Infinite Dimensional Analysis, Quantum Probability
and Related Topics
Vol. 9, No. 3 (2006) 331–351

World Scientific
www.worldscientific.com

SUFFICIENCY IN QUANTUM STATISTICAL INFERENCE. A SURVEY WITH EXAMPLES

ANNA JENČOVÁ

Mathematical Institute, Slovak Academy of Sciences,
Stefanikova 49, Bratislava, Slovakia
jenca@mat.savba.sk

DÉNES PETZ

Alfréd Rényi Institute of Mathematics,
Hungarian Academy of Sciences,
P. O. Box 127, H-1364 Budapest, Hungary
petz@renyi.hu

Received 12 April 2006
Communicated by F. Fagnola

This paper attempts to give an overview about sufficiency in the setting of quantum statistics. The basic concepts are treated in parallel to the measure theoretic case. It turns out that several classical examples and results have a noncommutative analogue. Some of the results are presented without proof (but with exact references) and the presentation is intended to be self-contained. The main examples discussed in the paper are related to the Weyl algebra and to the exponential family of states. The characterization of sufficiency in terms of quantum Fisher information is a new result.

Keywords: Quantum statistics; coarse-graining; factorization theorem; exponential family; perturbation of states; sufficient subalgebra; quantum Fisher information.

AMS Subject Classification: 46L53, 81R15, 62B05

In order to motivate the concept of sufficiency, we first turn to the setting of classical statistics. Suppose we observe an N-dimensional random vector X, characterized by the density function $f(x|\theta)$, where θ is a p-dimensional vector of parameters and $p < N$. Assume that the densities $f(x|\theta)$ are known and the parameter θ completely determines the distribution of X. Therefore, θ is to be estimated. The N-dimensional observation X carries information about the p-dimensional parameter vector θ. One may ask the following question: Can we compress x into a low-dimensional statistic without any loss of information? Does there exist some function $t = Tx$, where the dimension of t is less than N, such that t carries all the useful information about θ? If so, for the purpose of studying θ, we could discard the measurements x and retain only the low-dimensional statistic t. In this case, we

call t a sufficient statistic. The following example is standard and simple. Suppose a binary information source emits a sequence of 0's and 1's, we have the independent variables X_1, $X_2, \ldots, X_N$ such that $\mathrm{Prob}(X_i = 1) = \theta$. In this case the empirical mean

$$T(x_1, x_2, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

can be used to estimate the parameter θ and it is a sufficient statistic.

1. Preliminaries

A quantum mechanical system is described by a C^*-algebra, the dynamical variables (or observables) correspond to the self-adjoint elements and the physical states of the system are modeled by the normalized positive functionals of the algebra, see Refs. 4 and 5. The evolution of the system $\mathcal{M}$ can be described in the *Heisenberg picture* in which an observable $A \in \mathcal{M}$ moves into $\alpha(A)$, where α is a linear transformation. α is an automorphism in case of the time evolution of a closed system but it could be the irreversible evolution of an open system. The *Schrödinger picture* is dual, it gives the transformation of the states, the state $\varphi \in \mathcal{M}^*$ moves into $\varphi \circ \alpha$. The algebra of a quantum system is typically noncommutative but the mathematical formalism supports commutative algebras as well. A simple *measurement* is usually modeled by a family of pairwise orthogonal projections, or more generally, by a partition of unity, $(E_i)_{i=1}^n$. Since all E_i are supposed to be positive and $\sum_i E_i = I$, $\beta : \mathbb{C}^n \to \mathcal{M}$, $(z_1, z_2, \ldots, z_n) \mapsto \sum_i z_i E_i$ gives a positive unital mapping from the commutative C^*-algebra $\mathbb{C}^n$ to the noncommutative algebra $\mathcal{M}$. Every positive unital mappings occur in this way. The essential concept in quantum information theory is the state transformation which is affine and the dual of a positive unital mapping. All these and several other situations justify the study of positive unital mappings between C^*-algebras from a quantum statistical viewpoint.

If the algebra $\mathcal{M}$ is "small" and $\mathcal{N}$ is "large", and the mapping $\alpha : \mathcal{M} \to \mathcal{N}$ sends the state φ of the system of interest to the state $\varphi \circ \alpha$ at our disposal, then loss of information takes place and the problem of statistical inference is to reconstruct the real state from partial information. In this paper we mostly consider parametric statistical models, a parametric family $\mathcal{S} := \{\varphi_\theta : \theta \in \Theta\}$ of states is given and on the basis of the partial information the correct value of the parameter should be decided. If the partial information is the outcome of a measurement, then we have statistical inference in the very strong sense. However, there are "more quantum" situations, to decide between quantum states on the basis of quantum data. The problem we discuss is not the procedure of the decision about the true state of the system but we want to describe the circumstances under which this is perfectly possible.

In this paper, C^*-algebras always have a unit I. Given a C^*-algebra $\mathcal{M}$, a state φ of $\mathcal{M}$ is a linear function $\mathcal{M} \to \mathbb{C}$ such that $\varphi(I) = 1 = \|\varphi\|$. (Note that the

second condition is equivalent to the positivity of φ.) The books[4,5] — among many others — explain the basic facts about C^*-algebras. The class of finite dimensional full matrix algebras forms a small and algebraically rather trivial subclass of C^*-algebras, but from the viewpoint of noncommutative statistics, almost all ideas and concepts appear in this setting. A matrix algebra $M_n(\mathbb{C})$ admits a canonical trace Tr and all states are described by their densities with respect to Tr. The correspondence is given by $\varphi(A) = \mathrm{Tr}\rho_\varphi A$ ($A \in M_n(\mathbb{C})$) and we can simply identify the functional φ by the density ρ_φ. Note that the density is a positive (semi-definite) matrix of trace 1.

Example 1. Let $\mathcal{X}$ be a finite set and $\mathcal{N}$ be a C^*-algebra. Assume that for each $x \in \mathcal{X}$ a positive operator $E(x) \in \mathcal{N}$ is given and $\sum_x E(x) = I$. In quantum mechanics such a setting is a model for a measurement with values in $\mathcal{X}$.

The space $C(\mathcal{X})$ of function on $\mathcal{X}$ is a C^*-algebra and the partition of unity E induces a coarse-graining $\alpha : C(\mathcal{X}) \to \mathcal{N}$ given by $\alpha(f) = \sum_x f(x)E(x)$. Therefore a coarse-graining defined on a commutative algebra is an equivalent way to give a measurement. (Note that the condition of two-positivity is automatically fulfilled on a commutative algebra.) □

Example 2. Let $\mathcal{M}$ be the algebra of all bounded operators acting on a Hilbert space $\mathcal{H}$ and let $\mathcal{N}$ be the infinite tensor product $\mathcal{M} \otimes \mathcal{M} \otimes \dots$. (To understand the essence of the example one does not need the very formal definition of the infinite tensor product.) If γ denotes the right shift on $\mathcal{N}$, then we can define a sequence α_n of coarse-grainings $\mathcal{M} \to \mathcal{N}$:

$$\alpha_n(A) := \frac{1}{n}(A + \gamma(A) + \dots + \gamma^{n-1}(A)) .$$

α_n is the quantum analogue of the *sample mean.* □

In this paper, the emphasis is on the definitions and the results. The results obtained in earlier works are typically not proved but several examples are presented to give a better insight. Fisher information is simply a widely used concept in classical statistics. The relation of sufficiency and quantum Fisher information is new and proved here in details. (However, the concept of quantum Fisher information is rather concisely discussed.)

2. Basic Definitions

In this section we recall some well-known results from classical mathematical statistics, Ref. 20 is our general reference, and the basic concepts of the quantum cases are discussed in parallel.

Let $(X_i, \mathcal{A}_i, \mu_i)$ be probability spaces ($i = 1, 2$). Recall that a positive linear map $M : L^\infty(X_1, \mathcal{A}_1, \mu_1) \to L^\infty(X_2, \mathcal{A}_2, \mu_2)$ is called a *Markov operator* if it satisfies $M1 = 1$ and $f_n \searrow 0$ implies $Mf_n \searrow 0$.

Let $\mathcal{M}$ and $\mathcal{N}$ be C^*-algebras. Recall that *two-positivity* of $\alpha : \mathcal{M} \to \mathcal{N}$ means that

$$\begin{pmatrix} \alpha(A) & \alpha(B) \\ \alpha(C) & \alpha(D) \end{pmatrix} \geq 0 \quad \text{if} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix} \geq 0$$

for 2×2 matrices with operator entries. It is well known that a 2-positive unit-preserving mapping α satisfies the *Schwarz inequality*

$$\alpha(A^*A) \geq \alpha(A)^*\alpha(A)\,. \tag{1}$$

A 2-positive unital mapping between C^*-algebras will be called *coarse-graining*. All Markov operators (defined above) are coarse-grainings. For mappings defined between von Neumann algebras, the monotone continuity is called *normality*. When $\mathcal{M}$ and $\mathcal{N}$ are von Neumann algebras, a coarse-graining $\mathcal{M} \to \mathcal{N}$ will always be assumed to be normal. Therefore, our concept of coarse-graining is the analogue of the Markov operator.

