Let $\mathbb{H}$ be a 2-dimensional Hilbert space with the standard basis $\{|0\rangle, |1\rangle\}$. Let $X, Y, Z$ be Pauli 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 nonclassical.

### 4 Quantum Source and Quantum Strategy

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

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

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

$$
dvol_\mathcal{M} = |g|^{1/2}dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n
$$

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

The **canonical measure** of $\mathcal{M}$ is the **counting measure** when $\mathcal{M}$ is “countable” and is the measure induced by the **volume element** $dvol_\mathcal{M}$ when, $\mathcal{M}$ is a “Riemannian submanifold” of $B_T$. 

12
Definition 7
Let $\mathbb{H}$ be a separable Hilbert space and let $\rho \in B_+(\mathbb{H})$. We call $\rho$ a semi-density matrix if $\text{Tr}(\rho) \leq 1$.

Definition 8

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

2. More generally, if $\mathcal{M}$ consists of semi-density operators, then $(\mathcal{M}, \Sigma, \mu)$ or simply $\mathcal{M}$ will be called a generalized quantum model on $\mathbb{H}$.

3. For each $Q \in \pi(\mathbb{H})$, the set $\mathcal{M}_Q = \{\rho_Q | \rho \in \mathcal{M}\}$ is a $Q$-(generalized) quantum model.

Remark 1
Let $\mathcal{M}$ be a generalized quantum model consisting of nonzero semi-density matrices and let $n$ be a nonzero positive integer. Then, for each $\rho \in \mathcal{M}$ we set

$$\rho^{(n)} := \rho \otimes \left( \frac{\rho}{\text{Tr}(\rho)} \right) \otimes \cdots \otimes \left( \frac{\rho}{\text{Tr}(\rho)} \right).$$

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

Let $\mathcal{M}'$ be a generalized quantum model which does not contain the trivial semi-density matrix $0$. Then, the set $\mathcal{M} = \{\frac{\rho}{\text{Tr}(\rho)} | \rho \in \mathcal{M}'\}$ is a quantum model and the mapping $\rho \rightarrow \frac{\rho}{\text{Tr}(\rho)}$ from $\mathcal{M}'$ into $\mathcal{M}$ will be denoted by $\omega$.

Definition 9
The generalized quantum model $\mathcal{M}$ will be called Bayesian if $\int_{\mathcal{M}} \rho d\mu(\rho)$ exists and is a semi-density matrix.

Definition 10

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

\[
tr_n(\tilde{\rho}^{(n)}) = \tilde{\rho}^{(n-1)}.
\]

**Notation:** In this paper, we will denote any (generalized) quantum source by \( \tilde{\rho} = (\tilde{\rho}^{(n)})_{n \in \mathbb{N}} \) where \( \tilde{\rho}^{(n)} \in B_T(\mathbb{H}^{(n)}) \).

Let \( Q = \{q_1, q_2, \ldots\} \) be a complete set of mutually orthogonal projections of \( \mathbb{H} \). By setting \( I = (i_1, i_2, \ldots, i_n) \in \mathbb{N}^n \) the projection \( \otimes_1^n q_i \) on \( \mathbb{H}^{(n)} \) will be denoted by \( q_i^{(n)} \) or simply by \( q_i \).

**Lemma 8**

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

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

\[
tr_{n+1}(\tilde{\rho}^{(n+1)}) = \int_\mathcal{M} tr_{n+1}(\rho^{(n+1)}) d\mu(\rho) = \int_\mathcal{M} \rho^{(n)} d\mu(\rho) = \tilde{\rho}^{(n)}.
\]

\[\blacksquare\]

**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 \( T_1 \in B_+(\mathbb{H}_1) \).

In this case for simplicity the operator \( (T_1^{1/2} \otimes I_2)T(T_1^{1/2} \otimes I_2) \) will be denoted by \( T_1.T \). Here \( I_2 \) is the identity mapping of \( \mathbb{H}_2 \).

**Definition 11**

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

\[
\rho_{(1)} = \rho_1^{-1}\rho
\]

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

\[
\rho_1.\rho_{(1)} = \rho.
\]
Remark 2

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 9

Let $Q \in \pi_0(\mathbb{H}_1)$ and $P \in \pi_0(\mathbb{H}_2)$. Assume that the density matrix $\rho$ 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

$$\rho_i = tr_1(q_i \cdot \rho_{(1)})$$

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 $\text{Tr}(\rho_i)$ is equal to 1. Clearly, we have

$$\rho_1 = tr_2(\rho) = \sum_{i=1}^{\infty} (\sum_{j=1}^{\infty} \lambda_{ij}) q_i$$

$$\implies (\rho_1)^{\frac{1}{2}} = \sum_{i=1}^{\infty} (\sum_{j=1}^{\infty} \lambda_{ij})^{\frac{1}{2}} q_i.$$ 

Therefore,

$$\rho_{(1)} = (\rho_1^{-\frac{1}{2}}) \cdot \rho = (\rho_1^{-\frac{1}{2}} \otimes I) \rho (\rho_1^{-\frac{1}{2}} \otimes I)$$

$$= ((\sum_{i=1}^{\infty} (\sum_{k=1}^{\infty} \lambda_{ik})^{-\frac{1}{2}} q_i) \otimes I) \times (\sum_{i,j=1}^{\infty} \lambda_{ij} (q_i \otimes p_j)) \times ((\sum_{i=1}^{\infty} (\sum_{k=1}^{\infty} \lambda_{ik})^{-\frac{1}{2}} q_i) \otimes I)$$

