CONDITIONAL EXPECTATION IN AN OPERATOR ALGEBRA, IV (ENTROPY AND INFORMATION)

By Hisaharu Umegaki

1. Introduction.

The theory of information, created by Shannon [23], is developed by Feinstein, Kullback, MacMillan, Wiener and other American statisticians (e. g., cf. [10]), and also advanced into the ergodic theory by Gelfand, Khinchin, Kolmogorov, Yaglom and other Russian probabilists (e. g., cf. [8]). Through recent years, the theory is regarded as a new chapter in the theory of probability.

Recently, Segal [22] gave a mathematical formulation of the entropy of state of a von Neumann algebra, which contains both the cases for the theory of information and the theory of quantum statistics. Segal's theorem was reformulated in operator algebraic form by Nakamura and Umegaki [16] and independently by Davis [3].

Since the summer in 1954, Nakamura and Umegaki have investigated the concept of the conditional expectation in von Neumann algebra as a non-commutative extension of probability theory (cf. for example [13~18] and [25 ~28]), and in the most recent paper [18] it was applied to the theory of measurements of quantum statistics which is regarded as a non-commutative case of the theory of entropy and information. Furthermore, it may be very intersesting to develope the theory of information under functional-analysistic and operator-theoretic methods. From these points of views, we shall discuss the measure of information of integrable operators or of normal states of a von Neumann algebra. Davis [3] has independently studied on the almost same theme with Nakamura-Umegaki [16] and [18], in which he developed the theory of entropy and he simplified the proof of the theorem relative to the operator-entropy.

Now, we shall give the basic notations and describe the fundamental concepts in a von Neumann algebra which will be used throughout the present paper.

Let A be a von Neumann algebra, that is, A is a weakly closed self-adjoint algebra of bounded operators acting over a complex Hilbert space H, which contains the identity operator I. A linear functional ρ of A is said to be *positive* if $\rho(aa^*)\geqq 0$ for every $a \varepsilon A$. Such ρ is said to be *state* if $\rho(I)=1$, to be *normal* in the terminology of Dixmier [4] if $\rho(a_\alpha)\uparrow\rho(a)$ for $a_\alpha\uparrow a$, and to be *trace* if $\rho(ab)=\rho(ba)$ for every pair $a, b \varepsilon A$. The normality of state is equivalent to the complete additivity: $\sum\rho(p_\alpha)=\rho(\sum p_\alpha)$ for any disjoint family of projections $\{p_\alpha\}\subset A$ (cf. Dixmier [4]).

Received December 1, 1961.

Throughout this paper, as far as we give no exposition on each part, it will be assumed that A *have a faithful normal trace* τ. It is known, that A has such τ if and only if A is of finite class and sigma-finite, and in general that a von Neumann algebra has sufficiently many normal traces if and only if it is of finite class (cf. Dixmier [4]).

Denote $L^p(A,\tau)$ or simply $L^p(A)$ ($p \geqq 1$) the space of all measurable operators x with finite integral $(\tau(|x|^p))^{1/p}$ ($=\|x\|_p$, say L^p-norm) in the sense of Segal [21]. Then $L^p(A)$ is a Banach space. If ρ is a normal state of A, then there is so-called Radon-Nikodym derivative $d\rho/d\tau$ in the sense of Dye [6] and Segal [21] such that $\rho(a) = \tau((d\rho/d\tau)a)$ for every $a \varepsilon A$. Then $d\rho/d\tau$ is a self-adjoint non-negative operator ($d\rho/d\tau \geqq 0$, say) belonging to $L^1(A)$ and has norm one.

Denote $s(a)$ the support-projection of a self-adjoint operator a. For a pair of such operators a, b, if $s(a) \leqq s(b)$, then denote $a \prec b$. When both $a \prec b$ and $b \prec a$ hold, denote $a \sim b$. For a pair of normal states σ, ρ of A, if σ is absolutely continuous with respect ρ, then denote $\sigma \prec \rho$. When both $\sigma \prec \rho$ and $\rho \prec \sigma$ hold, denote $\sigma \sim \rho$. If a and b are Radon-Nikodym derivatives $d\sigma/d\tau$ and $d\rho/d\tau$, respectively, then $a \prec b$ or $a \sim b$ are equivalent to $\sigma \prec \rho$ or $\sigma \sim \rho$, respectively.

For any family F of bounded operators, denote F' the von Neumann algebra consists of all bounded operators commuting with the operators in F. For any self-adjoint $a = \int \lambda de_\lambda$, denote $R(a)$ the von Neumann algebra generated by a, i.e., $R(a) = \{e_\lambda\}_\lambda''$.

In the present paper, §§2 and 3 contain the basic notions and the preliminaries for §4 ~ §9. In §2, the back-grand and the fundamental properties of the conditional expectation in A will be described and as preliminary lemma the domain of it will be extended to a family of measurable operators which are not necessarily bounded or integrable. In §3, it will be stated on the operator-functions $\log a$ and $a \log a$ for $a \varepsilon A$, $a \geqq 0$, which will become the basic notations to discuss the measures of entropy and information, and it will be also stated on the two basic theorems of the operator-entropy which were proved previously by Nakamura-Umegaki [16] and Davis [3].

In §§4 and 5, the notions of the measures of information $I(\cdot,\cdot)$ and of divergence $J(\cdot,\cdot)$ between two probability measures, in the sense of Kullback-Leibler [11], are introduced for the von Neumann algebra A and the main theorems relative to the measure of information are extended, from which a minimal property of entropy follows. For the proof of these theorems, the properties of the conditional expectation in non-commutative case and the theorems relative to entropy in §3 are essentially used. In §6, using the method of direct (= tensor) product von Neumann algebra, Shannon-Wiener theorem relative to the additivity of the measure of information over independent events will be extended to the von Neumann algebra A. In §7 ~ §9, the results obtained in the preceding sections will be applied to various cases. In §7, it will be proved that the measure of information between certain restricted operators is not increasing under operating a conditional expectation. In §8, a characterization of sufficient subalgebra is given as a generalization of Kullback-Leibler's theorem.