We mostly mean that a coarse-graining transforms observables to observables corresponding to the *Heisenberg picture* and in this case we assume that it is unit preserving. The dual of such a mapping acts on states or on density matrices and it will be called state transformation.

Let $(X, \mathcal{A})$ be a measurable space and let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a set of probability measures on $(X, \mathcal{A})$. Usually, $\mathcal{P}$ is called *statistical experiment*, if it contains only two measures, then we speak about a *binary experiment*. The aim of estimation theory is to decided about the true value of θ on the basis of data.

A sub-σ-algebra $\mathcal{A}_0 \subset \mathcal{A}$ is *sufficient* for the family $\mathcal{P}$ of measures if for all $A \in \mathcal{A}$, there is an $\mathcal{A}_0$-measurable function f_A such that for all θ,

$$f_A = P_\theta(A|\mathcal{A}_0) \quad P_\theta\text{-almost everywhere}\,,$$

that is,

$$P_\theta(A \cap A_0) = \int_{A_0} f_A dP_\theta \tag{2}$$

for all $A_0 \in \mathcal{A}_0$ and for all θ. It is clear from this definition that if $\mathcal{A}_0$ is sufficient, then for all P_θ there is a common version of the conditional expectations $E_\theta[g|\mathcal{A}_0]$ for any measurable step function g, or, more generally, for any function $g \in \cap_{\theta\in\Theta} L^1(X, \mathcal{A}, P_\theta)$.

In the most important case, the family $\mathcal{P}$ is *dominated*, that is there is a σ-finite measure μ such that P_θ is absolutely continuous with respect to μ for all θ, this will be denoted by $\mathcal{P} \ll \mu$. A finite family is always dominated.

For our purposes, it is more suitable to use the following characterization of sufficiency in terms of randomization.

Let $\mathcal{P}_i = \{P_{i,\theta} : \theta \in \Theta\}$ be dominated families of probability measures on $(X_i, \mathcal{A}_i)$, such that $\mathcal{P}_i \equiv \mu_i$, $i = 1, 2$. We say that $(X_2, \mathcal{A}_2, \mathcal{P}_2)$ is a *randomization*

of $(X_1, \mathcal{A}_1, \mathcal{P}_1)$, if there exists a Markov operator $M : L^\infty(X_2, \mathcal{A}_2, \mu_2) \to L^\infty(X_1, \mathcal{A}_1, \mu_1)$, satisfying

$$\int (Mf) dP_{\theta,1} = \int f dP_{\theta,2}\,, \qquad \theta \in \Theta, f \in L^\infty(X_2, \mathcal{A}_2, \mathcal{P}_2)\,.$$

If $(X_1, \mathcal{A}_1, \mathcal{P}_1)$ is also a randomization of $(X_2, \mathcal{A}_2, \mathcal{P}_2)$, then $(X_1, \mathcal{A}_1, \mathcal{P}_1)$ and $(X_2, \mathcal{A}_2, \mathcal{P}_2)$ are *statistically equivalent*.

For example, let $\mathcal{P} \equiv P_0$ and let $\mathcal{A}_0 \subseteq \mathcal{A}$ be a subalgebra. Then $(X, \mathcal{A}_0, \mathcal{P}|_{\mathcal{A}_0})$ is obviously a randomization of $(X, \mathcal{A}, \mathcal{P})$, where the Markov operator is the inclusion $L^\infty(X, \mathcal{A}_0, P_0|_{\mathcal{A}_0}) \to L^\infty(X, \mathcal{A}, P_0)$. On the other hand, if $\mathcal{A}_0$ is sufficient, then the map

$$f \mapsto E[f|\mathcal{A}_0]\,, \qquad E[f|\mathcal{A}_0] = E_\theta[f|\mathcal{A}_0]\,, \qquad P_\theta\text{-almost everywhere}\,,$$

is a Markov operator $L^\infty(X, \mathcal{A}, P_0) \to L^\infty(X, \mathcal{A}_0, P_0|_{\mathcal{A}_0})$ and

$$\int E[f|\mathcal{A}_0] dP_\theta|_{\mathcal{A}_0} = \int f dP_\theta\,, \qquad f \in L^\infty(X, \mathcal{A}, P_0)\,,\ \theta \in \Theta\,.$$

We have the following characterizations of sufficient sub-σ-algebras.

Proposition 1. *Let $\mathcal{P}$ be a dominated family and let $\mathcal{A}_0 \subseteq \mathcal{A}$ be a sub-σ-algebra. The following are equivalent*:

(i) *$\mathcal{A}_0$ is sufficient for $\mathcal{P}$,*
(ii) *There exists a measure P_0 such that $\mathcal{P} \equiv P_0$ and dP_θ/dP_0 is $\mathcal{A}_0$-measurable for all θ.*
(iii) *$(X, \mathcal{A}, \mathcal{P})$ and $(X, \mathcal{A}_0, \mathcal{P}|_{\mathcal{A}_0})$ are statistically equivalent.*

A classical *sufficient statistic* for the family $\mathcal{P}$ is a measurable mapping $T : (X, \mathcal{A}) \to (X_1, \mathcal{A}_1)$ such that the sub-σ-algebra $\mathcal{A}^T$ generated by T is sufficient for $\mathcal{P}$. To any statistic T, we associate a Markov operator

$$\tilde{T} : L^\infty(X_1, \mathcal{A}_1, P_0^T) \to L^\infty(X, \mathcal{A}, P_0)\,, \qquad (\tilde{T}g)(x) = g(T(x))\,.$$

Obviously, $(X_1, \mathcal{A}_1, \mathcal{P}^T)$ is a randomization of $(X, \mathcal{A}, \mathcal{P})$. As in the case of subalgebras, we have

Proposition 2. *The statistic $T : (X, \mathcal{A}) \to (X_1, \mathcal{A}_1)$ is sufficient for $\mathcal{P}$ if and only if $(X, \mathcal{A}, \mathcal{P})$ and $(X_1, \mathcal{A}_1, \mathcal{P}^T)$ are statistically equivalent.*

Example 3. Let P and Q be measures on the σ-algebra $\mathcal{A}$, i.e. $\{P, Q\}$ is a binary experiment which is dominated by $\mu := P + Q$. Let us define the function

$$T : X \ni x \mapsto \frac{dP}{d\mu}(x) \in [0, 1]$$

T is a minimal sufficient statistic for $\{P, Q\}$. For illustration, we prove this statement directly.

Let $\mathcal{A}_0 \subseteq \mathcal{A}$ be a sub-σ-algebra. For $A \in \mathcal{A}$, let us denote $f_A := P_0(A|\mathcal{A}_0)$. We show that f_A is a common version of $P(A|\mathcal{A}_0)$ and $Q(A|\mathcal{A}_0)$ if and only if T is $\mathcal{A}_0$-measurable. Indeed, for $A_0 \in \mathcal{A}_0$,

$$P(A \cap A_0) = \int_{A_0} \mathbf{1}_A \, dP = \int_{A_0} \mathbf{1}_A T \, d\mu = \int_{A_0} E_\mu[\mathbf{1}_A T|\mathcal{A}_0] d\mu$$

and similarly,

$$Q(A \cap A_0) = \int_{A_0} E_\mu[\mathbf{1}_A (1-T)|\mathcal{A}_0] d\mu \,.$$

The fact that T is $\mathcal{A}_0$-measurable is equivalent with

$$\int_{A_0} E_\mu[\mathbf{1}_A T|\mathcal{A}_0] d\mu = \int_{A_0} f_A T \, d\mu = \int_{A_0} f_A dP$$

for all $A_0 \in \mathcal{A}_0$, and similarly for Q.

Let $p := \frac{dP}{d\mu}$, $q := \frac{dQ}{d\mu}$. Then

$$\frac{dQ}{dP} := \frac{q}{p} \mathbf{1}_{\{p>0\}}$$

is called the *likelihood ratio* of Q and P.

Since

$$\frac{dQ}{dP} = \frac{1-T}{T} \mathbf{1}_{\{T>0\}} \,,$$

the likelihood ratio and T generates the same σ-algebra. It follows that the likelihood ratio is a minimal sufficient statistic as well. □

Proposition 3. (Factorization criterion) *Let $\mathcal{P} \ll \mu$. The statistic $T : (X, \mathcal{A}) \to (X_1, \mathcal{A}_1)$ is sufficient for $\mathcal{P}$ if and only if there is an $\mathcal{A}_1$-measurable function g_θ for all θ and an $\mathcal{A}$-measurable function h such that*

$$\frac{dP_\theta}{d\mu}(x) = g_\theta(T(x))h(x) \qquad P_\theta\text{-almost everywhere}\,.$$

Example 4. Let $X_1, X_2, \ldots, X_N$ be independent random variables with normal distribution $N(m, \sigma)$. It is well known that the empirical mean

$$T(x_1, x_2, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

is a sufficient statistic for the parameter m, when σ is fixed.