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

$$= \sum_{i=1}^{\infty} (q_i \otimes \rho_{(i)})$$

Where,

$$\rho_i := \sum_{j=1}^{\infty} ((\sum_{k=1}^{\infty} \lambda_{ik})^{-1}) \lambda_{ij} p_j.$$

Therefore $\rho_i$ is a density matrix. $\blacksquare$
Lemma 10
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 \( i \in \mathbb{N} \), \( \text{tr}_1(q_i.T) \) be a density matrix on \( \mathbb{H}_2 \), and
\[ \rho_1 = \sum_{i=1}^{\infty} \lambda_i q_i \]
be a density matrix on \( \mathbb{H}_1 \). Then \( \rho = \rho_1.T \) is a density matrix on \( \mathbb{H}_1 \otimes \mathbb{H}_2 \), moreover, \( T = \rho||1). \)

Proof. Clearly, \( \rho \) is a positive operator of \( \mathbb{H}_1 \otimes \mathbb{H}_2 \). On the other hand,
\[ \text{Tr}(\rho) = \text{Tr}(\rho_1.T) = \sum_{i,j=1}^{\infty} \lambda_i \lambda_{ij} = \sum_{i=1}^{\infty} \lambda_i \sum_{j=1}^{\infty} \lambda_{ij} = \sum_{i=1}^{\infty} \lambda_i = 1. \]
Therefore, \( \rho \) is a density matrix on \( \mathbb{H}_1 \otimes \mathbb{H}_2 \). Now,
\[ T = \rho_1^{-1} \cdot \rho = \rho||1). \]

Since for all \( Q \in \pi_{0}(\mathbb{H}) \), \( B_Q(\mathbb{H}^{(2)}) \) is a commutative \( \mathcal{C}^* \)-algebra, in general,
\[ T \in B_T(\mathbb{H}^{(2)}) \text{ and } (\text{tr}_2 T) \otimes I \text{ weakly commute.} \]

Convention 2
Let \( \bar{\rho} = (\bar{\rho}^{(n)})_{n\in\mathbb{N}} \) be a quantum source. Then, for each \( n \in \mathbb{N} \), the operator \( (\bar{\rho}^{(n-1)})^{-1} \cdot \bar{\rho}^{(n)} \) will be denoted by \( \bar{\rho}_{n\cdot(n-1)} \).

Remark 3
We set \( \otimes^0 \mathbb{H} = \{0\} \). Clearly, the only linear transformation of this vector space is the identity \( I_0 \). For consistency we set \( \text{Tr}(I_0) = 1 \). Moreover, for each \( T \in B(\mathbb{H}), T^0 = I_{\mathbb{H}} \).

Lemma 11
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

$$\bar{\rho}^{(n+1)} = \sum_{i,j} R_{i,j} \otimes \langle i | j \rangle$$

where $R_{i,j} \in B(\mathbb{H}^{(n)})$. Because $\bar{\rho}^{(n+1)}$ is a prequential quantum source we have

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

So,

$$(U \bar{\rho} U^\dagger)^{(n+1)} = U^{(n+1)} \bar{\rho}^{(n+1)} (U^\dagger)^{(n+1)}$$

$$= \sum_{i,j=1}^\infty (U^{(n)} R_{i,j} (U^\dagger)^{(n)}) \otimes U| i \rangle \langle j | U^\dagger.$$ 

Therefore,

$$\text{tr}_{n+1}(U \bar{\rho} U^\dagger)^{(n+1)} = \sum_{i,j=1}^\infty (U^{(n)} R_{i,j} (U^\dagger)^{(n)}) \text{Tr}(U| i \rangle \langle j | U^\dagger)$$

$$= \sum_{i=1}^\infty (U^{(n)} R_{i,i} (U^\dagger)^{(n)})$$

$$= U^{(n)} (\sum_{i=1}^\infty R_{i,i}) (U^\dagger)^{(n)}$$

$$= U^{(n)} \bar{\rho}^{(n)} (U^\dagger)^{(n)} = (U \bar{\rho} U^\dagger)^{(n)}$$

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

**Definition 12**

Let $\mathcal{M}$ be a (generalized) quantum model and $\bar{\rho}$ be a (generalized) 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 $\mathcal{M}$** if for each $\rho \in \mathcal{M}$ and for each $\epsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that for all $n > n_0$ we have:

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

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

$$D(\rho^{(n)} || \bar{\rho}^{(n)}) \leq n\epsilon.$$ 

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 $n > n_0$ we have:

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

17
In the above if $\epsilon$ does not depend on $\rho$, $\bar{\rho}$ is called uniformly (Q-)universal.

**Lemma 12**
With the above notations and conventions 1 implies 2.

**Proof.** Clearly we have
\[
\bar{\rho}^{(n)} - 2^{-n \epsilon} \rho^{(n)} \geq 0 \\
\Rightarrow n \epsilon + \log \bar{\rho}^{(n)} - \log \rho^{(n)} \geq 0 \\
\Rightarrow n \epsilon + (\rho^{(n)})^{1/2} \log \bar{\rho}^{(n)} - \log \rho^{(n)} (\rho^{(n)})^{1/2} \geq 0 \\
\Rightarrow n \epsilon + \text{Tr} \rho^{(n)} (\log \bar{\rho} - \log \rho) \geq 0. \\
\Rightarrow D(\rho^{(n)} \| \bar{\rho}^{(n)}) \leq n \epsilon.
\]

\[\blacksquare\]

