PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 4, Fasc. 2 (1984), p. 237-244

ON THE INFORMATION-THEORETICAL FOUNDATIONS OF QUANTUM STATISTICS

BY

RYSZARD JAJTE (ŁÓDŹ)

Abstract. In the paper we study a few basic definitions and facts concerning statistical quantities for a quantum-theoretic measurement problem. The discussion is in the spirit of the information-theoretical foundations of classical statistics.

0. We are going to discuss in a very simple way some basic problems of quantum statistics from the point of view of information theory. This leads us naturally to the W^*-algebra setting of quantum theory, where the bounded observables of a physical system are represented by the self-adjoint elements of a W^*-algebra, and the states are represented by positive linear functionals on this algebra [2], [3]. Our approach has been suggested by some works of Rényi ([10], [11]) on the foundations of classical statistics. Another approach has been presented in [5]-[9] where rich bibliographical information can be found.

We shall consider a compound system which consists, roughly speaking, of two parts: parameter hypothesis θ and measurement ξ. That is why we shall deal with the tensor product of two von Neumann algebras (and a state on this product algebra).

1. Let us begin with some notation and definitions. Throughout the paper, $\mathscr{A}$ will denote a von Neumann algebra (acting in a Hilbert space H). $\mathscr{A}\otimes\mathscr{B}$ will stand for the W^*-tensor product of two W^*-algebras $\mathscr{A}$ and $\mathscr{B}$ (see [1] and [12]). μ will denote a normal faithful state on $\mathscr{A}\otimes\mathscr{A}$.

Now, we are in a position to introduce suitable definitions.

1.1. The *parameter hypothesis* is given by a self-adjoint operator of the form

$$\theta = \sum_{k=1}^{N} \theta_k p_k.$$

The projectors p_k corresponding to the eigenvalues θ_k play the role of "questions" about the values of the parameter θ, and p_k is interpreted as the hypothesis that the true value of the parameter θ is equal to θ_k.

The number

$$H(\theta) = -\sum_{k=1}^{N} \mu(p_k \otimes 1) \lg \mu(p_k \otimes 1)$$

will be called the *(a priori) entropy* of θ at the state μ.

1.2. The measurement is represented by a self-adjoint operator ξ affiliated to $\mathscr{A}$, with the spectral representation

$$\xi = \int_{-\infty}^{\infty} \mu \mathrm{E}_{\xi}(du),$$

where $\mathrm{E}_{\xi}(\cdot)$ denotes the spectral measure of ξ. Thus $\mu\big(1 \otimes \mathrm{E}_{\xi}(\cdot)\big)$ is a Borel probability measure on the real line, and $\mu\big(1 \otimes \mathrm{E}_{\xi}(Z)\big)$ will be interpreted as the probability that the measurement ξ (in our compound system) gives values from the Borel set Z.

1.3. We define the *a posteriori entropy* $H(\theta|\xi)$ (of θ after the measurement ξ has been accomplished) by the formula

$$H(\theta|\xi) = -\sum_{k=1}^{N} q_k \lg q_k = \sum_{k=1}^{N} \psi(q_k)$$

with $\psi \colon x \to x \lg x$ $\big(x \in (0, 1)\big)$ and

$$q_k = \mathrm{E}_{\mu}(p_k \otimes 1 \mid 1 \otimes \xi), \tag{1}$$

where $\mathrm{E}_{\mu}(\cdot|\cdot)$ denotes the conditional expectation ([13], [14]) and $\psi(q_k)$ is defined by the Spectral Theorem. Because of the commutativity of the operators $p \otimes 1$ and $1 \otimes \xi$, the conditional expectation E_{μ} can be described in terms of the Radon-Nikodym derivatives. Therefore, we can write

$$q_k = \int f_k(u) \mathrm{E}_{1 \otimes \xi} \qquad (k = 1, \ldots, N). \tag{2}$$

The operator $H(\theta|\xi)$ is interpreted as the physical quantity describing the amount of information concerning θ still missing after having observed ξ. The number

$$R(\theta|\xi) = \mu\big(H(\theta|\xi)\big)$$

is called the *average amount of this information.*

2. The "separation" of the operators p and ξ by taking their commuting extensions $p \otimes 1$ and $1 \otimes \xi$ makes it possible to imitate the classical procedures in a rather true way. When one repeats the measurement ξ, the state μ defined on the product $\mathscr{A} \otimes \mathscr{A}$ can (and should) reflect the alterations of the state in the algebra $\mathscr{A}$, caused by the measurements ξ. The arbitrariness of the choice of the state μ in the tensor product of algebras enables us to establish the entropy measure somewhat according to the physicist's wish, which seems to permit him to accentuate those aspects of the experiment which are more interesting for him. Connections with the classical situation are conspicuous. For a formal proof of the properties discussed in the article, one can construct suitable commutative von Neumann algebras, make use of their isomorphism with the algebras $L^{\infty}(M, m)$ and, in consequence, of the classical results (for the algebras $L^{\infty}(M, m)$ and elements affiliated with them). It seems, however, that because of the simplicity of direct considerations, it is more purposeful to disregard the possibilities the theorem on representation of W^*-algebras gives and to carry out the proofs directly. We shall only confine ourselves to outlining them since they imitate the classical procedures.