In §9, it will be discussed the information $I(\cdot:B)$ with respect to a von Neumann subalgebra B, which was introduced by Nakamura-Umegaki[16], and will be studied the relation between $I(\cdot:B)$ and $I(\cdot,\cdot)$. The faithfulness of $I(\cdot:B)$ is not plain, and in order to make up for this point, a new notion of divergence $J(\cdot,B)$ will be introduced. By these mathematical formulation combined with the operator theoretic characterization of measurements due to Nakamura-Umegaki [18], our method of informations in operator algebra may be applicable to the von Neumann theory of measurement in quantum statistics.

In the present paper, for the sake of notational convinience and in order to develope the direct extension of one chapter of probability theory in which the based space is a measure space with total mass 1, we assume that the von Neumann algebra A is of finite class and sigma-finite, having a (fixed) faithful normal trace τ. However, this assumption is not necessarily essential. Indeed, *for the case A being of semi-finite and not necessarily finite, and for τ being a semi-trace, all theorems and propositions in this paper can be shown by a little or simply modified proofs.* For such semi-finite A, the von Neumann subalgebras discussed there (for example, $R(a)$, $R(b)$, B, $\cdots$) should be taken such that the property τ to be semi-trace is preserving on each subalgebra. The abstract of this paper was published in [28].

The author is indebted to Professor M. Nakamura who suggested the possibility of the non-commutative development of the Kullback-Leibler's information and gave valuable discussions to him.

2. Conditional expectation.

Let B be a von Neumann subalgebra of A. An operation ϵ from A onto B ($a \varepsilon A \rightarrow a^{\epsilon} \varepsilon B$) is said to be an *expectation* onto B (cf. Nakamura-Turumaru [13]), if for any operators $a, a_1, a_2 \varepsilon A$ and complex numbers α_1, α_2

(i) (linearity) $(\alpha_1 a_1 + \alpha_2 a_2)^{\epsilon} = \alpha_1 a_1^{\epsilon} + \alpha_2 a_2^{\epsilon}$,

(ii) (Reynolds' identity) $(a_1 a_2^{\epsilon})^{\epsilon} = a_1^{\epsilon} a_2^{\epsilon} = (a_1^{\epsilon} a_2)^{\epsilon}$,

(iii) (non-negative preserving) $a^{\epsilon} \geqq 0$ for $a \geqq 0$,

(iv) (identity preserving) $I^{\epsilon} = I$.

An expectation ϵ is said to be *normal*, if

(v) $a_{\alpha}^{\epsilon} \uparrow a^{\epsilon}$ for $a_{\alpha} \uparrow a$.

Let T_B be the set of all normal states ρ such that

$$\rho(ab) = \rho(ba) \quad \text{for every } a \varepsilon A \text{ and } b \varepsilon B, \tag{2.1}$$

and it is called the *B-tracelet space* (cf. Umegaki [27]).[1] Then for every $\rho \varepsilon S_B$, there exists uniquely a normal expectation $\epsilon = \epsilon_{\rho} = \epsilon(\rho)$ within the support of ρ

1) In the preceding paper [27], the B-tracelet space was denoted by S_B.

such that

$$\rho(a^{\epsilon}b)=\rho(ab) \qquad \text{for every } a \in A \text{ and } b \in B. \tag{2.2}$$

Such an expectation is called by *B-expectation* [27]. Among such expectations, the *conditional expectation* (cf. Umegaki [25], [26])

$$E[a \mid B] = a^{e} \tag{2.3}$$

of an operator $a \in A$ *conditioned by* B is most important where a^e is defined by

$$\tau(a^{e}b)=\tau(ab) \qquad \text{for every } a \in A \text{ and } b \in B, \tag{2.2$'$}$$

that is, $e=\epsilon(\tau)$. The domain of the conditional expectation a^e is uniquely extended to $L^p(A)$, $p \geqq 1$, which is projection of norm one from $L^p(A)$ onto $L^p(B)$, where $L^p(B)$ is the metric hull of B with respect to the L^p-norm.

The concept of the conditional expectation on A is recognized as a non-commutative extension of that in probability space. Indeed, let $(\Omega, \mathfrak{A}, \mu)$ be a probability space and $A(\Omega)$ be the space of all bounded measurable functions. Then $A(\Omega)$ is a multiplication algebra over the Hilbert space $L^2(\Omega, \mu)$ and $A(\Omega)$ can be considered as a commutative von Neumann algebra with a faithful normal trace τ defined by

$$\tau(a(\cdot)) = \int_{\Omega} a(\omega)\, d\mu(\omega). \tag{2.4}$$

Conversely, if A is a commutative von Neumann algebra with a faithful normal trace τ, then there corresponds essentially uniquely a probability space $(\Omega, \mathfrak{A}, \mu)$ such that A is unitarily equivalent to the multiplication algebra $A(\Omega)$, and denoting $a \to a(\cdot)$ the correspondence between A and $A(\Omega)$, then $\tau(a)$ coincides with $\tau(a(\cdot))$ defined by (2.4). For a von Neumann subalgebra B of A there exists uniquely a Borel subfield $\mathfrak{B}$ of $\mathfrak{A}$ such that B is unitarily equivalent to the multiplication algebra $B(\Omega)$ of all bounded measurable functions with respect to $\mathfrak{B}$ over Ω. The conditional expectation $E[\cdot \mid B]$ conditioned by B in the sence of (2.3) coincides with the conditional expectation $E(\cdot \mid B)$ conditioned by $\mathfrak{B}$ in the sense of usual probability theory (cf. Doob [5]):

$$E[a \mid B](\omega) = E(a(\cdot) \mid B)(\omega) \qquad \text{a. e. } \omega \in \Omega \tag{2.5}$$

for every $a \in A$.