**Example 4**
Let $\mathcal{M}$ 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 $\rho^* \in \mathcal{M}$ and each $n \in \mathbb{N}$ we have
\[
\bar{\rho}^{(n)} = \sum_{\rho \in \mathcal{M}} \rho^{(n)} \geq \rho^*(n).
\]
Now let $\epsilon$ be given and let $n_0 \in \mathbb{N}$ be such that
\[
\text{Tr}(\rho^*) \geq 2^{-(n_0)\epsilon}.
\]
Let $\hat{\rho}$ be the density matrix associated with $\rho^*$. Then, for each $n \geq n_0$ we have
\[
\bar{\rho}^{(n)} - 2^{-n \epsilon} \hat{\rho}^{(n)} \geq 0.
\]
Therefore, $\bar{\rho}$ is universal for $\mathcal{M}'$.

**Example 5**
Let $Q \in \pi_0(\mathbb{H})$ and $\mathcal{M}$ be a quantum model on the Hilbert space $\mathbb{H}$. For each $I \in \mathbb{N}^n$ let
\[
T^{(n)} = \sum_{I \in \mathbb{N}^n} (\max_{\rho \in \mathcal{M}} q_I^{(n)} \rho^{(n)} q_I^{(n)})^{(n)}.
\]
Clearly, $T^{(n)} \in B(\mathbb{H}^n)$. Assume that there exists $K > 0$ such that $\text{Tr}(T^{(n)}) < K$. Let $\bar{\rho}^{(n)} = T^{(n)}/\text{Tr}(T^{(n)})$. It is straightforward to see that $\bar{\rho} = (\bar{\rho}^{(n)})_{n \in \mathbb{N}}$ is a uniformly universal quantum source relative to $\mathcal{M}$.
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 submanifold of $B_{T}(\mathbb{H})$, consisting of regular density matrices. Assume that

1. $p: \mathcal{M} \rightarrow ]0, \infty[\text{ is a continuous function.}$

2. There exists a positive real $c > 0$ such that for all $q \in Q$, $q\rho q \geq cq$.

Where $\rho_q \in \mathcal{M}$ is such that we have

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

Moreover, for each $n \in \mathbb{N}$ let $A^{(n)} = \int_{\mathcal{M}} p(\rho) \rho^{(n)} d\text{vol}_{\mathcal{M}}(\rho)$, and $A^{(1)} = A$. Then the sequence $(A^{(n)})_{n \in \mathbb{N}}$ 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

$$\text{Tr}(q\rho_q q) - \text{Tr}(q\rho_q^* q) = \text{Tr}(q(\rho_q - \rho_q^*) q) = \|q(\rho_q - \rho_q^*) q\|$$

$$\leq \|\rho_q - \rho_q^*\| \leq \|\rho_q - \rho_q^*\|_1 \leq d(\rho_q, \rho_q^*) \leq \delta.$$ 

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

$$\text{vol}_{\mathcal{M}} B(\rho, \delta) \geq v.$$ 

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

$$qAq = \int_{\mathcal{M}} p(\rho)q\rho q d\rho \geq \int_{B(\rho_q, \delta)} p(\rho)q\rho q d\rho$$

$$\geq 2^{-\epsilon/2} \beta v \text{vol}(B(\rho_q, \delta)) q\rho q q \geq 2^{-(k+1)\epsilon/2} q\rho q q.$$ 

Let us denote $\otimes^n \rho$ by $\tilde{\rho}^{(n)}$ and $\int_{\mathcal{M}} p(\rho) \rho^{(n)} d\text{vol}_{\mathcal{M}}(\rho)$ by $\tilde{A}^{(n)}$. 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
\[ q^{(n)} A^{(n)} q^{(n)} \geq 2^{-(k+n)/2} q^{(n)} \tilde{\rho}^{(n)} q^{(n)} \geq 2^{-n \epsilon} q^{(n)} \tilde{\rho}^{(n)} q^{(n)}. \]

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

\[ A^{(n)}_Q = \sum_{q(n) \in Q(n)} (q^{(n)} A^{(n)} q^{(n)}) \geq 2^{-n \epsilon} \sum_{q(n) \in Q(n)} (q^{(n)} \tilde{\rho}^{(n)} q^{(n)}) = 2^{-n \epsilon} \tilde{\rho}_Q. \]

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

\[ \blacksquare \]

**Corollary 2**

Let \( \mathbb{H} \) be a finite dimensional Hilbert space and \( Q \in \pi_0(\mathbb{H}) \). Let \( \mathcal{M} \) be a quantum model which is a compact Riemannian submanifold of \( B_T(\mathbb{H}) \) consisting of regular density matrices. Assume that \( p : \mathcal{M} \rightarrow [0, \infty) \) is a continuous function. Then, with the above notations the sequence \((A^{(n)})_{n \in \mathbb{N}}\) is \(Q\)-universal for \(\mathcal{M}\).