The joint distribution is

$$\prod_{i=1}^{N} C \exp\left(-\frac{(x_i - m)^2}{2\sigma^2}\right)$$
$$= C^N \exp\left(-\frac{m}{\sigma^2} \sum_{i=1}^{N} x_i - \frac{nm^2}{2\sigma^2}\right) \exp\left(-\frac{\sum_{i=1}^{N} x_i^2}{2\sigma^2}\right)$$

and we observe the factorization:

$$f(x,m) = g(T(x),m)h(x)\,.$$

According to Proposition 3, this is enough for the sufficiency. □

Next we formulate the noncommutative setting. Let $\mathcal{M}$ be a von Neumann algebra and $\mathcal{M}_0$ be its von Neumann subalgebra. Assume that a family $\mathcal{S} := \{\varphi_\theta : \theta \in \Theta\}$ of normal states are given. $(\mathcal{M}, \mathcal{S})$ is called *statistical experiment*. The subalgebra $\mathcal{M}_0 \subset \mathcal{M}$ is *sufficient* for $(\mathcal{M}, \mathcal{S})$ if for every $a \in \mathcal{M}$, there is $\alpha(a) \in \mathcal{M}_0$ such that

$$\varphi_\theta(a) = \varphi_\theta(\alpha(a))\,, \qquad \theta \in \Theta \tag{3}$$

and the correspondence $a \mapsto \alpha(a)$ is a coarse-graining. (Note that a positive mapping is automatically completely positive if it is defined on a commutative algebra.)

We will now define sufficient coarse-grainings. Let $\mathcal{N}$, $\mathcal{M}$ be C^*-algebras and let $\sigma : \mathcal{N} \to \mathcal{M}$ be a coarse-graining. By Proposition 2, the classical definition of sufficiency can be generalized in the following way: we say that σ is sufficient for the statistical experiment $(\mathcal{M}, \varphi_\theta)$ if there exists a coarse-graining $\beta : \mathcal{M} \to \mathcal{N}$ such that $\varphi_\theta \circ \sigma \circ \beta = \varphi_\theta$ for every θ.

The next example is the analogue of Example 4 on the algebra of the canonical commutation relation. Note that the bilinear form α plays the role of the variance (while σ denotes a simplectic form).

Example 5. Let σ be a nondegenerate symplectic form on a linear space $\mathcal{H}$. Typically, $\mathcal{H}$ is a complex Hilbert space and $\sigma(f,g) = \mathrm{Im}\langle f, g\rangle$. The *Weyl algebra* $\mathrm{CCR}(\mathcal{H})$ is generated by unitaries $\{W(f) : f \in \mathcal{H}\}$ satisfying the Weyl form of the *canonical commutation relation*:

$$W(f)W(g) = e^{i\sigma(f,g)}W(f+g)\,, \qquad f,g \in \mathcal{H}\,,$$

see Refs. 5 and 13 about the details. Since the linear hull of the unitaries $W(f)$ is dense in $\mathrm{CCR}(\mathcal{H})$, any state is determined uniquely by its values taken on the Weyl unitaries. The most important states of the Weyl algebra are the Gaussian (or quasifree) states which are given as

$$\varphi_{m,\alpha}(W(f)) = \exp\left(m(f)i - \frac{1}{2}\alpha(f,f)\right)\,, \qquad f \in \mathcal{H}\,,$$

where m is a linear functional and α is a bilinear functional on $\mathcal{H}$ and $\mathcal{H} \times \mathcal{H}$, respectively. Note that α should satisfy the constraint

$$\sigma(f,g)^2 \le \alpha(f,f)\alpha(g,g)\,, \tag{4}$$

see Theorem 3.4 and its proof in Ref. 13.

It is well known that

$$\mathrm{CCR}(\mathcal{K}_1) \otimes \mathrm{CCR}(\mathcal{K}_2) \otimes \cdots \otimes \mathrm{CCR}(\mathcal{K}_n)$$

may be regarded as

$$\mathrm{CCR}(\mathcal{K}_1 \oplus \mathcal{K}_2 \oplus \cdots \oplus \mathcal{K}_n)$$

for any Hilbert spaces $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_n$. Now we assume that all these spaces coincide with $\mathcal{H}$ and we write $\mathcal{H}_n$ for $\mathcal{H} \oplus \mathcal{H} \oplus \cdots \oplus \mathcal{H}$. The bilinear forms α_n and σ_n defined on $\mathcal{H}_n$ are induced by α and σ.

There exists a completely positive (so-called quasifree) mapping

$$T : \mathrm{CCR}(\mathcal{H}) \to \mathrm{CCR}(\mathcal{H}_n)$$

such that

$$T(W(f)) = W\left(\frac{1}{\sqrt{n}}(f \oplus f \oplus \cdots \oplus f)\right)$$

(p. 73 in Ref. 13). We claim that T is sufficient for the family

$$\{\psi_{m,\alpha} := \varphi^{(1)}_{m,\alpha} \otimes \varphi^{(2)}_{m,\alpha} \otimes \cdots \otimes \varphi^{(n)}_{m,\alpha} : m\}$$

of states on $\mathrm{CCR}(\mathcal{H}_n)$, when α is fixed.

Consider the quasi-free mapping $S_\alpha : \mathcal{A}^{(n)} \to \mathrm{CCR}(\mathcal{H})$ given as

$$S_\alpha(W(f_1 \oplus f_2 \oplus \cdots \oplus f_n)) = W\left(\frac{1}{\sqrt{n}}\sum_i f_i\right) \exp\left(\frac{1}{2n}\alpha\left(\sum_i f_i, \sum_i f_i\right) - \frac{1}{2}\sum_i \alpha(f_i, f_i)\right).$$

Then

$$(T \circ S_\alpha)(W(f_1 \oplus f_2 \oplus \cdots \oplus f_n)) = W\left(\frac{1}{n}\sum_i f_i \oplus \cdots \oplus \frac{1}{n}\sum_i f_i\right) \times \exp\left(\frac{1}{2n}\alpha\left(\sum_i f_i, \sum_i f_i\right) - \frac{1}{2}\sum_i \alpha(f_i, f_i)\right)$$

and

$$\psi_{m,\sigma}(T \circ S_\alpha) = \psi_{m,\sigma}$$

holds for every m. We will show that S_α is completely positive.

We can write $S_\alpha : \mathcal{A}^{(n)} \to \mathrm{CCR}(\mathcal{H})$ as

$$S_\alpha(W(f^n)) = W(A_n f^n) F(f^n), \tag{5}$$

where $f^n = f_1 \oplus \cdots \oplus f_n \in \mathcal{H}_n$,

$$F(f^n) = \exp\left(\frac{1}{2n}\alpha\left(\sum_i f_i, \sum_i f_i\right) - \frac{1}{2}\sum_i \alpha(f_i, f_i)\right)$$

and $A_n : \mathcal{H}_n \to \mathcal{H}$ is the linear map $f_1 \oplus \cdots \oplus f_n \mapsto \sum_i f_i$. By Theorem 8.1 in Ref. 13, S_α is completely positive if and only if the kernel

$$(f^n, g^n) \mapsto F(g^n - f^n) \exp i(\sigma_n(g^n, f^n) - \sigma(A_n g^n, A_n f^n)) \tag{6}$$

is positive definite.

It is easy to see that $A_n^* : f \mapsto \frac{1}{\sqrt{n}}(f \oplus f \oplus \cdots \oplus f)$ and

$$\alpha_n(f^n, (I - A_n^* A_n) g^n) = \alpha_n((I - A_n^* A_n) f^n, g^n) = \sum_i \alpha(f_i, g_i) - \frac{1}{n}\alpha\left(\sum_i f_i, \sum_i g_i\right).$$

Since A_n is a contraction, $I - A_n^* A_n$ is positive. Setting $B_n = (I - A_n^* A_n)^{1/2}$, we have

$$F(f^n) = \exp\left(-\frac{1}{2}\alpha_n(B_n f^n, B_n f^n)\right)$$

and

$$\sigma_n(g^n, f^n) - \sigma(A_n g^n, A_n f^n) = -\sigma_n(B_n f^n, B_n g^n).$$

The kernel (6) has the form

$$\exp\left(-\frac{1}{2}\alpha_n(B_n(g^n - f^n), B_n(g^n - f^n) + i\sigma_n(B_n f^n, B_n g^n))\right).$$

The positive definiteness follows from that of the exponent which is so due to

$$\sigma_n^2(f^n, g^n) \le \alpha_n(f^n, f^n)\alpha_n(g^n, g^n). \qquad \square$$

3. Sufficient Subalgebras and Coarse-Grainings

In the study of sufficient subalgebras monotone quasi-entropy quantities play an important role. The *relative α-entropies* are examples of Refs. 9 and 11.

Let φ and ω be normal states of a von Neumann algebra and let ξ_φ and ξ_ω be the representing vectors of these states from the natural positive cone (see below). Let

$$f_\alpha(t) = \frac{1}{\alpha(1-\alpha)}(1 - t^\alpha).$$

It is well known that this function is operator monotone decreasing for $\alpha \in (-1, 1)$. The relative α-entropy

$$S_\alpha(\varphi \| \omega) = \langle \xi_\varphi, f_\alpha(\Delta)\xi_\varphi \rangle \tag{7}$$

is a particular quasi-entropy corresponding to the function f_α, Δ is the relative modular operator $\Delta(\omega/\varphi)$. When ρ_1 and ρ_2 are statistical operators, this formula can be written as

$$S_\alpha(\rho_1 \| \rho_2) = \frac{1}{\alpha(1-\alpha)}\mathrm{Tr}(I - \rho_2^\alpha \rho_1^{-\alpha})\rho_1 \tag{8}$$

(for details, see Chap. 7 in Ref. 9).