In case the algebra A is a function algebra, the expectation has been called the *averaging operator*, and it is one of the the most important operator among the *Reynolds operators* which have been studied by Birkhoff, Dubreil-Jacotin, Kampé de Fériet, Kelley, Rota and others (cf. [1], [9] and [20]). While in case A being finite dimensional, the study of such operators has been carried by Davis [2], who has introduced the pinching operations and proved a uniqueness condition of such operation. On the other hand, from more general point of view, the uniqueness of expectation onto a subalgebra was independently proved for A being of finite class von Neumann algebra by Umegaki [27] which is a *non-commutative extension of the concept of sufficient statistics* in probability space. The expectation onto the subalgebra B is characterized as a projection onto B of norm one, this is due to Tomiyama [24].

The typical examples of the conditional expectations of non-commutative and infinite dimensional case are given by

$$a^{|p} = pap + (1-p)a(1-p),$$
$$a^{|p|q} = (a^{|p})^{|q}, \qquad a^{|p|q|r} = (a^{|p|q})^{|r}, \qquad \cdots$$

in the notation of von Neumann, where $p, q, r, \cdots$ are projections in A commuting each other. If the underlying Hilbert space H is separable, B is maximal abelian subalgebra of A and $p, q, \cdots, r, \cdots$ a sequence of projections in B which generates B, then $a^{|p|q\cdots|r}$ $(a \in A)$ converges strongly to $E[a \mid B]$ which denotes

$$a^{|p|q\cdots} \tag{2.7}$$

and was called von Neumann's operation (cf. Nakamura-Umegaki [14]).

When A is a semi-finite von Neumann algebra, the notion of the conditional expectation conditioned by B is quite analogously introduced for B being the von Neumann subalgebra on which the restriction of τ is also semi-trace (cf. Umegaki [27]). All propositions stated or proved in this section hold for the case A being semi-finite. For such A, a typical example of the conditional expectation is given by a measurement, cf. Nakamura-Umegaki [18], that is, it is the case A being I_∞-factor and whence a measurement of operator corresponding to a physical quantity is just characterized by the conditional expectation of that operator conditioned by the von Neumann subalgebra which is generated by the operator corresponding to the measurement.

Finally, as a preliminary lemma we shall extend the domain of the conditional expectation to certain class of measurable operators.

Let $\mathcal{M}$ be the family of all measurable operators with respect to the von Neumann algebra A. Then $\mathcal{M}$ is self-adjoint algebra (cf. Segal [21]). Let $\mathcal{M}(B)$ be the algebra of all measurable operators affiliated with the von Neumann subalgebra B, and put $e = E[a \mid B]$. Then we prove:

LEMMA 2.1. *Let $\mathcal{A}_B$ be the family of all operators $x \in \mathcal{M}$ for which there exists a sequence of projections $\{p_n\}$ dependent on x and belonging to B such that*

$$p_n \uparrow I \quad and \quad xp_n \in L^1(A) \qquad (n = 1, 2, \cdots). \tag{2.8}$$

Then $\mathcal{A}_B$ is a self-adjoint linear subspace of $\mathcal{M}$, containing $L^1(A)$ and $\mathcal{M}(B)$ as subspaces, and also for every $x \in \mathcal{A}_B$ and $y \in \mathcal{M}(B)$ the operators xy and yx belong to $\mathcal{A}_B$. Furthermore there exists a linear mapping $\bar{e}$ from $\mathcal{A}_B$ onto $\mathcal{M}(B)$ such that, for every x, $x_n \in \mathcal{A}_B$ $(n=1, 2, \cdots)$,

$$(2.2'') \qquad \tau(x^{\bar{e}}b) = \tau(xb) \qquad \textit{for every } b \in B,$$ [2]

$$\text{(i)} \qquad x^{\bar{e}} = x^{e} \qquad \textit{if } x \in L^1(A),$$

$$\text{(ii)} \qquad (x_1^{\bar{e}} x_2)^{\bar{e}} = x_1^{\bar{e}} x_2^{\bar{e}} = (x_1 x_2^{\bar{e}})^{\bar{e}},$$

$$\text{(iii)} \qquad x^{\bar{e}} \geqq 0 \qquad \textit{for } x \geqq 0,$$

$$\text{(iv)} \qquad x_n^{e} \uparrow x^{e} \qquad \textit{for } x_n \uparrow x \quad a.\, e.$$

Proof. It can be proved that, $\mathcal{A}_B$ is a self-adjoint linear subspace of $\mathcal{M}$, containing $L^1(A)$ and $\mathcal{M}(B)$ as subspaces and xy, yx belong to $\mathcal{A}_B$ for each $x \in \mathcal{A}_B$ and $y \in \mathcal{M}(B)$, and since the proof is elementary we omit it. Taking $x, y \in \mathcal{A}_B$ with the sequences of projections $\{p_n\}$ and $\{q_n\}$ satisfying (2.8) respectively, and putting $r_n = p_n \wedge q_n$, then $r_n \in B$, $r_n^{\perp} = p_n^{\perp} \vee q_n^{\perp}$ and

$$\tau(r_n^{\perp}) = \tau(p_n^{\perp} \vee q_n^{\perp}) \leqq \tau(p_n^{\perp}) + \tau(q_n^{\perp}) \to 0,$$ [3]

where $p^{\perp} = 1 - p$. These facts imply that $r_n \uparrow I$,

$$(x+y)r_n = (x\, p_n r_n + y\, q_n r_n) \in L^1(A), \qquad n=1,2,\cdots$$

and $x + y \in \mathcal{A}_B$. For $x \in \mathcal{A}_B$, putting $x_n = x\, p_n$, x_n^e is defined and $x_n^e p_m = x_m^e$ for all $n \geqq m$. By this, we can define an operator y' such as

$$y'\xi = x_n^e \xi, \qquad \text{for every } \xi \in \mathcal{D}(x_n^e) \cap p_n H,\ n=1,2,\cdots,$$

then y' is well-defined and a linear operator on the dense domain $\cup_{n=1}^{\infty} (\mathcal{D}(x_n^e) \cap p_n H)$, and it has a unique extension $y \in \mathcal{M}(B)$ satisfying $y p_n = x_n^e$ $(n=1,2,\cdots)$. Put $x^{\bar{e}} = y$. Then by the construction of $x^{\bar{e}}$

$$(2.2''') \qquad \tau(x\, p_n b) = \tau(x^{\bar{e}} p_n b), \qquad \text{for every } b \in B,\ n=1,2,\cdots.$$