**Lemma 13**

\( S_{\mathcal{M}} \) 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 \( \tilde{\rho}_1 \) and \( \tilde{\rho}_2 \) 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 \( q \in Q \). Then there exists \( n_0 \in \mathbb{N} \) such that for \( k = 1, 2 \) we have:

\[ \tilde{\rho}_k^{(n)} - 2^{-n \epsilon} \rho^{(n)} \geq 0. \]

Let \( \alpha \) and \( \beta \) be two positive real numbers such that \( \alpha + \beta = 1 \). Then

\[ \alpha \tilde{\rho}_1^{(n)} + \beta \tilde{\rho}_2^{(n)} - 2^{-n \epsilon} \rho^{(n)} = \alpha (\tilde{\rho}_1^{(n)} - 2^{-n \epsilon} \rho^{(n)}) + \beta (\tilde{\rho}_2^{(n)} - 2^{-n \epsilon} \rho^{(n)}) \geq 0. \]

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

\[ (\alpha \tilde{\rho}_1 + \beta \tilde{\rho}_2)^{(n)} = \alpha \tilde{\rho}_1^{(n)} + \beta \tilde{\rho}_2^{(n)} \in (S_{\mathcal{M}})^{(n)}, \]

where \((S_{\mathcal{M}})^{(n)} = \{ \tilde{\rho}^{(n)} | \tilde{\rho} \in S_{\mathcal{M}} \} \). Therefore

\[ \alpha \tilde{\rho}_1 + \beta \tilde{\rho}_2 \in S_{\mathcal{M}}. \]

\[ \blacksquare \]
Definition 13
Let $Q = \{q_1, q_2, \ldots\}$ be a complete set of mutually orthogonal minimal projections of the Hilbert space $\mathbb{H}$ and for each $n \in \mathbb{N}$ let $\hat{\rho}^{(n)} = \hat{\rho}_Q^{(n)} \in B_+(\mathbb{H}^{(n)})$ be such that for all $n > 1$ and $q_i^{(n-1)} \in Q^{(n-1)}$, $\text{Tr}(q_i^{(n-1)}.(\hat{\rho}^{(n)})) = 1$. 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 $q_i^{(n-1)} \cdot (\hat{\rho}^{(n)})$ will be denoted by $\hat{\rho}_I^{(n)}$.

Theorem 3
Let $\hat{\rho}$ be a $Q$-quantum strategy. Let us define the sequence $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ recursively as follows:
1) $\hat{\rho}^{(1)} = \hat{\rho}^{(1)}$,
2) $\hat{\rho}^{(n+1)} = \hat{\rho}^{(n)} \cdot \hat{\rho}^{(n+1)}$.
Then the sequence $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ is a prequential $Q$-quantum source.
Conversely, let $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ be a prequential $Q$-quantum source. Let $\hat{\rho}^{(n+1)} = (\hat{\rho}^{(n)})^{-1} \cdot \hat{\rho}^{(n+1)}$. Then, $\hat{\rho}^{(n+1)} = \hat{\rho}^{(n+1)}$ and the sequence $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ is a $Q$-quantum strategy.

The proof is a direct consequence of Lemma 11 and Lemma 12.

Lemma 14
Let $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ be a $Q$-quantum strategy and $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ be its associated $Q$-quantum source. Then for each $T \in B(\mathbb{H})$, $T \hat{\rho} = \hat{\rho} T$ if and only if $T \hat{\rho} = \hat{\rho} T$.

The proof is straightforward.

Remark 4
For future applications we mention that because of the equality $\hat{\rho}^{(n+1)} = \hat{\rho}^{(n+1)}$, $Q$-quantum strategies are also called $Q$-quantum estimators. Let $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ be a prequential quantum source. It is straightforward to see that $(\hat{\rho}_Q^{(n)})_{n \in \mathbb{N}}$ is a prequential $Q$-quantum source and gives rise to a $Q$-quantum strategy.

Definition 14
A quantum estimator $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ is called good with respect to a quantum model $\mathcal{M}$ if its associated prequential source is universal for $\mathcal{M}$.
5 Quantum Exponential Families

The material in this section is extracted from [10] with necessary modifications.

Let $H$ be a separable Hilbert space, $n \in \mathbb{N}$ and $\Theta_{can} \subseteq \mathbb{R}^n$. Let $\psi : \Theta_{can} \rightarrow \mathbb{R}$ and $T = (T_1, T_2, ..., T_n)$, where for $n \geq k \geq 1$, $T_k$ is a bounded self-adjoint operator. Let $\rho_0$ be a regular density matrix of $H$. Assume that $\rho_{\theta} = \exp(\frac{\theta T - \psi(\theta)}{2})\rho_0 \exp(\frac{\theta T - \psi(\theta)}{2})$, where for $\theta = (\theta_1, \theta_2, ..., \theta_n) \in \Theta_{can}$, $\theta T = \sum_{k=1}^{n} \theta_k T_k$. and 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], [22].