The relative α-entropy is monotone under coarse-graining:

$$S_\alpha(\rho_1\|\rho_2) \geq S_\alpha(\mathcal{E}(\rho_1)\|\mathcal{E}(\rho_2))\,.$$

If follows also from the general properties of quasi-entropies that $S_\alpha(\rho_1\|\rho_2)$ is jointly convex and positive. The *transition probability*

$$P_A(\varphi,\omega) = \langle \xi_\varphi, \xi_\omega\rangle$$

corresponds to $\alpha = 1/2$ (up to additive and multiplicative constants).

The next theorem is essentially Theorem 9.5 from Ref. 9.

Theorem 1. *Let $\mathcal{M}_0 \subset \mathcal{M}$ be von Neumann algebras and let $(\mathcal{M}, \{\varphi_\theta : \theta \in \Theta\})$ be a statistical experiment. Assume that there are states $\varphi_n \in \mathcal{S} := \{\varphi_\theta : \theta \in \Theta\}$ such that*

$$\omega := \sum_{n=1}^{\infty} \lambda_n \varphi_n$$

is a faithful normal state for some constants $\lambda_n > 0$. Then the following conditions are equivalent.

(i) *$\mathcal{M}_0$ is sufficient for $(\mathcal{M}, \varphi_\theta)$.*
(ii) *$S_\alpha(\varphi_\theta, \omega) = S_\alpha(\varphi_\theta|\mathcal{M}_0, \omega|\mathcal{M}_0)$ for all θ and for some $0 < |\alpha| < 1$.*
(iii) *$[D\varphi_\theta, D\omega]_t = [D(\varphi_\theta|\mathcal{M}_0), D(\omega|\mathcal{M}_0)]_t$ for every real t and for every θ.*
(iv) *$[D\varphi_\theta, D\omega]_t \in \mathcal{M}_0$ for all real t and every θ.*
(v) *The generalized conditional expectation $E_\omega : \mathcal{M} \to \mathcal{M}_0$ leaves all the states φ_θ invariant.*

Since ω is assumed to be faithful and normal, it is convenient to consider a representation of $\mathcal{M}$ on a Hilbert space $\mathcal{H}$ such that ω is induced by a cyclic and separating vector Ω. Given a normal state ψ, the quadratic form $a\Omega \mapsto \psi(aa^*)$ $(a \in \mathcal{M})$ determines the relative modular operator $\Delta(\psi/\omega)$ as

$$\psi(aa^*) = \|\Delta(\psi/\omega)a\Omega\|^2\,, \qquad a \in \mathcal{M}\,.$$

The vector $\Delta(\psi/\omega)^{1/2}\Omega$ is the representative of ψ from the so-called natural positive cone (which is actually the set of all such vectors). The *Connes' cocycle*

$$[D\psi, D\omega]_t = \Delta(\psi/\omega)^{it}\Delta(\omega/\omega)^{-it}$$

is a one-parameter family of contractions in $\mathcal{M}$, unitaries when ψ is faithful. The modular group of ω is a group of automorphisms defined as

$$\sigma_t(a) = \Delta(\omega/\omega)^{it}a\Delta(\omega/\omega)^{-it}\,, \qquad t \in \mathbb{R}\,.$$

The Connes' cocycle is the quantum analogue of the Radon–Nikodym derivative of measures.

The generalized conditional expectation $E_\omega : \mathcal{M} \to \mathcal{M}_0$ is defined as

$$E_\omega(a)\Omega = J_0 P J a\Omega\,,$$

where J is the modular conjugation on the Hilbert space $\mathcal{H}$, J_0 is that on the closure $\mathcal{H}_0$ of $\mathcal{M}_0\Omega$ and $P : \mathcal{H} \to \mathcal{H}_0$ is the orthogonal projection.[1] There are several equivalent conditions which guarantee that E_ω is a conditional expectation, for example, $\sigma_t(\mathcal{M}_0) \subset \mathcal{M}_0$ (*Takesaki's theorem*, Ref. 9).

More generally, let $\mathcal{M}_1$ and $\mathcal{M}_2$ be von Neumann algebras and let $\sigma : \mathcal{M}_1 \to \mathcal{M}_2$ be a coarse-graining. Suppose that a normal state φ_2 is given and $\varphi_1 := \varphi_2 \circ \sigma$ is normal as well. Let Φ_i be the representing vectors in given natural positive cones and J_i be the modular conjugations ($i = 1, 2$).

From the modular theory we know that

$$p_i := \overline{J_i \mathcal{M}_i \Phi_i}$$

is the support projection of φ_i ($i = 1, 2$).

The dual $\sigma^*_{\varphi_2} : p_2\mathcal{M}_2p_2 \to p_1\mathcal{M}_1p_1$ of σ is characterized by the property

$$\langle a_1\Phi_1, J_1\sigma_{\varphi_2}(a_2)\Phi_1\rangle = \langle \sigma(a_1)\Phi_2, J_2a_2\Phi_2\rangle\,, \qquad a_i \in \mathcal{M}_i, i = 1, 2 \tag{9}$$

(see Proposition 8.3 in Ref. 9).

Example 6. Let $\mathcal{M}$ be a matrix algebra with a family of states $\{\varphi_\theta : \theta \in \Theta\}$ and let $\mathcal{M}^{n\otimes} := \mathcal{M} \otimes \cdots \otimes \mathcal{M}$ and $\varphi_\theta^{n\otimes} := \varphi_\theta \otimes \cdots \otimes \varphi_\theta$ be n-fold products. Each permutation of the tensor factors induces an automorphism of $\mathcal{M}^{n\otimes n}$ and let $\mathcal{N}$ be the fixed point subalgebra of these automorphisms. Then $\mathcal{N}$ is sufficient for the family $\{\varphi_\theta^{n\otimes} : \theta \in \Theta\}$. Indeed, the Cones' cocycle of any two of these states is a homogeneous tensor product, therefore they are in the fixed point algebra $\mathcal{N}$. □

Let us return to the Weyl algebra.

Example 7. Let $\mathcal{H}$ be a real Hilbert space with inner product $\alpha(f, g)$ ($f, g \in \mathcal{H}$) and let σ be a nondegenerate symplectic form on $\mathcal{H}$. Assume that (4) holds. Then there exists an invertible contraction D on $\mathcal{H}$, such that

$$\sigma(f, g) = \alpha(Df, g)\,, \qquad f, g \in \mathcal{H}\,.$$

Let $D = J|D|$ be the polar decomposition, then $JD = DJ$, $J^2 = -I$. The unitary J defines a complex structure on $\mathcal{H}$. We introduce a complex inner product by

$$\langle f, g\rangle := \sigma(f, Jg) + i\sigma(f, g)\,,$$

then

$$\sigma(f, g) = \mathrm{Im}\langle f, g\rangle\,, \quad \text{and} \quad \alpha(f, g) = \mathrm{Re}\langle |D|^{-1}f, g\rangle\,.$$

For each linear form m on $\mathcal{H}$, there is an element $g_m \in \mathcal{H}$, such that

$$m(f) = 2\sigma(g_m, f)\,, \qquad f \in \mathcal{H}\,.$$

Let φ_m be the quasi-free state on $\mathrm{CCR}(\mathcal{H}, \sigma)$ given by

$$\varphi_m(W(f)) = \exp\left(im(f) - \frac{1}{2}\alpha(f, f)\right)\,.$$

Then

$$\varphi_m(W(f)) = \varphi_0(W(g_m)W(f)W(-g_m))\,, \qquad f \in \mathcal{H}\,.$$

Let H be a subset of $\mathcal{H}$. The family of states $\mathcal{S}_H = \{\varphi_m : g_m \in H\}$, is the quantum counterpart of the classical Gaussian shift on $\mathcal{H}$.

Let us now suppose that $\|D\| < 1$, then there is an operator $L \geq \varepsilon I$ for some $\varepsilon > 0$, such that $|D|^{-1} = \coth L$. It was proved in Ref. 13 that the state φ_0 satisfies the KMS condition with respect to the automorphism group

$$\sigma_t(W(f)) = W(V_t f)\,, \qquad t \in \mathbb{R}, f \in \mathcal{H}\,,$$

where $V_t = \exp(-2itL)$. Therefore, σ_t is the modular group of φ_0. It is not difficult to prove that

$$u_t^g = \exp(i\sigma(V_t g, g))W(V_t g - g)\,, \qquad g \in \mathcal{H}, t \in \mathbb{R}$$

is the Connes' cocycle $[D\varphi_m, D\varphi_0]_t$. It follows that the algebra $\mathrm{CCR}(\mathcal{K}, \sigma|\mathcal{K})$ is minimal sufficient for $\mathcal{S}_\mathcal{K}$ when $\mathcal{K}$ is the subspace generated by $\{V_t(g) - g : g \in K\}$. In particular, we see that if $K = \mathcal{H}$, then there is no nontrivial sufficient subalgebra for the Gaussian shift.