By (2.2'''), it is obvious that (2.2'') holds for every $x \in \mathcal{A}_B$ and $x^{\bar{e}}$ is uniquely determined by x and independent of the choice of the sequence of projections $\{p_n\}$ within B. It is also obvious that, the operation $x \to x^{\bar{e}}$ from $\mathcal{A}_B$ onto $\mathcal{M}(B)$ is well-defined, linear and satisfies (i). The property (ii) follows from that: Since $x_1^{\bar{e}} x_2 \in \mathcal{A}_B$ and $x_1 x_2^{\bar{e}} \in \mathcal{A}_B$, there exists a sequence of projections $\{p_n\} \subset B$ simultaneously satisfying (2.8) for x_1, x_2, $x_1^{\bar{e}} x_2$ and $x_1 x_1^{\bar{e}}$ and also satisfying

$$(x_i p_n)^e = x_i^{\bar{e}} p_n, \qquad i=1,2 \text{ and } n=1,2,\cdots,$$

whence for every $b \in B$

$$\tau(x_1^{\bar{e}} x_2 p_n b) = \tau(x_1^{\bar{e}} x_2^{\bar{e}} p_n b) = \tau(x_1 x_2^{\bar{e}} p_n b), \qquad n=1,2,\cdots$$

and this implies (ii). (iii) is obvious by (2.2'''). Finally (iv) is proved such as:

---

2) The equality (2.2'') is meant by that the left-side exists if and only if the right-side exists and they are equal each other.

3) This is a method of Segal [21] which is applicable to the case that A is of semi-finite and τ is semi-trace, by taking the dimension function $d(\cdot)$ of him.

By (iii), $x_n^{\bar{e}} \leqq x_{n+1}^{\bar{e}} \leqq x^{\bar{e}}$ $(n=1,2,\cdots)$, and restricting $x^{\bar{e}}$ onto a strongly dense domain on which $x^{\bar{e}}$ is bounded, then we can find an operator $y \in \mathcal{M}(B)$ such that

$$x_n^e \uparrow y \leqq x^{\bar{e}} \qquad \text{(a. e.)}.$$

Taking a sequence of projection $\{p_k\} \subset B$ satisfying (2.8) for x, x_n and y, then

$$\begin{aligned} 0 \leqq \tau(p_k(x^{\bar{e}} - y)p_k) &= \lim_{n\to\infty} \tau(p_k(x^{\bar{e}} - x_n^{\bar{e}})p_k) \\ &= \lim_{n\to\infty} \tau(p_k(x - x_n)p_k) = 0 \end{aligned}$$

and $p_k(x^{\bar{e}} - y)p_k = 0$ $(k=1,2,\cdots)$. Since $p_k \uparrow I$, $x^{\bar{e}} = y$. Q. E. D.

The following Corollary immediately follows from Lemma 2.1:

COROLLARY 2.1. *For every $a \in L^1(A)$ and $y \in \mathcal{M}(B)$, the equality $(ay)^{\bar{e}} = a^e y$ holds.*

When $B = \{p, q, \cdots, r\}' \cap A$ for a finite number of projections $\{p, q, \cdots, r\} \subset A$ which are commuting each others, the space $\mathcal{A}_B$ coincides with the space of all measurable operators $\mathcal{M}$. However, in general, the equality $\mathcal{A}_B = \mathcal{M}$ not necessarily holds (for example, when B consists of only the scalar multiples of identity I, then $\mathcal{A}_B = L^1(A)$).

In order to simplify the notation, we denote $x^{\bar{e}} = E[x \mid B] = x^e$ and call it the conditional expectation of x conditioned by B as for the case x belonging to A or $L^1(A)$.

3. Entropy.

Let a be a self-adjoint operator with the spectral resolution

$$a = \int_I \lambda \, de_\lambda, \tag{3.1}$$

where I is a finite or infinite interval, and let $f(\lambda)$ be a continuous function over I, then the operator-function $f(a)$, for operators a with spectra $\subset I$, is defined by

$$f(a) = \int_I f(\lambda) de_\lambda. \tag{3.2}$$

Now, we discuss two operator-functions $\log a$ and $a \log a$. Denote

$$\log^- \lambda = \begin{cases} \log \lambda & (0 < \lambda \leqq 1), \\ 0 & (\lambda > 1) \end{cases} \quad \text{and} \quad \log^+ \lambda = \begin{cases} 0 & (\lambda \leqq 1), \\ \log \lambda & (\lambda > 1). \end{cases} \tag{3.3}$$

If an operator

$$a = \int_0^\infty \lambda \, de_\lambda$$

has not 0 as its spectrum, then, it will be denoted by $a \gg 0$. For $a \gg 0$, the operators

$$
(3.4) \qquad \begin{cases} \log a = \int_0^\infty \log \lambda \, de_\lambda, \\ \log^- a = \int_0^\infty \log^- \lambda \, de_\lambda, \\ \log^+ a = \int_0^\infty \log^+ \lambda \, de_\lambda \end{cases}
$$

are self-adjoint and satisfy

$$
\log a = \log^- a + \log^+ a,
$$

$$
\log^- a \leqq 0, \quad \log^+ a \geqq 0 \quad \text{and} \quad (\log^- a) \cdot (\log^+ a) = 0 .
$$