In this paper for simplicity we assume that $\Theta_{can}$ is an open interval of $\mathbb{R}$ and $\psi$ is smooth on a neighborhood of the closure of $\Theta_{can}$. Now, assume that $Q \in \pi_0(H)$ and $T = T_Q$. From the equality $1 = \text{Tr}(\rho_{\theta Q}) = \text{Tr}(\exp(\theta T - \psi(\theta))\rho_0 Q)$, we have $\text{Tr}(\exp(\theta T)\rho_0 Q) = \exp(\psi(\theta))$. And from this we have

$$\frac{d}{d\theta} \psi(\theta) = \text{Tr}(T \exp(\theta T - \psi(\theta))\rho_0 Q) = E_\theta(T),$$

and $\text{Var}_\theta(T) = E_\theta(T^2) - (E_\theta(T))^2 = \frac{d^2}{d\theta^2} \psi(\theta) = I_\theta$. Where $I_\theta$ is the Fisher information of the associated $Q$-exponential family $\bar{Q}(\rho_\theta)$ at $\theta$. It is not difficult to prove that $I_\theta$ is strictly positive on $\Theta_{can}$.

Therefore, the mapping $\mu : \Theta_{can} \rightarrow \Theta_{mean} = \mu(\Theta_{can})$ defined by $\mu(\theta) = E_\theta(T)$ is a diffeomorphism. For each $t \in \Theta_{mean}$ let us denote $\rho_{\mu^{-1}(t)}$ by $\rho_t$.

Clearly we have \{\rho_\theta | \theta \in \Theta_{can}\} = \{\rho_t | t \in \Theta_{mean}\}.

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

$$\hat{t}_{(\alpha,t_0,I)} = \frac{\alpha t_0 + \sum_{k=1}^{n} \langle q_{ik} | T q_{ik} \rangle}{\alpha + n}.$$  

Assume that $\hat{t}_{(\alpha,t_0,I)}$ is in $\Theta_{mean}$ and $(\rho^{(n)})_{n \in \mathbb{N}}$ is the quantum strategy defined by

$$q_{t}^{(n-1)} \cdot \rho^{(n)} = \rho_{\hat{t}_{(\alpha,t_0,I)}}.$$
This quantum strategy is a good quantum estimator.
Let $(\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ 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

$$D(\rho^{(n)} \| \hat{\rho}^{(n)}) = O(\log(n)).$$

6 Quantum Prediction and Quantum Estimation

As we said in the introduction Quantum Prediction and Quantum Estimation are important subjects of Quantum Statistical Inference. To do this “Following the general MDL principle our strategy is invariably to design a universal quantum source relative to the given quantum model and base all inferences on it.” [10] page 470 with some modifications.

Quantum strategy $\hat{\rho}^{(n)}$:

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

$$q_I^{(n-1)} \hat{\rho}^{(n)} = \omega(\arg\max_{\rho \in \mathcal{M}} q_I^{(n-1)} \rho^{(n-1)} q_I^{(n-1)}).$$

Clearly, $\hat{\rho} = (\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ is a Q-quantum strategy for $\mathcal{M}$ the quantum model associated to $\mathcal{M}$.

In many cases, for example when $\mathcal{M}$ consists of a quantum exponential family, as we have mentioned in Section 6, $\hat{\rho}$ is a good Q-quantum strategy for $\mathcal{M}$. The Q-quantum strategy $\hat{\rho} = (\hat{\rho}^{(n)})_{n \in \mathbb{N}}$ enables us to predict the $n-th$ outcome of the measurement $Q$ knowing the $(n-1)$ previous ones. Moreover, we can consider $\hat{\rho}$ as an estimator which enables us to estimate the state of the quantum system knowing the outcomes of $(n-1)$ measurements.

Now, let $\mathcal{M}$ be an n-dimensional compact Riemannian submanifold of the Hilbert space $(B_+(\mathbb{H}), \langle . | . \rangle)$ consisting of semi-density matrices, where for $\rho$ and $\rho'$ in $B_+(\mathbb{H})$, $\langle \rho | \rho' \rangle = \text{Tr}(\rho \rho')$. Let the (generalized) quantum model $(\mathcal{M}, \Sigma, \mu)$ be its associated canonical measure space. To obtain $q_I^{(n-1)}, \hat{\rho}^{(n)}$, let $Z$ be the set of all extremum points of the smooth function $h : \rho \rightarrow$
\[ \text{Tr}(q_{(n-1)}^{(n-1)} \rho^{(n-1)}) \] on \( \mathcal{M} \), and let \( Z' \) be the set of all elements \( \rho \in Z \) at which the bundle map \( \text{Hessian}(h) : T\mathcal{M} \to T\mathcal{M} \) is negative. Clearly, \( Z' \) is the set of all maximum points of \( h \). Now, let \( \rho_0 \) be the element of \( Z' \) with least trace. Then, \( q_{(n-1)}^{(n-1)} \rho_0^{(n)} = \omega(\rho_0) \). If there are more than one \( \rho_0 \) in \( Z' \) we do not have any further preference among them.

### 7 Consistency of Our Method

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 \( \mathbb{H} \) at the state \( \rho_0 \) repeatedly and obtain more and more data the state yielded by the method is more and more close to the state \( \rho_0 \) in some sense.

In this section we investigate some different notions of 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 \( \rho \) 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 \( \sum_{q^{(n)} \in Q^{(n)}|P_n(q^{(n)})} \rho^{(n)} q^{(n)} \) will be denoted by \( \rho(P_n) \). Now let \( \rho' \) 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

\[ q^{(n)} \rho' q^{(n)}/q^{(n)} \rho q^{(n)} > \delta \]

on \( Q^{(n)} \).