The amount of information $I(\theta|\xi)$, contained in the observation ξ (concerning the parameter θ), is given by the difference

$$I(\theta|\xi) = H(\theta)-R(\theta|\xi).$$

We have

$$R(\theta|\xi) \leqslant \sum_k \psi\big(\mu(q_k)\big) = \sum_k \psi|\mu(p_k \otimes 1)| = H(\theta);$$

so $I(\theta|\xi) = H(\theta)-R(\theta|\xi) \geqslant 0$. It is easy to show that if μ is a product state $\mu = \nu_1 \otimes \nu_2$ over $\mathscr{A}\otimes\mathscr{A}$, then $I(\theta|\xi) = 0$.

2.1. Let $g\colon R \to R^m$ be a Borel function. The vector with operator-valued components $g(\xi) = \big(g_1(\xi), \ldots, g_m(\xi)\big)$ is called a *statistic*. Instead of $R(\theta|\xi)$ we can consider the quantity $R\big(\theta|g(\xi)\big)$. The definition of $R\big(\theta|g(\xi)\big)$ is evident. Namely, we put

$$H\big(\theta|g(\xi)\big) = -\sum_k Q_k \lg Q_k = \sum_k \psi(Q_k)$$

with

$$Q_k = \mathscr{E}_\mu\big(p_k \otimes 1 \mid 1 \otimes g(\xi)\big) = \mathscr{E}_\mu\big(p_k \otimes 1 \mid W^*\big(1\otimes g_1(\xi), \ldots, 1 \otimes g_m(\xi)\big)\big), \tag{3}$$

where $\mathscr{E}_\mu\big(\cdot|W^*(x, \ldots)\big)$ denotes the conditional expectation with respect to the W^*-subalgebra generated by x's (generated by the elements $1\otimes g_1(\xi), \ldots, 1\otimes g_m(\xi)$ in our case).

If we take into consideration $g(\xi)$ instead of ξ, then only the information about θ contained in $g(\xi)$ is taken into account and, as a rule, some information is lost. That is why, putting

$$I\big(\theta|g(\xi)\big) = H(\theta)-R\big(\theta|g(\xi)\big),$$

we may expect that

$$I(\theta|g(\xi)) \leqslant I(\theta|\xi)$$

or, equivalently,

$$R(\theta|\xi) \leqslant R(\theta|g(\xi)). \tag{4}$$

We prove (4). Note that the operators q_k and Q_k (defined by (1) and (3)) are positive, and

$$\sum_{k=1}^{N} q_k = \sum_{k=1}^{N} Q_k = 1.$$

Moreover, q_k and Q_k belong to the commutative algebra $W^*(1\otimes\xi)$. More exactly, they can be represented by functions defined on the interval $(a, b) \supset \mathrm{spectrum}(1\otimes\xi)$ (a, b may be infinite):

$$q_k = \int_a^b f_k(u)\,\mathrm{E}_{1\otimes\xi}(du), \qquad Q_k = \int_a^b F_k(u)\,\mathrm{E}_{1\otimes\xi}(du).$$

It is easy to check that for $\alpha_k \geqslant 0$ and $\beta_k \geqslant 0$ such that

$$\sum_k \alpha_k = \sum_k \beta_k = 1$$

the inequality

$$\sum_k \alpha_k \lg\alpha_k \geqslant \sum_k \alpha_k \lg\beta_k$$

holds. Thus

$$-\sum_k f_k(u)\lg f_k(u) \leqslant -\sum_k f_k(u)\lg F_k(u).$$

Integrating this inequality with respect to $\nu_\xi(\cdot) = \mu\big(1\otimes\mathrm{E}_\xi(du)\big)$, we obtain

$$R(\theta|\xi) = \mu(-\sum_k q_k\lg q_k) \leqslant \mu(-\sum_k q_k\lg Q_k).$$

To get (4), it is enough to show that

$$\mu(q_k\lg Q_k) = \mu(Q_k\lg Q_k).$$

But $q_k \in W^*(1\otimes\xi)$, $Q_k \in W^*\big(1\otimes g_j(\xi), j = 1,\ldots,N\big)$ and, evidently,

$$W^*(1\otimes\xi) \supseteq W^*\big(1\otimes g_j(\xi), j = 1,\ldots,N\big),$$

which completes the proof.