Furthermore denote

$$
(3.5) \qquad \begin{cases} a' = \int_0^1 \lambda \, de_\lambda, \\ a'' = \int_1^\infty \lambda \, de_\lambda, \end{cases}
$$

and

$$
(3.6) \qquad \begin{cases} a'_n = \int_{1/n}^1 \lambda \, de_\lambda + e_{1/n-0}, \\ a''_n = \int_1^n \lambda \, de_\lambda + I - e_n, \end{cases}
$$

put $\mathit{log}^- a'_n = \log^- (a'_n + a'')$ and $\mathit{log}^+ a''_n = \log^+ (a''_n + a')$, then

$$
(3.7) \qquad \begin{cases} 0 \geqq \mathit{log}^- a'_n \downarrow \log^- a, \\ 0 \leqq \mathit{log}^+ a''_n \uparrow \log^+ a. \end{cases}
$$

When $a \geqq 0$ but not necessarily $a \gg 0$, the operators $\log a$, $\log^- a$ and $\log^+ a$ are defined over the subspace $s(a)H$ and these are denoted by $s(a) \log a$, $s(a) \log^- a$ and $s(a) \log^+ a$, and can be regarded as self-adjoint operators over H, where $s(a)$ is the support-projection of a.

Let $h(\lambda)$ be a function defined by

$$
(3.8) \qquad h(\lambda) = \begin{cases} -\lambda \log \lambda & \text{for } \lambda > 0, \\ 0 & \text{for } \lambda = 0, \end{cases}
$$

then $h(\lambda)$ is continuous for $\lambda \geqq 0$. Whence the operator function $h(a)$, denote $-a \log a$, is defined for $a \geqq 0$ and it will be called *operator-entropy of* a. *The operator function* $a \log a$ $(= -h(a))$ *is expressed by*

$$
a \log a = a \cdot s(a) \log a \qquad \textit{for } a \geqq 0
$$

and $a \log a$ for $a \gg 0$, here $\cdot$ is the notation of product of operators. This operator function satisfies the following fundamental theorems:

THEOREM A. *The operator-entropy* $h(a)$, $0 \leqq a \varepsilon A$, *is an operator-concave* *function, that is,*

$$h(\alpha a+\beta b)\geqq \alpha h(a)+\beta h(b) \tag{3.9}$$

for every $a, b \varepsilon A,\ a, b \geqq 0$ *and* $\alpha, \beta \geqq 0$ *with* $\alpha+\beta=1$.

THEOREM B. *Let B be a fixed von Neumann subalgebra, then the operator function $h(a)$ satisfies*

$$h(E[a \mid B])\geqq E[h(a) \mid B] \tag{3.10}$$

for every $a \varepsilon A,\ a\geqq 0$. *(This may be simply written by*

$$a^e \log a^e \leqq (a \log a)^e \tag{3.11}$$

where $e=E[\cdot \mid B]$.)

These theorems, for the case A being semi-finite, are proved by Nakamura-Umegaki [16] and also independently by Davis [3], which are generalizations of theorems of Segal [22] on the numerical entropy. In order to apply to another occasion, Theorem B, for B being commutative, is extended for operators not necessarily bounded, and in fact we get Theorem B′ which will become necessary to discuss the parts below in this paper:

THEOREM B′. *Let B be commutative von Neumann subalgebra and $\mathcal{A}_B$ be the space of measurable operators determined by B (cf. Lemma 2.1). Then the inequality (3.10) holds for every $a\geqq 0$, $a \log a \varepsilon \mathcal{A}_B$, hence it holds also for every $a \geqq 0$, $a \log a \varepsilon L^1(A)$.*

Proof. For the given operator a with

$$0\leqq a=\int_0^\infty \lambda\, de_\lambda \,\varepsilon\, \mathcal{A}_B,$$

putting

$$a_n=\int_0^n \lambda\, de_\lambda,$$

then $a_n \varepsilon R(a)$ and

$$0\leqq a_n \uparrow a, \qquad a_n \log a_n \uparrow a \log a. \tag{3.12}$$

By this and Lemma 2.1,

$$0\leqq a_n^e \uparrow a^e, \qquad (a_n \log a_n)^e \uparrow (a \log a)^e. \tag{3.13}$$

Since B is commutative,

$$a_n^e \log a_n^e \rightarrow a^e \log a^e \qquad \text{a. e.} \tag{3.14}$$

By Theorem B and boundedness of a_n,

$$a_n^e \log a_n^e \leqq (a_n \log a_n)^e, \qquad n=1,2,\cdots.$$

Therefore by (3.13) and (3.14), the required inequalities (3.11), i. e. (3.10) are obtained.

Define the *entropy* of an operator $a \varepsilon L^1(A)$, $a \geqq 0$, by

$$H(a) = -\tau(a \log a)(=\tau(h(a))), \tag{3.15}$$

cf. Nakamura-Umegaki [16] and Davis [3], and the *entropy* of a normal state ρ of A by

$$H(\rho) = -\tau\left(\frac{d\rho}{d\tau}\log\frac{d\rho}{d\tau}\right)\left(=\tau\left(h\left(\frac{d\rho}{d\tau}\right)\right)=\dot{H}\left(\frac{d\rho}{d\tau}\right)\right). \tag{3.15'}$$

It is obvious that the entropies $H(a)$ and $H(\rho)$ are well-defined as finite values or $-\infty$. From the Theorem A, it follows immediately that the function $H(a)$, $0 \leqq a \varepsilon L^1(A)$, is *concave*, that is,

$$H(\alpha a + \beta b) \geqq \alpha H(a) + \beta H(b) \tag{3.16}$$

for every $a, b \varepsilon L^1(A)$, $a, b \geqq 0$ and scalars $\alpha, \beta \geqq 0$ with $\alpha+\beta=1$. Using this fact, it holds also that $H(\rho)$ is concave over the set of all normal states S, that is,

$$H(\alpha\sigma + \beta\rho) \geqq \alpha H(\sigma) + \beta H(\rho) \tag{3.16'}$$

for every $\sigma, \rho \varepsilon S$ and the scalars α, β.