Let us now recall the situation in Example 5. There, we studied the algebra $\mathrm{CCR}(\mathcal{H}_n, \sigma_n)$ with the family of states $\mathcal{S}_\mathcal{L}$, where $\mathcal{L} = \{g \oplus \cdots \oplus g : g \in \mathcal{H}\}$. It follows from our analysis that the minimal sufficient subalgebra is $\mathrm{CCR}(\mathcal{L}, \sigma_n)$. □

Let us recall the following well known property of coarse-grainings, see Sec. 9.2 in Ref. 19.

Lemma 1. *Let $\mathcal{M}$ and $\mathcal{N}$ be C^*-algebras and let $\sigma : \mathcal{N} \to \mathcal{M}$ be a coarse-graining. Then*

$$\mathcal{N}_\sigma := \{a \in \mathcal{N} : \sigma(a^*a) = \sigma(a)\sigma(a)^* \text{ and } \sigma(aa^*) = \sigma(a)^*\sigma(a)\} \tag{10}$$

is a subalgebra of $\mathcal{N}$ and

$$\sigma(ab) = \sigma(a)\sigma(b) \quad \text{and} \quad \sigma(ba) = \sigma(b)\sigma(a) \tag{11}$$

hold for all $a \in \mathcal{N}_\sigma$ and $b \in \mathcal{N}$.

We call the subalgebra $\mathcal{N}_\sigma$ the *multiplicative domain* of σ.

Now let $\mathcal{N}$ and $\mathcal{M}$ be von Neumann algebras and let ω be a faithful normal state on $\mathcal{M}$ such that $\omega \circ \sigma$ is also faithful. Let

$$\mathcal{N}_1 = \{a \in \mathcal{N}, \sigma_\omega^* \circ \sigma(a) = a\}\,.$$

It was proved in Ref. 12 that $\mathcal{N}_1$ is a subalgebra of $\mathcal{N}_\sigma$, moreover, $a \in \mathcal{N}_1$ if and only if $\sigma(a^*a) = \sigma(a)^*\sigma(a)$ and $\sigma(\sigma_t^{\omega\circ\sigma}(a)) = \sigma_t^\omega(\sigma(a))$. The restriction of σ to $\mathcal{N}_1$ is an isomorphism onto

$$\mathcal{M}_1 = \{b \in \mathcal{M}, \sigma \circ \sigma_\omega^*(b) = b\}\,.$$

The following theorem was proved in Ref. 12 in the case when φ_θ are faithful states. See Ref. 6 concerning the general case.

Theorem 2. *Let $\mathcal{M}$ and $\mathcal{N}$ be von Neumann algebras and let $\sigma : \mathcal{N} \to \mathcal{M}$ be a coarse-graining. Suppose that $(\mathcal{M}, \varphi_\theta)$ is a statistical experiment dominated by a state ω such that both ω and $\omega \circ \sigma$ are faithful and normal. Then the following properties are equivalent*:

(i) $\sigma(\mathcal{N}_\sigma)$ *is a sufficient subalgebra for* $(\mathcal{M}, \varphi_\theta)$,
(ii) σ *is a sufficient coarse-graining for* $(\mathcal{M}, \varphi_\theta)$,
(iii) $S_\alpha(\varphi_\theta \| \omega) = S_\alpha(\varphi_\theta | \mathcal{M}_0 \| \omega | \mathcal{M}_0)$ *for all* θ *and for some* $0 < |\alpha| < 1$,
(iv) $\sigma([D\varphi_\theta \circ \sigma, D\omega \circ \sigma]_t) = [D\varphi_\theta, D\omega]_t$,
(v) $\mathcal{M}_1$ *is a sufficient subalgebra for* $(\mathcal{M}, \varphi_\theta)$,
(vi) $\varphi_\theta \circ \sigma \circ \sigma_\omega^* = \varphi_\theta$.

The previous theorem applies to a measurement which is essentially a positive mapping $\mathcal{N} \to \mathcal{M}$ from a commutative algebra. The concept of sufficient measurement appeared also in Ref. 3. For a noncommuting family of states, there is no sufficient measurement.

We also have the following characterization of sufficient coarse-grainings in terms of relative entropy, see Ref. 10.

Proposition 4. *Under the conditions of Theorem* 2, *suppose that $S(\varphi_\theta \| \omega)$ is finite for all θ. Then σ is a sufficient coarse-graining if and only if*

$$S(\varphi_\theta \| \omega) = S(\varphi_\theta \circ \sigma \| \omega \circ \sigma)\,.$$

The equality in inequalities for entropy quantities was also studied in Refs. 17 and 18. For density matrices, it was shown that the equality in Proposition 4 is equivalent to

$$\sigma(\log \sigma^*(D_\theta) - \log \sigma^*(D_{\omega_0})) = \log D_\theta - \log D_\omega\,, \tag{12}$$

where σ^* is the dual mapping of σ on density matrices.

Let us now show how Theorems 1 and 2 can be applied if the dominating state ω is not faithful. Suppose that $p = \operatorname{supp}\omega$, $q = \operatorname{supp}\omega \circ \sigma$. We define the map $\alpha : q\mathcal{N}q \to p\mathcal{M}p$ by $\alpha(a) = p\sigma(a)p$. Then α is a coarse-graining such that $\alpha_\omega^* = \sigma_\omega^*$ and $\varphi_\theta \circ \sigma(a) = \varphi_\theta \circ \alpha(qaq)$ for all θ. We check that α is sufficient for $(p\mathcal{M}p, \varphi_\theta|_{p\mathcal{M}p})$ if and only if σ is sufficient for $(\mathcal{M}, \varphi_\theta)$. Indeed, let $\tilde{\beta} : p\mathcal{M}p \to q\mathcal{N}q$ be a coarse-graining such that $\varphi_\theta|_{p\mathcal{M}p} \circ \alpha \circ \tilde{\beta} = \varphi_\theta|_{p\mathcal{M}p}$ and let $\beta : \mathcal{M} \to \mathcal{N}$ be defined by

$$\beta(a) = \tilde{\beta}(pap) + \omega(a)(1 - q)\,.$$

Then β is a coarse-graining and

$$\varphi_\theta \circ \sigma \circ \beta(a) = \varphi_\theta \circ \sigma(q\beta(a)q)) = \varphi_\theta \circ \alpha \circ \tilde{\beta}(pap) = \varphi_\theta(pap) = \varphi_\theta(a)\,.$$

The converse is proved similarly, taking $\tilde{\beta}(a) = q\beta(a)q$ for $a \in p\mathcal{M}p$.

4. Exponential Families and Fisher Information

Let $\mathcal{M}$ be a von Neumann algebra and ω be a normal state. For $a \in \mathcal{M}^{sa}$ define the (perturbed) state $[\omega^a]$ as the minimizer of the functional

$$\psi \mapsto S(\psi\|\omega) - \psi(a) \tag{13}$$

defined on normal states of $\mathcal{M}$.

We define the *quantum exponential family* as

$$\mathcal{S} = \{\varphi_\theta := [\omega^{\sum_i \theta_i a_i}] : \theta \in \Theta\}\,, \tag{14}$$

where $a_1, a_2, \ldots, a_n$ are self-adjoint operators from $\mathcal{M}$ and $\Theta \subseteq \mathbb{R}^n$ is the parameter space. Let $\mathcal{M}$ be finite dimensional, and assume that the density of ω is written in the form e^H, $H = H^* \in \mathcal{M}$. Then the density of φ_θ is nothing but

$$\rho_\theta = \frac{\exp(H + \sum_i \theta_i a_i)}{\operatorname{Tr}\exp(H + \sum_i \theta_i a_i)}\,, \tag{15}$$

which is a direct analogue of the classical exponential family.

Returning to the general case, note that the support of the states φ_θ is $\operatorname{supp}\omega$. For more details about perturbation of states, see Chap. 12 of Ref. 9, here we recall the analogue of (15) in the general case. We assume that the von Neumann algebra is in a standard form and the representative of ω is the vector Ω from the positive cone of the Hilbert space. Let $\Delta_\omega \equiv \Delta(\omega/\omega)$ be the modular operator of ω, then φ_θ of (15) is the vector state induced by the unit vector

$$\Phi_\theta := \frac{\exp\frac{1}{2}(\log\Delta_\omega + \sum_i \theta_i a_i)\Omega}{\|\exp\frac{1}{2}(\log\Delta_\omega + \sum_i \theta_i a_i)\Omega\|}\,. \tag{16}$$

(This formula holds in the strict sense if ω is faithful, since Δ_ω is invertible in this case. For non-faithful ω the formula is modified by the support projection.)

In the next theorem σ_t^ω denotes the modular automorphism group of ω, $\sigma_t^\omega(a) = \Delta_\omega^{it} a \Delta_\omega^{-it}$.