2.2. We call a statistic $g = (g_1,\ldots,g_m)$ *sufficient* for θ if

$$I(\theta|g(\xi)) = I(\theta|\xi).$$

The statistic $q(\xi) = (q_1, \ldots, q_N)$, where q_j are defined by (1) is sufficient. Indeed, it is enough to verify that

$$q_k = \mathscr{E}_\mu(p_k \otimes 1 \mid q_1, \ldots, q_N),$$

and this follows immediately from the definition of q_j.

3. In this section we discuss the choice of an optimal decision concerning the parameter θ (made on the basis of the observation ξ). By a *decision* we mean an operator of the form

$$\Delta = \sum_{k=1}^{N} \theta_k \Pi_k(\xi), \tag{5}$$

where $\Pi_1(\xi), \ldots, \Pi_N(\xi)$ are pairwise orthogonal self-adjoint projectors from $W^*(\xi)$ such that

$$\sum_{k=1}^{N} \Pi_k(\xi) = 1.$$

The operator $\Pi_k(\xi)$ will be interpreted as a decision that the true value of θ is equal to θ_k. The probability of such a decision is equal to $\mu(1 \otimes \Pi_k(\xi))$.

We shall compare decision operators (5) with the parameter operator

$$\theta = \sum_{k=1}^{N} \theta_k p_k.$$

3.1. A *standard decision operator* Δ_0 is defined by the formula

$$\Delta_0 = \sum_{k=1}^{N} \theta_k \Pi_k^0,$$

where $\Pi_j^0 = \mathrm{E}_\xi(Z_j)$ and $(Z_1, \ldots, Z_N)$ is an arbitrary (Borel measurable) decomposition of the interval $[a, b]$ such that

$$Z_j \subset \{u \in [a, b] : f_j(u) = \max_{1 \leqslant k \leqslant N} f_k(u)\}, \tag{6}$$

the functions f_k being the Radon-Nikodym derivatives defined by (2). Of course, $\Pi_j^0 \in W^*(\xi)$ for $j = 1, \ldots, N$. From what follows it will be clear that the standard decision does not depend on the choice of the decomposition $(Z_1, \ldots, Z_N)$ (in the case where there is more than one decomposition of the kind described above, satisfying condition (6)).

3.2. In this section we show that the standard decision Δ_0 is optimal in a natural sense. We treat a pair (θ, Δ) as a random vector by putting

$$\Pr(\theta = \theta_k, \Delta = \theta_l) = \mu(p_k \otimes \Pi_l) \quad \text{for } k, l = 1, \ldots, N.$$

Then the error δ of the decision Δ is defined by $\delta = \Pr(\Delta \neq \theta)$.

Denoting by δ_0 the error of the standard decision Δ_0, we show that

$$\delta_0 \leqslant \delta. \tag{7}$$

Indeed, let (5) be an arbitrary decision. Then Π_k must be of the form $\Pi_k = \mathrm{E}_\xi(V_k)$ $(k = 1, \ldots, N)$, where $(V_1, \ldots, V_N)$ is a decomposition of the interval $[a, b]$ (because $\Pi_k \in W^*(\xi)$), and we have

$$\delta = \Pr(\Delta \neq \theta) = 1 - \Pr(\Delta = \theta) = 1 - \sum_{k=1}^{N} \mu\big(p_k \otimes \mathrm{E}_\xi(V_k)\big).$$

But

$$\mu(p_k \otimes \Pi_k) = \int_{V_k} f_k(u)\, \nu_\xi(du),$$

where $\nu_\xi(\cdot) = \mu\big(\mathrm{E}_{1\otimes\xi}(\cdot)\big)$ and f_k is the corresponding Radon-Nikodym derivative. Thus

$$\delta = 1 - \sum_k \int_{V_k} f_k(u)\, \nu_\xi(du).$$

In particular,

$$\delta_0 = 1 - \sum_k \int_{Z_k} f_k(u)\, \nu_\xi(du).$$

By the definition of the sets Z_k, we have obviously (7).

3.3. In this section we prove that the information-theoretical point of view is in accordance with the usual procedures of statistics. Namely, we show that for every sequence of observations $\{\xi_n\}$ the following conditions are equivalent:

$$R(\theta|\xi_n) \to 0 \quad \text{and} \quad \delta_n = \delta\big(\Delta_0(\xi_n)\big) \to 0,$$

where δ_0 stands for the corresponding errors of standard decisions $\Delta_0(\xi_n)$.