REMARK 3.1.[4] (i) *If the operator a is bounded, then the operator-entropy $h(a)$ of a is necessarily bounded.* (ii) *If $a \varepsilon L^p(A)$ for some $p>1$, then $h(a)$ belongs to $L^1(A)$.* Indeed, (i) is obvious, and (ii) follows from that, there exists a projection $q \varepsilon R(a)$ such that

$$0 \leqq qa \log a \leqq qa^p \quad \text{and} \quad (I-q)(a \log a) \leqq \alpha(I-q)$$

for some scalor $\alpha \geqq 0$.

Denote $\mathcal{E} = \mathcal{E}(A)$ the set of all operators $a \varepsilon L^1(A)$, $a \geqq 0$, with the finite entropy $|H(a)| < \infty$ and with $\tau(a) = 1$.

4. Information between operators.

Now we define a measure of Kullback-Leibler's information in the von Neumann algebra A.

DEFINITION 1. Let a and b be a pair of operators in $L^1(A)$, $a, b \geqq 0$, with

4) For τ being semi-trace, the statement (ii) in this Remark is not necessarily satisfied, because $L^p(A)$ $(p>1)$ is not contained in $L^1(A)$.

$a \prec b$ and $\tau(a) = \tau(b) = 1$. Then

$$(4.1) \qquad I(a, b) = \tau(a \log a - a \log b)$$

is said to be the *information*[5] between a and b. Let σ and ρ be a pair of normal states with $\sigma \prec \rho$, take the Radon-Nikodym derivatives $d\sigma/d\tau$ and $d\rho/d\tau$, and put

$$(4.1') \qquad I(\sigma, \rho) = I\left(\frac{d\sigma}{d\tau}, \frac{d\rho}{d\tau}\right),$$

then it is said to be the *information* between σ and ρ. In (4.1), the operator $a \log b$ is defined by (cf. § 3)

$$a \log b = as(b) \log b \,\varepsilon\, \mathcal{M}$$

and in general, the operator $a \log a - a \log b$ will be simply denoted by

$$(4.2) \qquad a(\log a - \log b).$$

As preparation, we shall prove the following two propositions.

PROPOSITION 4.1. *For any pair of operators a, b in Definition 1, the information $I(a, b)$ is uniquely determined as finite or $+\infty$, if they satisfy*

(i) $$ab = ba,$$

or

(ii) *the entropy $H(a)$ is finite and b is bounded.*

The proposition, for the pair a, b being $a \sim b$, has been proved in the previous paper [28]. Here, for the sake of completeness, we recall its proof.

Proof. The case (i): Put $p = I - s(b)$, $s(b)$ being the support-projection of b. Then the operators $(b+p)^{-1}$ and $a(b+p)^{-1}$ are defined as measurable and $\geqq 0$, and hence

$$a \log a - a \log b = a(b+p)^{-1}(\log a(b+p)^{-1})\, b$$

is a measurable operator. Furthermore

$$a(b+p)^{-1} \log (a(b+p)^{-1}) \geqq -1$$

and $\tau(b) = 1$ imply that

$$I(a, b) = \tau(a(b+p)^{-1}[\log (a(b+p)^{-1})b])$$

is uniquely determined as finite or $+\infty$.

The case (ii): This condition implies that $a \log a \,\varepsilon\, L^1(A)$ and $\tau(a \log b) \leqq \| b \|_\infty$ ($\| \cdot \|_\infty$ being operator bound) and hence

---

5) When A is of semi-finite and τ is a semi-trace and not trace, $I(\cdot, \cdot)$ can be defined for $0 \leqq a \,\varepsilon\, L^1(A)$ and $0 \leqq b \,\varepsilon\, A$ with $a \prec b$, where b does not necessarily belong to $L^1(A)$. Hence, at any rate, the entropy $H(a)$ of $0 \leqq a \,\varepsilon\, L^1(A)$ is given by $H(a) = -I(a, I)$.

$$I(a, b) = \tau(a \log a) - \tau(a \log b)$$

is uniquely determined as finite or $+\infty$.

It is obvious that the information $I(a, b)$ for the pair of operators a, b, depends on the choice of the normal trace τ. However, for the information between normal states, the following holds:

PROPOSITION 4.2. *The information $I(\sigma, \rho)$ between σ and ρ, $\sigma \prec \rho$, is independent of the choice of the trace τ.*

Proof.[6] Let τ' be another faithful normal trace. Then the Radon-Nikodym derivatives satisfy

$$\frac{d\tau'}{d\tau}\frac{d\sigma}{d\tau'} = \frac{d\sigma}{d\tau'}\cdot\frac{d\tau'}{d\tau} = \frac{d\sigma}{d\tau}$$

because $d\tau'/d\tau$ is affiliated with the center $A \cap A'$ and similarly

$$\frac{d\rho}{d\tau'}\frac{d\tau'}{d\tau} = \frac{d\rho}{d\tau}.$$

Therefore

$$\begin{aligned}
&\tau'\left(\frac{d\sigma}{d\tau'}\left[\log\frac{d\sigma}{d\tau'} - \log\frac{d\rho}{d\tau'}\right]\right)\\
&= \tau\left(\frac{d\tau'}{d\tau}\frac{d\sigma}{d\tau'}\left[\log\frac{d\sigma}{d\tau'} - \log\frac{d\rho}{d\tau'}\right]\right)\\
&= \tau\left(\frac{d\sigma}{d\tau}\left[\log\frac{d\sigma}{d\tau} - \log\frac{d\tau'}{d\tau} - \log\frac{d\rho}{d\tau} + \log\frac{d\tau'}{d\tau}\right]\right)\\
&= \tau\left(\frac{d\sigma}{d\tau}\left[\log\frac{d\sigma}{d\tau} - \log\frac{d\rho}{d\tau}\right]\right).
\end{aligned}$$

This implies that, in these equalities, the left-side exists if and only if the right-side exists and they are equal, that is, $I(\sigma, \rho)$ is independent of the choice of τ.