Theorem 3.[10] *Let $\mathcal{M}$ be a von Neumann algebra with a faithful normal state ω and $\mathcal{M}_0$ be a subalgebra. For $a \in \mathcal{A}^{sa}$ the following conditions are equivalent*:

(i) $[D[\omega^a], D\omega]_t \in \mathcal{M}_0$ *for all* $t \in \mathbb{R}$,
(ii) $\sigma_t^\omega(a) \in \mathcal{M}_0$ *for all* $t \in \mathbb{R}$,
(iii) *For the generalized conditional expectation* $E_\omega : \mathcal{M} \to \mathcal{M}_0$, $E_\omega(a) = a$ *holds.*

Corollary 1. *Let $\mathcal{S}$ be the exponential family* (14) *and let $\mathcal{M}_0 \subseteq \mathcal{M}$ be a subalgebra. Then the following are equivalent*:

(i) $\mathcal{M}_0$ *is sufficient for* $(\mathcal{M}, \mathcal{S})$,
(ii) $\sigma_t^\omega(a_i) \in \mathcal{M}_0$ *for all* $t \in \mathbb{R}$ *and* $1 \le i \le n$,
(iii) $\mathcal{M}_0$ *is sufficient for* $(\mathcal{M}, \{[\omega^{a_1}], \ldots, [\omega^{a_n}]\})$.

Let us denote by $c(\omega, a)$ the minimum in (13), i.e. $c(\omega, a) = S([\omega^a]\|\varphi) - [\omega^a](a)$. Then the function $a \mapsto c(\omega, a)$ is analytic and concave. We recall that for $a,\ h \in \mathcal{M}^{sa}$,

$$\frac{d}{dt} c(\omega, a + th)|_{t=0} = -[\omega^a](h)\,.$$

Let us define for $a,\ h,\ k \in \mathcal{M}^{sa}$

$$\gamma_\omega(h,k) = -\frac{\partial^2}{\partial s \partial t} c(\omega, th + sk)|_{s=t=0} = -\frac{d}{dt}[\omega^{a+th}](k)|_{t=0}\,.$$

Then γ_ω is a positive bilinear form on $\mathcal{M}^{sa}$. It has an important monotonicity property: If $\alpha : \mathcal{N} \to \mathcal{M}$ is a faithful coarse-graining, then we have for any faithful state ω on $\mathcal{M}$ and a self-adjoint element $a \in \mathcal{N}$ that

$$\gamma_\omega(\alpha(a), \alpha(a)) \leq \gamma_{\omega\circ\alpha}(a, a)\,.$$

Note also that for $h,\ k \in \mathcal{M}^{sa}$ and $\lambda_1,\ \lambda_2 \in \mathbb{R}$,

$$\gamma_\omega(h + \lambda_1, k + \lambda_2) = \gamma_\omega(h, k)$$

and $\gamma_\omega(h, h) = 0$ implies $h = \lambda \in \mathbb{R}$.

Let now $\mathcal{S} = \{\varphi_\theta : \theta \in \Theta\}$ be a family of normal states on $\mathcal{M}$ and assume that the parameter space is an open subset $\Theta \subset \mathbb{R}^k$. Furthermore, we suppose that there exists a faithful normal state ω on $\mathcal{M}$, such that there are some constants $\lambda,\ \mu > 0$ satisfying

$$\lambda\omega \leq \varphi_\theta \leq \mu\omega \tag{17}$$

holds for every θ. If this condition holds, it remains true if we take any element in $\mathcal{S}$ in place of ω, we may therefore assume that $\omega \in \mathcal{S}$.

Condition (17) implies that for each $\theta \in \Theta$, there is some $a(\theta) \in \mathcal{M}^{sa}$, such that $\varphi_\theta = [\omega^{a(\theta)}]$. We will further assume that the function $\theta \mapsto a(\theta)$ is continuously differentiable and denote by ∂_i the partial derivative with respect to θ_i.

If $\alpha : \mathcal{N} \to \mathcal{M}$ be a coarse-graining, then for $\theta \in \Theta$, we have

$$\lambda\omega \circ \alpha \leq \varphi_\theta \circ \alpha \leq \mu\omega \circ \alpha\,,$$

so that the induced family again satisfies condition (17) and there are self-adjoint elements $b(\theta) \in \mathcal{N}$, such that $\varphi_\theta \circ \alpha = [\omega \circ \alpha^{b(\theta)}]$.

We have the following characterization of sufficient coarse-grainings under the above conditions:

Theorem 4. *Let $\alpha : \mathcal{N} \to \mathcal{M}$ be a faithful coarse-graining and let $\mathcal{S}$ be as above. Then α is sufficient for $(\mathcal{M}, \mathcal{S})$ if and only if for each θ there is some $b(\theta) \in \mathcal{N}^{sa}$, such that*

$$\varphi_\theta = [\omega^{\alpha(b(\theta))}] \quad \textit{and} \quad \varphi_\theta \circ \alpha = [\omega \circ \alpha^{b(\theta)}]\,. \tag{18}$$

Proof. Let $\omega = \varphi_{\theta_0} \in \mathcal{S}$ and let $\varphi_\theta = [\omega^{a(\theta)}]$. Let α be sufficient for $(\mathcal{M}, \mathcal{S})$ and let

$$\mathcal{N}_1 = \{a \in \mathcal{N} : \alpha^*_\omega \circ \alpha(a) = a\} = \{a \in \mathcal{N}_\alpha : \alpha(\sigma^{\omega\circ\alpha}_t(a)) = \sigma^\omega_t(\alpha(a))\}\,.$$

Then $\alpha(\mathcal{N}_1)$ is a sufficient subalgebra and by Theorems 3 and 2, $\sigma_t^\omega(a(\theta)) \in \alpha(\mathcal{N}_1)$ for all t, θ, in particular, $a(\theta) = \alpha(b(\theta))$, for some elements $b(\theta) \in \mathcal{N}_1$. Consider the expansion:

$$\begin{aligned}[D\omega^{\alpha(b(\theta))}, D\omega]_t &= \sum_{n=0}^{\infty} i^n \int_0^t dt_1 \cdots \int_0^{t_{n-1}} dt_n \sigma_{t_n}^\omega(\alpha(b(\theta))) \cdots \sigma_{t_1}^\omega(\alpha(b(\theta))) \\ &= \sum_{n=0}^{\infty} i^n \int_0^t dt_1 \cdots \int_0^{t_{n-1}} dt_n \alpha(\sigma_{t_n}^{\omega\circ\alpha}(b(\theta))) \cdots \alpha(\sigma_{t_1}^{\omega\circ\alpha}(b(\theta))) \\ &= \alpha([D\omega \circ \alpha^{b(\theta)}, D\omega \circ \alpha]_t)\,.\end{aligned}$$

On the other hand, α is sufficient, therefore $[D\varphi_\theta, \omega]_t \in \alpha(\mathcal{N}_\alpha)$ and

$$\alpha([D\varphi_\theta \circ \alpha, D\omega \circ \alpha]_t) = [D\varphi_\theta, D\omega]_t\,.$$

As α is invertible on $\mathcal{N}_\alpha$, it follows that $[D\varphi_\theta \circ \alpha, D\omega \circ \alpha]_t = [D[\omega \circ \alpha^{b(\theta)}], D\omega \circ \alpha]_t$ and we have (18).

Conversely, assume (18) holds, then

$$\partial_j c(\omega \circ \alpha, b(\theta)) = -[\omega \circ \alpha^{b(\theta)}](\partial_j b(\theta)) = -\varphi_\theta(\alpha(\partial_j b(\theta)) = \partial_j c(\omega, \alpha(b(\theta)))$$

for all θ and j. Putting $\theta = \theta_0$, it follows that $c(\omega \circ \alpha, b(\theta)) = c(\omega, \alpha(b(\theta)))$ for all θ. Hence

$$S(\varphi_\theta \| \omega) = c(\omega, \alpha(b(\theta))) - \varphi_\theta(\alpha(b(\theta))) = c(\omega\circ\alpha, b(\theta)) - \varphi_\theta\circ\alpha(b(\theta)) = S(\varphi_\theta\circ\alpha \| \omega\circ\alpha)$$

and α is sufficient. □

Note that the above theorem implies that if $\mathcal{S}$ is the exponential family (15) for some $a_1, \ldots, a_k \in \mathcal{M}^{sa}$, then the coarse-graining is sufficient if and only if $\varphi_\theta \circ \alpha$ is again an exponential family, $\varphi_\theta \circ \alpha = [\omega \circ \alpha^{\sum_i \theta_i b_i}]$ and $a_i = \alpha(b_i)$. In finite dimensions, the theorem reduces to equality (12).

Let us denote

$$\ell_i = \partial_i(a(\theta) - c(\omega, a(\theta))) = \partial_i a(\theta) - \varphi_\theta(\partial_i a(\theta))\,.$$

Then ℓ_i is a quantum version of the *score* in classical statistics. We define a Riemannian metric tensor on Θ by

$$g_{i,j}(\theta) = \gamma_{\varphi_\theta}(\ell_i, \ell_j)\,.$$

This is one of the quantum versions of the *Fisher information.*[16] Note that $g_{i,j}(\theta) = \gamma_{\varphi_\theta}(\partial_i a(\theta), \partial_j a(\theta))$ and if $a(\theta)$ is twice differentiable, then

$$g_{i,j}(\theta) = -\partial_i\partial_j c(\omega, a(\theta)) + \varphi_\theta(\partial_i\partial_j a(\theta))\,.$$

Next we show how sufficiency can be characterized by the Fisher information.