In other words, if the information about parameter θ, missing after having observed ξ_n, is small, then the error of the corresponding standard decision $\Delta_0(\xi_n)$ is small, and conversely.

To show this we prove the following two inequalities:

$$-\lg(1-\delta_\xi) \leqslant R(\theta|\xi) \tag{8}$$

and

$$R(\theta|\xi) \leqslant \delta_\xi \lg(N-1) + \delta_\xi \lg \delta_\xi + (1-\delta_\xi)\lg(1-\delta_\xi) \tag{9}$$

(cf. [4], Section 3.3).

We first prove (8). We have

$$R(\theta|\xi) = -\sum_{j,k} \int_{Z_j} f_k(u) \lg f_k(u)\, \nu_\xi(du),$$

where $\nu_\xi(du) = \mu[\mathrm{E}_{1\otimes\xi}(du)]$. Using the definition of Z_j, we can write

$$R(\theta|\xi) \geqslant -\sum_j \int_{Z_j} \lg f_j(u)\,\nu_\xi(du).$$

Using twice the Jensen inequality, we obtain

$$\begin{aligned} P(\theta|\xi) &\geqslant -\sum_j \nu_\xi(Z_j)\lg\frac{1}{\nu_\xi(Z_j)}\int_{Z_j} f_j(u)\,\nu_\xi(du) \\ &\geqslant -\lg\sum_j \int_{Z_j} f_j(u)\,\nu_\xi(du) = -\lg\Pr\big(\varDelta(\xi)=\theta\big) = -\lg(1-\delta_\xi), \end{aligned}$$

which completes the proof of (8).

To prove (9) observe that

$$R(\theta|\xi) \leqslant \sum_j \int_{Z_j}\left[-f_j(u)\lg f_j(u) + \big(1-f_j(u)\big)\lg\frac{N-1}{1-f_j(u)}\right]\nu_\xi(du).$$

Since

$$(10)\qquad \sum_j \int_{Z_j} f_j(u)\,\nu_\xi(du) = 1-\delta,$$

we get

$$\begin{aligned} R(\theta|\xi) \leqslant \delta\lg(N-1) &- \sum_j \int_{Z_j} f_j(u)\lg f_j(u)\,\nu_\xi(du) - \\ &- \sum_j \int_{Z_j}\big(1-f_j(u)\big)\lg\big(1-f_j(u)\big)\,\nu_\xi(du). \end{aligned}$$

The last formula leads to (8) by (9) and the formula for the conditional entropy. The proof is completed.

REFERENCES

[1] J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien*, 2ème éd., Gautier-Villars, Paris 1969.

[2] G. G. Emch, *Algebraic methods in statistical mechanics and quantum field theory*, New York 1972.

[3] D. E. Evans and J. T. Lewis, *Dilations of irreversible evolutions in algebraic quantum theory*, Comm. Dublin Inst. Adv. Studies, Series A (Theoretical Physics), No. 24, Dublin 1977.

[4] A. Feinstein, *Foundations on information theory*, New York 1958.

[5] S. Gudder, *Statistical inference in quantum mechanics*, Rep. Math. Phys. 17 (1980), p. 265-274.

[6] C. W. Helstrom, *Noncommuting observables in quantum detection and estimation theory*, IEEE Trans. Information Theory, IT-20, No. 1 (1974), p. 16-24.

[7] A. S. Holevo, *Information theory aspects of quantum measurement*, Problems Inform. Transmission 9 (1973), p. 31-42.

[8] — *Statistical decision theory for quantum systems*, J. Multivariate Analysis 3 (4) (1973), p. 337-394.

[9] G. Lindblad, *Entropy, information and quantum measurements*, Comm. Math. Phys. 33 (1973), p. 305-322.

[10] A. Rényi, *On the amount of missing information and the Neyman-Pearson lemma*, p. 281-288 in: *Festschrift for J. Neyman*, Wiley, London 1966.

[11] — *Statistics and information theory*, Studia Sci. Math. Hungar. 2 (1967), p. 249-256.

[12] S. Sakai, *C^*-algebras and W^*-algebras*, Springer-Verlag, Berlin 1971.

[13] M. Takesaki, *Conditional expectation in von Neumann algebras*, J. Functional Analysis 9 (1972), p. 306-321.

[14] J. Tomijama, *On the projection of norm one in W^*-algebras*, Proc. Japan Acad. 33 (1957), p. 608-612.

Institute of Mathematics
Łódź University
Banach 22, 90-456 Łódź, Poland

Received on 18. 5. 1981;
revised version on 5. 11. 1982