The following theorem is fundamental for Kullback-Leibler's theory of information:

THEOREM 1. *For any pair of operators $a, b \varepsilon \mathcal{E}$, i.e. non-negative, self-adjoint operators with finite entropies $H(a)$, $H(b)$ satisfying $\tau(a) = \tau(b) = 1$ and $a \prec b$, the information $I(a, b)$ is uniquely determined. Unless $I(a, b)$ is infinite, it is a non-negative real number.*

Proof. At first, assume the operator b being bounded. Whence, by Pro-

6) When A is semi-finite and τ, τ' are semi-trace, the Radon-Nikodym derivative $d\tau'/d\tau$ exists by Segal [21].

position 4.1, $I(a, b)$ is uniquely determined as finite or $+\infty$. Let $R(b)$ be the von Neumann algebra generated by b, and denote $e = E[\cdot \mid R(b)]$ the conditional expectation conditioned by $R(b)$. Then by Theorem B′,

$$a^e \log a^e \leqq (a \log a)^e. \tag{4.4}$$

Since $a \prec b$ implies $a^e \prec b$ (this fact holds for more general case, cf. Lemma 7.1 below), the operator $a^e \log b$ is defined by $a^e s(b) \log b$ and, by Corollary 2.1,

$$(a \log b)^e = (as(b) \log b)^e = a^e s(b) \log b = a^e \log b \qquad (\text{by } a^e \prec b), \tag{4.5}$$

where $s(b)$ is the support-projection $s(b)$ of b. By Lemma 2.1, (4.4) and (4.5) imply the inequality

$$a^e \log a^e - a^e \log b \leqq (a \log a)^e - (a \log b)^e = (a \log a - a \log b)^e. \tag{4.6}$$

Therefore, in order to show $I(a, b) \geqq 0$ (for $b \in A$), it is sufficient to prove

$$\tau(a^e \log a^e - a^e \log b) \geqq 0. \tag{4.7}$$

Since $\tau(a^e) = \tau(a) = 1$ and $a^e \geqq 0$, the left-side of (4.7) equals to $I(a^e, b)$. Furthermore, since $a^e b = b a^e$, it is sufficient to see that $I(a, b) \geqq 0$ for the case a and b being permutable. This is the case of Kullback-Leibler [11]. For the sake of completeness, we shall describe it in some details.

Using the second order mean value theorem, the function $-h(\lambda)$ for $\lambda \geqq 0$ defined by (3.8) is expressed by

$$-h(\lambda) = (\lambda - 1) + \frac{1}{2}(\lambda - 1)^2 \theta(\lambda), \tag{4.8}$$

where $\theta(\lambda)$ is a bounded continuous and positive valued function of $\lambda \geqq 0$.

Since $ab = ba$,

$$\begin{aligned} &a \log a - a \log b \\ &= a(b + s'(b))^{-1}[\log (a(b + s'(b))^{-1})]b \\ &= -h(a(b + s'(b))^{-1})b \\ &= [a(b + s'(b))^{-1} - 1]b + \frac{1}{2}[a(b + s'(b))^{-1} - 1]^2 \theta(a(b + s'(b))^{-1})b \end{aligned}$$

by (4.8),

$$= (a - b) + \frac{1}{2}(a - b)^2 \theta(a(b+s'(b))^{-1})(b+s'(b))^{-1},$$

where $s'(b) = I - s(b)$ $(= s(b)^{\perp})$. Therefore, by $\tau(a) = \tau(b) = 1$, we obtain

$$I(a, b) = \frac{1}{2}\tau([(a - b)^2 \theta(a(b + s'(b))^{-1})(b + s'(b))^{-1}]) \geqq 0 \quad (\text{admitting } +\infty). \tag{4.9}$$

Thus the assertion of Theorem 1 are proved for the operators a and bounded b.

In general case, suppose $I(a, b)$ is not uniquely determined for some pair $a, b \varepsilon \mathcal{E}$ with $a \prec b$. Then, taking a projection p in $R(b)$ such as

$$0 \leqq b < 1 \text{ over } pH \quad \text{and} \quad 1 \leqq b \text{ over } (1-p)H,$$

there exist two sequences y'_n and $y''_n \varepsilon R(b)$ with $\tau(y'_n) = \tau(y''_n) = 1$ and $y'_n, y''_n \geqq 0$ such that each y'_n and y''_n have bounded inverses and satisfy

$$py'_n \downarrow pb, \qquad py''_n \downarrow pb$$

and

$$(1-p)y'_n \uparrow (1-p)b, \qquad (1-p)y''_n \uparrow (1-p)b,$$

and furthermore

$$I(a, y'_n) \rightarrow \beta', \qquad I(a, y''_n) \rightarrow \beta'', \qquad \beta' \neq \beta''.$$