Theorem 5. *Let $\alpha : \mathcal{N} \to \mathcal{M}$ and $\mathcal{S}$ be as in the previous theorem. Let $g(\theta)$ and $h(\theta)$ be the Fisher information matrix for $\mathcal{S}$ and the induced family $\{\varphi_\theta \circ \alpha : \theta \in \Theta\}$, respectively. Then the matrix inequality*

$$h(\theta) \leq g(\theta)$$

holds. Moreover, equality is attained if and only if α is sufficient for $(\mathcal{M}, \mathcal{S})$.

Proof. Let $c = (c_1, \ldots, c_k) \in \mathbb{R}^k$, we have to show that

$$\sum_{i,j} c_i c_j h_{i,j}(\theta) \leq \sum_{i,j} c_i c_j g_{i,j}(\theta)$$

for all θ. Let $\varphi_\theta = [\omega^{a(\theta)}]$, $\varphi_\theta \circ \alpha = [\omega \circ \alpha^{b(\theta)}]$ and let us denote

$$\dot{b} = \frac{d}{dt} b(\theta + tc)|_{t=0} \in \mathcal{N}\,, \qquad \dot{a} = \frac{d}{dt} a(\theta + tc)|_{t=0} \in \mathcal{M}\,.$$

We have

$$\sum_{i,j} c_i c_j h_{i,j}(\theta) = \gamma_{\varphi_\theta \circ \alpha}(\dot{b}, \dot{b}) = -\frac{d}{dt}[\omega \circ \alpha^{b(\theta+tc)}](\dot{b})|_{t=0}$$

$$= -\frac{d}{dt}\varphi_{\theta+tc}\alpha((\dot{b}))|_{t=0} = \gamma_{\varphi_\theta}(\dot{a}, \alpha(\dot{b}))\,.$$

By Schwarz inequality and monotonicity of γ, we get

$$\gamma_{\varphi_\theta}(\dot{a}, \alpha(\dot{b}))^2 \leq \gamma_{\varphi_\theta}(\dot{a}, \dot{a})\gamma_{\varphi_\theta}(\alpha(\dot{b}), \alpha(\dot{b})) \leq \gamma_{\varphi_\theta}(\dot{a}, \dot{a})\gamma_{\varphi_\theta \circ \alpha}(\dot{b}, \dot{b})\,.$$

This implies that

$$\sum_{i,j} c_i c_j h_{i,j}(\theta) \leq \gamma_{\varphi_\theta}(\dot{a}, \dot{a}) = \sum_{i,j} c_i c_j g_{i,j}(\theta)\,.$$

Suppose that α is sufficient, then there is a coarse-graining $\beta : \mathcal{M} \to \mathcal{N}$, such that $\varphi_\theta = \varphi_\theta \circ \alpha \circ \beta$ and, by the first part of the proof, $g(\theta) \leq h(\theta)$, hence $g(\theta) = h(\theta)$.

Conversely, let $g(\theta) = h(\theta)$, and let us denote $a_i = \partial_i a(\theta)|_\theta$ and $b_i = \partial_i b(\theta)|_\theta$. Then

$$\partial_i \varphi_\theta(a_j)|_\theta = -g_{i,j}(\theta) = -h_{i,j}(\theta) = \partial_i \varphi_\theta \circ \alpha(b_j)\,.$$

It follows that

$$0 = \partial_i \varphi_\theta(\alpha(b_j) - a_j)|_\theta = \gamma_{\varphi_\theta}(a_i, a_j - \alpha(b_j))$$

for all i, j and θ. Therefore, we have for all i and θ,

$$\gamma_{\varphi_\theta}(\alpha(b_i), \alpha(b_i)) = \gamma_{\varphi_\theta}(\alpha(b_i) - a_i, \alpha(b_i) - a_i) + \gamma_{\varphi_\theta}(a_i, a_i)\,.$$

On the other hand, by monotonicity and the assumption, we have

$$\gamma_{\varphi_\theta}(\alpha(b_i), \alpha(b_i)) \leq \gamma_{\varphi_\theta \circ \alpha}(b_i, b_i) = \gamma_{\varphi_\theta}(a_i, a_i)\,.$$

This implies that $\partial_i \alpha(b(\theta)) - \partial_i a(\theta) = \lambda_i(\theta)$ for some $\lambda_i(\theta) \in \mathbb{R}$, for all i and θ. Since $a(\theta)$ and $b(\theta)$ are only determined up to a scalar multiple of 1 and we may assume that $b(\theta_0) = 0$, $a(\theta_0) = 0$, we may choose $b(\theta)$ so that $a(\theta) = \alpha(b(\theta))$ for all θ. By Theorem 4, α is sufficient. □

5. Factorization

Let $\mathcal{M}$ be a von Neumann algebra with a standard representation on a Hilbert space $\mathcal{H}$ and let ω be a faithful state on $\mathcal{M}$. Let $\mathcal{M}_0 \subset \mathcal{M}$ be a subalgebra and assume that it is invariant under the modular group σ_t^ω of ω. Let ω_0 be the restriction of ω to $\mathcal{M}_0$, then $\sigma_t^\omega|\mathcal{M}_0 = \sigma_t^{\omega_0}$.

Let ϕ and ϕ_0 be faithful normal semifinite weights on $\mathcal{M}$ and $\mathcal{M}_0$, respectively, then for $a \in \mathcal{M}_0$, we have

$$\Delta_{\omega,\phi}^{it} a \Delta_{\omega,\phi}^{-it} = \sigma_t^\omega(a) = \sigma_t^{\omega_0}(a) = \Delta_{\omega_0,\phi_0}^{it} a \Delta_{\omega_0,\phi_0}^{-it} .$$

It follows that there is a unitary $w_t \in \mathcal{M}_0'$, such that

$$\Delta_{\omega,\phi}^{it} = \Delta_{\omega_0,\phi_0}^{it} w_t .$$

Theorem 6. *Let $(\mathcal{M}, \mathcal{S})$ be a statistical experiment dominated by a faithful normal state ω. Let $\mathcal{M}_0 \subset \mathcal{M}$ be a von Neumann subalgebra invariant with respect to the modular group σ_t^ω. Then $\mathcal{M}_0$ is sufficient for $\mathcal{S}$ if and only if for each $t \in \mathbb{R}$, there is a unitary element $w_t \in \mathcal{M}_0'$, such that*

$$\Delta_{\varphi_\theta,\phi}^{it} = \Delta_{\varphi_{\theta,0},\phi_0}^{it} w_t , \qquad t \in \mathbb{R} \tag{19}$$

where $\varphi_{\theta,0} = \varphi_\theta|_{\mathcal{M}_0}$.

Proof. Let $\mathcal{M}_0$ be sufficient for $(\mathcal{M}, \mathcal{S})$, then $[D\varphi_\theta, D\omega]_t = [D\varphi_{\theta,0}, D\omega_0]_t$ for all θ and t. It follows that

$$\Delta_{\varphi_\theta,\phi}^{it} = [D\varphi_\theta, D\omega]_t \Delta_{\omega,\phi}^{it} = [D\varphi_{\theta,0}, D\omega_0]_t \Delta_{\omega_0,\phi_0}^{it} w_t = \Delta_{\varphi_{\theta,0},\phi_0}^{it} w_t .$$

Conversely, suppose (19), then

$$[D\varphi_\theta, D\omega]_t = \Delta_{\varphi_\theta,\phi}^{it} \Delta_{\omega,\phi}^{-it} = \Delta_{\varphi_{\theta,0},\phi_0}^{it} w_t w_t^* \Delta_{\omega_0,\phi_0}^{-it} = [D\varphi_{\theta,0}, D\omega_0]_t$$

and $\mathcal{M}_0$ is sufficient. □

Let $\mathcal{M}_1 = \mathcal{M}_0' \cap \mathcal{M}$ be the relative commutant, then $\mathcal{M}_1$ is invariant under σ_t^ω as well and $\sigma_t^\omega|\mathcal{M}_1 = \sigma_t^{\omega_1}$, where $\omega_1 = \omega|_{\mathcal{M}_1}$. Suppose further that the subalgebra $\mathcal{M}_0$ is semifinite and let ϕ_0 be a trace. Then $\Delta_{\omega_0,\phi_0}^{it}$, $\Delta_{\varphi_{\theta,0},\phi_0}^{it} \in \mathcal{M}_0$ for all θ and there is an operator Δ affiliated with $\mathcal{M}_0'$, such that $w_t = \Delta^{it}$. Moreover, for $a \in \mathcal{M}_1$,

$$\sigma_t^{\omega_1}(a) = \sigma_t^\omega(a) = w_t a w_t^* = \Delta^{it} a \Delta^{-it} .$$

The factorization (19) has a special form, if we require that the entropy of the state ω is finite. Recall that the *entropy* of a state φ of a C^*-algebra is defined as

$$S(\varphi) := \sup \left\{ \sum_i \lambda_i S(\varphi_i \| \varphi) : \sum_i \lambda_i \varphi_i = \varphi \right\} ,$$

see (6.9) in Ref. 9. If $S(\omega) < \infty$, then $\mathcal{M}$ must be a countable direct sum of type I factors, see Theorem 6.10 in Ref. 9. As the subalgebras $\mathcal{M}_0$ and $\mathcal{M}_1$ are invariant under σ_t^ω, we have by Proposition 6.7 in Ref. 9 that $S(\omega_0)$, $S(\omega_1) \leq S(\omega) < \infty$. It follows that both $\mathcal{M}_0$ and $\mathcal{M}_1$ must be countable direct sums of type I factors as well.