**Definition 15**

*Under the above notations and conventions we say \( \rho' \) is asymptotically distinguishable from \( \rho \) if for all \( \delta > 0 \) we have

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

Now let \( \mathcal{M} \) be a generalized quantum model on \( \mathbb{H} \) and \( \bar{\rho}_n \) be defined as follows
\[ \tilde{\rho}_n(q^{(n)}) = \omega(\arg\max_{\rho \in M} q^{(n)} \rho^{(n)} q^{(n)}). \]

Assume that \( M' \) is a quantum model. Then the unary relation \( \tilde{\rho}_n(q^{(n)}) \in M' \) on \( Q^{(n)} \) will be denoted by \( \tilde{\rho}_n \in M' \).

Now we have the following important consistency theorem.

**Theorem 4**

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

\[ \lim_{n \to \infty} \rho^*(\tilde{\rho}_n) \in \mathcal{M} \]

**Proof.** Let \( D \subseteq Q^* \) be defined as follows:

\( q^{(n)} \in Q^* \), is in \( D \) if and only if there exists an element \( \rho \in \mathcal{M} \) different from \( \rho' \) such that

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

Let the mapping \( \eta : D \rightarrow \mathcal{M} \) be a function with the following property

For each \( q^{(n)} \in D \), if there exists \( \rho \in \mathcal{M} \) satisfying the above inequality and \( \omega(\rho) \in \mathcal{M} \) then \( \eta(q^{(n)}) = \rho \). Otherwise, \( \eta(q^{(n)}) = \rho \) is an element of \( \mathcal{M} \) 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

\[ \rho^*(\tilde{\rho}_n) \in \mathcal{M}_2 \]

\( \rho^*(\tilde{\rho}_n) \in \mathcal{M}_2 \). Now, in the countable quantum model \( \mathcal{M}_2 \), the set of all elements asymptotically distinguishable from \( \rho^* \) is \( \mathcal{M}_2 \). The rest of the proof is the same as the proof of Theorem 5.1 of \[10\].

\[ \blacksquare \]
7.2 Consistency in the sense of Cezaro, Hellinger and Re’nyi divergence

Theorem 5

Let $\bar{\rho}$ and $\rho^*$ be Quantum sources. Then $D(\rho^*(n)\|\bar{\rho}^{(n)})$ is

$$
= \sum_{i=1}^{n} E_{\rho^*(i-1)} D(\rho^*_i(i-1)\|\bar{\rho}_i(i-1)).
$$

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

$$
D(\rho^*(n)\|\bar{\rho}^{(n)}) = \text{Tr} \rho^*(n) \log \rho^*(n) - \rho^*(n) \log \bar{\rho}^{(n)}
$$

$$
= \text{Tr} \rho^*(n) (\log \rho^*(n) - \log \bar{\rho}^{(n)})
$$

$$
= \Pi_{i=1}^{n} (\rho^*_i(i-1))(\log \Pi_{i=1}^{n} \rho^*_i(i-1) - \log \Pi_{i=1}^{n} \bar{\rho}_i(i-1))
$$

$$
= \sum_{i=1}^{n} \text{Tr}(\rho^*(i-1)) \rho^*_i(i-1)(\log \rho^*_i(i-1) - \log \bar{\rho}_i(i-1))
$$

$$
= \sum_{i=1}^{n} E_{\rho^*(i-1)} D(\rho^*_i(i-1)\|\bar{\rho}_i(i-1)).
$$

(See also [10].)

Theorem 6 (Convergence Theorem for quantum Estimators) (Barron 1998)

Let $H$ be a separable Hilbert space and $Q \in \pi_0(H)$. Let $\mathcal{M}$ be a set of quantum sources on $H$, 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 $\bar{\rho}_Q^{(n)}$ is Cesaro consistent with respect to $\mathcal{M}$. In other words

$$
\lim_{n \to \infty} \frac{1}{n} D(\hat{\rho}_Q^{(n)}\|\rho_Q^{(n)}) = 0.
$$

The proof is a consequence of the definition of $Q$-universal element and lemma 14.

Before going further in this direction, we give the following definitions.
Definition 16
Let $\rho, \rho' \in D(\mathcal{H})$, and $\lambda > 0$ be a real number. Then

1) The Hellinger distance of $\rho$ and $\rho'$ is defined as follows

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

2) The Rényi divergence of order $\lambda$ of $\rho$ and $\rho'$ is defined as follows

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

Now let $\alpha > 1$ and $\mathcal{M}$ be a generalized quantum model. For each $\rho \in \mathcal{M}$ let

$$\rho' = [\text{Tr}^\alpha]^{-1}\rho \quad \text{and} \quad \mathcal{M}' = \{\rho' | \rho \in \mathcal{M}\}.$$ 

and let $\tilde{\rho}_n = \omega(\arg\max_{\rho \in \mathcal{M}} q^{(n)}(\rho) \cdot \rho^{(n)} q^{(n)})$ and if the maximum is achieved by more than one $\rho$ we take the one with the greatest trace. Let $\tilde{\rho}_\alpha^{(n)}$ be defined as $q^{(n)} \tilde{\rho}_\alpha^{(n)} q^{(n)} = \max_{\rho \in \mathcal{M}} q^{(n)}(\rho) \cdot \rho^{(n)} q^{(n)}$.

Assume that $\bar{\rho} = (\tilde{\rho}_n)_{n \in \mathbb{N}}$ is $Q$-universal for the quantum model associated with $\mathcal{M}$. Then we have the following theorem

Theorem 7
Under the above notations and conventions for all $0 < \lambda \leq 1 - 1/\alpha$ we have

$$E_{\tilde{\rho}_Q^{(n)}}(\bar{d}_\lambda(\rho_Q^{(n)} \parallel \tilde{\rho}_Q^{(n)})) \leq \frac{1}{n} D(\rho_Q^{(n)} \parallel \tilde{\rho}_Q^{(n)}).$$

And for $\alpha = 2$ we have

$$E_{\tilde{\rho}_Q^{(n)}}(He^2(\rho_Q^{(n)} \parallel \tilde{\rho}_Q^{(n)})) \leq \frac{1}{n} D(\rho_Q^{(n)} \parallel \tilde{\rho}_Q^{(n)}).$$

Proof. Let $\tilde{\mathcal{M}}_n$ be a Bayesian countable subset of the quantum model associated with $\mathcal{M}$ and $\tilde{\rho}_n$ be in it. The rest of the proof is the same as the proof of Theorem 15.3 of [10] with obvious modifications.

Remark 5
Let $\epsilon < 1$ be a nonzero positive real number and let $t : [\epsilon, \infty] \rightarrow [1, \infty]$ be defined as follows

27
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 15**

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 $g : [1, \infty[ \to [0, \infty]$ be a continuous function such that for all $x \geq 1$ we have $g(x) = f \circ t^{-1}(x)$. 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 $\sum_{i=1}^{n} a_i = O(F(n) + 1)$. Then there exists a constant $c \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ greater than some $n_0$ we have $\sum_{i=1}^{n} a_i \leq c(F(n) + 1)$. Then,

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

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
\[ F(n) = f(\epsilon) + \sum_{i=1}^{n-1} f(i) + O(1) = \sum_{i=1}^{n} g(i) + O(1). \]

Therefore,

\[
\sum_{i=1}^{n} a_i = O(\sum_{i=1}^{n} g(i)) = O(F(n) + O(1)) = O(F(n) + 1)
\]

2. From the equality \( \lim_{n \to \infty} a_n = 0 \) 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 \( \epsilon \) is arbitrary we have \( \lim_{n \to \infty} \sum_{i=1}^{n} a_i/n = 0 \).

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

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

Hence,

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

Therefore, \( \lim_{n \to \infty} a_n = 0 \). \( \blacksquare \)

**Remark 6**

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

**Definition 17**

Let \( \rho^* \) and \( \bar{\rho} \) be quantum sources. Then,

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

\[
\text{Risk}_n(\rho^*, \bar{\rho}) = \mathbb{E}_\rho [D(\rho^*_{n|n-1} \| \bar{\rho}_{n|n-1})]
\]

2. The Cesaro risk of \( \rho^* \) and \( \bar{\rho} \) is

\[
\overline{\text{Risk}}_n(\rho^*, \bar{\rho}) = \mathbb{E}_\rho \left\{ \frac{1}{n} D(\rho^*^{(n)} \| \bar{\rho}^{(n)}) \right\}
\]
The following equality is a consequence of Lemma 17 and the definition of KL-risk.

\[
\overline{\text{Risk}}_n(\rho^*, \bar{\rho}) = w \frac{1}{n} \sum_{j=1}^{n} \text{Risk}_j(\rho^*, \bar{\rho})
\]

See also [10].

**Theorem 8**

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

2. Let \( \epsilon \) 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, \( \text{Risk}_n(\rho^*, \bar{\rho}) = w O(g(n)) \) if and only if \( \overline{\text{Risk}}_n(\rho^*, \bar{\rho}) = w O((F(n) + 1)/n) \). 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].

**Example 6**

From what we have said in Section 6 the Bayesian predictive density matrix for exponential family converges at rate \( O(\log(n)/n) \) in terms of Cesaro risk.

**8 Classical or Quantum Probability**

Assume that \( A \) is a quantum system with the state space \( \mathbb{H} \) and \( \mathcal{M} \) is an appropriate quantum model and \( Q \in \pi_0(\mathbb{H}) \) is a quantum measurement. Let \( q_I^{(n)} \in Q^{(n)} \) be the outcome of \( n \) times measurement of the system at state \( \rho \). We want to know whether \( \rho \) is \( Q \)-classical or not. To solve the problem we proceed as follows:

Let \( \mathcal{M}' = \{ e^{-\hat{\gamma}(\rho)} \rho | \rho \in \mathcal{M} \} \).

Where, \( \hat{\gamma}(\rho) = \nu(\rho) + S(\rho) \), \( \nu(\rho) \) is the quantum complexity of \( \rho \) and \( S(\rho) \) is the von Neumann entropy of \( \rho_Q \)
Finally we try to obtain a maximum element of the set $M'$, with the help of the given data $q^{(n)}_1$.

If there exist more than one maximum elements we choose $\rho$ with minimum $\hat{\gamma}(\rho)$. If there exists only one maximum element $\rho$ with this property, then:

i) If $\nu(\rho) \neq 0$, the state of the system is $Q$-nonclassical.
ii) If $\nu(\rho) = 0$, the state of the system is $Q$-classical.