Let ϕ and ϕ_0 be the canonical traces and let ρ_ω, ρ_θ and $\rho_{\theta,0}$, ρ_{ω_0} be the density operators. Then $w_t = \rho_{\omega_0}^{-it}\rho_\omega^{it} \in \mathcal{M}_0' \cap \mathcal{M} = \mathcal{M}_1$ and since $\sigma_t^{\omega_1}(a) = w_t a w_t^*$, we have $w_t = \rho_{\omega_1}^{it} z^{it}$ for a central element z in $\mathcal{M}_1$ and a density operator ρ_{ω_1} in $\mathcal{M}_1$. Putting all together, we get that sufficiency is equivalent with

$$\rho_\theta = \rho_{\theta,0}\rho_{\omega_1} z\,, \qquad \theta \in \Theta\,. \tag{20}$$

The essence of this factorization is that the first factor is the reduced density and the rest is independent of θ.

Since $\mathcal{M}_1$ is a countable direct sum of factors of type I, there is an orthogonal family of minimal central projections p_n, $\sum_n p_n = 1$. Moreover, there is a decomposition

$$\mathcal{H}_n = p_n\mathcal{H} = \mathcal{H}_n^L \otimes \mathcal{H}_n^R\,,$$

such that

$$\mathcal{M}_1 = \bigoplus_n \mathbb{C}I_{\mathcal{H}_n^L} \otimes B(\mathcal{H}_n^R)\,, \qquad (\mathcal{M}_1)' = \bigoplus_n B(\mathcal{H}_n^L) \otimes \mathbb{C}I_{\mathcal{H}_n^R}\,.$$

From this, we get

$$\rho_\theta = \rho_{\theta,0}\rho_{\omega_1} z = \sum_n \varphi_\theta(p_n)\rho_n^L(\theta) \otimes \rho_n^R\,, \tag{21}$$

where ρ_n^R is a density operator in $B(\mathcal{H}_n^R)$ and $\rho_n^L(\theta)$ is a density operator in $B(\mathcal{H}_n^L)$.

A particular example of a sufficient subalgebra is the subalgebra generated by the partial isometries $\{[D\varphi_\theta, D\omega]_t : t \in \mathbb{R}\}$, this subalgebra is minimal sufficient and invariant under σ_t^ω. If $S(\omega) < \infty$, the decomposition (20), corresponding to this subalgebra is a maximal such decomposition, in the sense that the density operator $\rho_{\theta,0}$ cannot be decomposed further, in a nontrivial way.

Example 8. Let $\mathcal{H}$ be a finite dimensional Hilbert space, let $\mathcal{S}$ be a family of pure states induced by the unit vectors $\{\xi_\theta : \theta \in \Theta\}$. Suppose that the vectors ξ_θ generate $\mathcal{H}$, then there is a faithful state ω, dominating $\mathcal{S}$. Let

$$\mathcal{A}_0 = \oplus_{j=1}^m B(\mathcal{H}_j^L) \otimes \mathbb{C}I_{\mathcal{H}_j^R}$$

be a subalgebra in $B(\mathcal{H})$, invariant under σ_t^ω and suppose that $\mathcal{A}_0$ is sufficient for $\mathcal{S}$. Then, we have from (21) that for each θ, there is some $1 \leq j \leq m$ and unit vectors $\xi_{\theta,j} \in \mathcal{H}_j^L$, $\xi_j \in \mathcal{H}_j^R$, such that

$$\xi_\theta = \xi_{\theta,j} \otimes \xi_j\,.$$

Suppose that there are θ_1, $\theta_2 \in \Theta$, such that $\xi_{\theta_i} = \xi_{\theta_i,j_i} \otimes \xi_{j_i}$, $i = 1, 2$ and $j_1 \neq j_2$, then ξ_{θ_1} and ξ_{θ_2} must be orthogonal. Consequently, if, for example, the family $\mathcal{S}$ contains no two orthogonal vectors, then $m = 1$, $\mathcal{A}_0$ must be of the form $\mathcal{A}_0 = B(\mathcal{H}_L) \otimes \mathbb{C}I_{\mathcal{H}_R}$ and $\xi_\theta = \xi_{\theta,L} \otimes \xi_R$ for all θ. □

Example 9. Let us return to the experiment $(\mathcal{M}^{\otimes n}, \{\varphi_\theta^{\otimes n}\})$ of Example 6. Let $\mathcal{M} = B(\mathcal{H})$ and let π be the unitary representation of the permutation group $S(n)$ on $\mathcal{H}^{\otimes n}$, then $\mathcal{N} = \pi(S(n))'$. There is a decomposition $\pi = \oplus_{i,j}\pi_{i,j}$, such that all $\pi_{i,j}$ are irreducible representations and $\pi_{i,j}$, $\pi_{k,l}$ are equivalent if and only if $i = k$. It follows that there is a decomposition $\mathcal{H}^{\otimes n} = \oplus_k \mathcal{H}_k^L \otimes \mathcal{H}_k^R$ such that

$$\mathcal{N} = \bigoplus_k B(\mathcal{H}_k^L) \otimes \mathbb{C}I_{\mathcal{H}_k^R}\,.$$

Let ω be a state dominating φ_θ, $\theta \in \Theta$, then $\omega^{\otimes n}$ dominates $\varphi^{\otimes n}$, $\theta \in \Theta$. Since $\mathcal{N}$ is also invariant under the modular group $\sigma_t^{\omega^{\otimes n}}$, we conclude that the densities decompose as

$$\rho_\theta^{\otimes n} = \sum_k \lambda_k \rho_k^L(\theta) \otimes \rho_k^R\,,$$

for density matrices $\rho_k^L(\theta) \in B(\mathcal{H}_k^L)$ and $\rho_k^R \in B(\mathcal{H}_k^R)$.

Acknowledgments

A.J. was supported by the EU Research Training Network Quantum Probability with Applications to Physics, Information Theory and Biology. D.P. was supported by the Hungarian grant OTKA T032662.

References

1. L. Accardi and C. Cecchini, Conditional expectations in von Neumann algebras and a theorem of Takesaki, *J. Funct. Anal.* **45** (1982) 245–273.
2. O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, Wiley Series in Probability and Mathematical Statistics (John Wiley & Sons, 1978).
3. O. E. Barndorff-Nielsen, R. Gill and P. E. Jupp, On quantum statistical inference, *J. R. Statist. Soc. Ser. B Stat. Methodol.* **65** (2003) 775–816.
4. J. Blank, P. Exner and M. Havliček, *Hilbert Space Operators in Quantum Physics* (Amer. Inst. Phys., 1994).
5. O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics. 1. C*- and W*-Algebras, Symmetry Groups, Decomposition of States*, 2nd edn., Texts and Monographs in Physics (Springer-Verlag, 1987).
6. A. Jenčová and D. Petz, Sufficient quantum coarse-grainings, *Commun. Math. Phys.* **263** (2006) 259–276.
7. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
8. M. A. Nielsen and D. Petz, A simple proof of the strong subadditivity inequality, *Quantum Inf. Comput.* **6** (2005) 507–513.
9. M. Ohya and D. Petz, *Quantum Entropy and Its Use*, 2nd edn. (Springer-Verlag, 2004).

10. D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, *Commun. Math. Phys.* **105** (1986) 123–131.
11. D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **21** (1986) 57–65.
12. D. Petz, Sufficiency of channels over von Neumann algebras, *Quart. J. Math.* **39** (1988) 907–1008.
13. D. Petz, *An Invitation to the Algebra of the Canonical Commutation Relation* (Leuven Univ. Press, 1990).
14. D. Petz, Geometry of canonical correlation on the state space of a quantum system, *J. Math. Phys.* **35** (1994) 780–795.
15. D. Petz, Discrimination between states of a quantum system by observations, *J. Funct. Anal.* **120** (1994) 82–97.
16. D. Petz, Covariance and Fisher information in quantum mechanics, *J. Phys. A: Math. Gen.* **35** (2002) 929–939.
17. D. Petz, Monotonicity of quantum relative entropy revisited, *Rev. Math. Phys.* **15** (2003) 79–91.
18. M. B. Ruskai, Inequalities for quantum entropy: A review with conditions with equality, *J. Math. Phys.* **43** (2002) 4358–4375.
19. Ş. Strătilă, *Modular Theory of Operator Algebras* (Abacus Press, 1981).
20. H. Strasser, *Mathematical Theory of Statistics. Statistical Experiments and Asymptotic Decision Theory* (Walter de Gruyter, 1985).