And if the cardinality of the set $M$ of elements with this property has more than one element, then

i) If $\nu(\rho) \neq 0$, for all $\rho \in M$ then the state of the system is $Q$-nonclassical.
ii) If $\nu(\rho) = 0$, for all $\rho \in M$ then the state of the system is $Q$-classical.
iii) Otherwise, the state of the system indifferently may be $Q$-classical or $Q$-nonclassical.

Let us illustrate our explanations about our method by the following example.

**Example 7**

*Let the quantum exponential family $(\rho_t)_{t \in \mathbb{R}}$ be defined as follows*

$$\rho_t = e^{\frac{1}{2}t[\sigma_z - \gamma(t)]}\rho_0 e^{\frac{1}{2}t[\sigma_z - \gamma(t)]},$$

*where*

$$\sigma = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \rho_0 = \frac{1}{2} \begin{bmatrix} 1 & x_0 \\ x_0 & 1 \end{bmatrix}, -1 \leq x_0 \leq 1, \gamma(t) = \log[\text{Tr}(\rho_0 e^{t\sigma_z})].$$

*now we perform measurement on the state space $\mathbb{H}$, the two dimensional Hilbert space, by the system of measurement $\Pi = \{q_0 = |0\rangle\langle 0|, q_1 = |1\rangle\langle 1|\}$ $n$ times and obtain $n_0$ times $q_0$ and $n_1$ times $q_1$. We want to estimate the state of the system. Let $\rho_T, T = \arg \max_t \{e^{-\hat{\gamma}(\rho_t)}(q_0 \rho_t q_0)^{n_0}(q_1 \rho_t q_1)^{n_1}\}$ be the estimator obtains by our method. Let $\hat{\gamma}(\rho) = \nu(\rho) + \gamma(\rho)$ be as above. Clearly*

$$T = \arg \max_t \{e^{-\hat{\gamma}(\rho_t)}(q_0 \rho_t q_0)^{n_0}(q_1 \rho_t q_1)^{n_1}\}$$

$$= \arg \max_t \{-\nu(\rho_t) + \gamma(\rho_t) + t(n_0 - n_1) - n \ln(e^t + e^{-t})\}$$

31
where,
\[
\rho_t = \frac{1}{e^t + e^{-t}} \begin{bmatrix} e^t & x_0 \\ x_0 & e^{-t} \end{bmatrix}
\]
\[
\gamma(\rho_t) = \frac{e^t}{e^t + e^{-t}} \ln\left(\frac{e^t + e^{-t}}{e^t + e^{-t}}\right) + \frac{e^{-t}}{e^t + e^{-t}} \ln\left(\frac{e^{-t}}{e^t + e^{-t}}\right) = t \tanh(t) - \ln(e^t + e^{-t})
\]
\[
\nu(\rho_t) = \frac{x_0}{e^t + e^{-t}}
\]

Therefore,
\[
T = \arg\max_t -\left(\frac{x_0}{e^t + e^{-t}} + t \tanh(t) - \ln(e^t + e^{-t})\right) + t(n_0 - n_1 - n \ln(e^t + e^{-t})).
\]
equivalently,
\[
d\left(\frac{x_0}{e^t + e^{-t}} + t \tanh(t) - \ln(e^t + e^{-t})\right) dt \bigg|_{t=T} = 0
\]
or
\[
(k - n)y^4 + x_0y^3 + 2ky^2 - x_0y - 4y^2 \ln y + k + n = 0
\]
where \(n_0 - n_1 = k\) and \(e^T = y\).

It is easy to see that the best estimation according to the MLE is \(y = \sqrt{\frac{n + k}{n - k}}\), which doesn’t depend on \(x_0\) and doesn’t get any information about it. As the following table shows the estimator obtained by our method is eventually the same as the ML estimator.

| n   | n_0 | n_1 | k   | x_0 | MLE results | Our methods |
|-----|-----|-----|-----|-----|-------------|-------------|
| 10  | 8   | 2   | 6   | 0.75| y=2        | y=1.92165   |
|     | 1   |     |     |     | y=2        | y=1.93858   |
|     | 0   |     |     |     | y=2        | y=1.87683   |
| 100 | 80  | 20  | 60  | 0.75| y=2        | y=1.99180   |
|     | 1   |     |     |     | y=2        | y=1.99366   |
|     | 0   |     |     |     | y=2        | y=1.98627   |
| 100 | 5   | 95  | -90 | 0.5 | y=0.2316   | y=0.2294    |
| 1000| 560 | 440 | 120 | 0.5 | y=1.1281   | y=1.1280    |
|     | 1   |     |     |     | y=1.1281   | y=1.1280    |
|     | 0   |     |     |     | y=1.1281   | y=1.1280    |
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. Grunwald, 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] M. Lobino, D. Korystov, C. Kupchak, E. Figueroa, B. C. Sanders and A. I. Lvovsky, Complete Characterization of Quantum-Optical Processes, Science journal, 2008

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

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

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

[25] J. Rissanen, Model Selection and Testing by the MDL Principle, in Information Theory and Statistical Learning, 2009.
[26] J. Rissanen, Stochastic Complexity and Modeling, Annals Statistics 14, 1986.

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