Of course, $I(a, y'_n)$ and $I(a, y''_n)$ are finite. Since

$$a^e (= E[a \mid R(b)]) \prec b \quad \text{and} \quad a^e b = b a^e,$$

by Proposition 4.1 $I(a^e, b)$ is uniquely determined and therefore

$$\lim_{n\to\infty} I(a^e, y'_n) = \lim_{n\to\infty} I(a^e, y''_n) = I(a^e, b) (= \beta, \text{ say}),$$

hence

$$\begin{aligned} \beta' &= \lim_{n\to\infty} (\tau(a \log a) - \tau(a \log y'_n)) \\ &= \lim_{n\to\infty} (\tau(a \log a) - \tau(a^e \log y'_n)) \\ &= \tau(a \log a) - \tau(a^e \log a^e) + \tau(a^e \log a^e) - \lim_{n\to\infty} \tau(a^e \log y'_n) \\ &= I(a, a^e) + \beta, \end{aligned}$$

and similarly

$$\beta'' = I(a, a^e) + \beta,$$

and hence $\beta' = \beta''$. This is a contradiction. Therefore $I(a, b)$ is uniquely determined for each pair $a, b \varepsilon \mathcal{E}$ with $a \prec b$. Moreover this fact and (4.9) imply that

$$0 \leqq I(a, y'_n) \rightarrow I(a, b) \qquad (\text{admitting } +\infty)$$

and $I(a, b) \geqq 0$ is satisfied. This completes the proof. Q.E.D.

By the proof of Theorem 1, we have

COROLLARY 4.1. *For any pair a, b satisfying the hypothesis of Theorem 1, the informations between a and b, and a^e and b are defined and satisfy*

$$I(a, b) \geqq I(a^e, b) \geqq 0,$$

where $a^e = E[a \mid R(b)]$.

From Theorem 1, it follows immediately

THEOREM 1′. *For any pair* σ, ρ *of normal states with* $\sigma \prec \rho$ *and with finite entropy,* $I(\sigma, \rho)$ *is uniquely determined. Unless* $I(\sigma, \rho)$ *is infinite, it is a non-negative real number.*

Moreover we get the following as a further corollary:

COROLLARY 4.2. *For any operator* $a \in \mathcal{E}$, *and for any normal state* σ *of* A *with finite entropy, the following equalities with respect to entropy hold:*

$$H(a) = \mathrm{Min}\, \{ -\tau(a \log b) \mid a \prec b, \ \| b \|_1 = 1 \} \tag{4.10}$$

and

$$H(\sigma) = \mathrm{Min}\, \left\{ -\tau\left(\frac{d\sigma}{d\tau} \log \frac{d\rho}{d\tau} \right) \middle| \sigma \prec \rho, \ \rho \in S \right\}, \tag{4.11}$$

where S *is the set of all normal states.*

5. Divergence between operators.

We introduce into the von Neumann algebra A the concept of the divergence in the Kullback-Leibler's information.

DEFINITION 2. For every pair of operators $a, b \in \mathcal{E}$ with $a \sim b$, let us define the *divergence* between a and b by

$$J(a, b) = I(a, b) + I(b, a). \tag{5.1}$$

Similarily, for every pair of normal states σ, ρ with $\sigma \sim \rho$, let us define the *divergence* between σ and ρ by

$$J(\sigma, \rho) = I(\sigma, \rho) + I(\rho, \sigma). \tag{5.1′}$$

Then, by Theorem 1, it is obvious that $J(a, b)$ and $J(\sigma, \rho)$ are well-defined over the each pair of the operators or the normal states with finite entropy and satisfy

$$J(a, b) = J(b, a) \geqq 0, \tag{5.2}$$

$$J(\sigma, \rho) = J(\rho, \sigma) \geqq 0 \tag{5.2′}$$

Whence we prove the following:

THEOREM 2. *For every pair $a, b \varepsilon \mathcal{E}$, the following conditions are equivalent each others*

(i) $$J(a, b) = 0,$$

(ii) $$ab = ba \quad and \quad I(a, b) = 0,$$

and

(iii) $$a = b.$$

Proof. Since it is obvious that (iii) implies (i) and (ii), we shall prove here that, (ii) implies (iii), and (i) implies (iii).

(ii) *implies* (iii): This is essentially the case of Kullback-Leibler (cf. Lemma 1 of [11]). Now prove this. (ii) implies

$$\tau(a \log b) = -H(a) \qquad (|H(a)| < \infty),$$

and by $ab = ba$, it is obvious $a \prec b$. Let M be a von Neumann algebra generated by $R(a)$ and $R(b)$. Then by the self-adjointness of a and b, and by $ab = ba$, M is commutative. Hence there exist two sequences of (boundedly invertible) operators $\{a_n\}$ and $\{b_n\}$ in M, $a_n, b_n \gg 0$, such that

$$\tau(a_n) = \tau(b_n) = 1, \quad a_n \to a \text{ and } b_n \to b$$

in the both senses of a. e. and L^1-mean, and further

$$I(a_n, b_n) \to I(a, b) = 0.$$

Since $a_n b_n = b_n a_n$, by (4.9)

$$I(a_n, b_n) = \tau((a_n - b_n)^2 \theta(a_n b_n^{-1}) b_n^{-1}) \to 0.$$

Since

$$(a_n - b_n)^2 \theta(a_n b_n^{-1}) b_n^{-1} \geqq 0,$$

we can take a suitable subsequence $\{n_k\}$ of $\{n\}$, such as

$$(a_{n_k} - b_{n_k})^2 \theta(a_{n_k} b_{n_k}^{-1}) b_{n_k}^{-1} \to 0 \qquad \text{a. e. as } k \to \infty,$$

while this converges to

$$(a - b)^2 \theta(a(b + s'(b))^{-1}) \cdot (b + s'(b))^{-1} s(b) \qquad \text{a. e.,}$$

where $s'(b) = 1 - s(b)$. Since $s(a) \leqq s(b)$ and

$$\theta(a(b + s'(b))^{-1}) \cdot (b + s'(b))^{-1} \gg 0 \qquad \text{over } s(b)H,$$

we obtain $a = b$ (a. e.).

(i) *implies* (iii): $J(a, b) = 0$ implies

$$I(a, b) = I(b, a) = 0.$$

Hence $a \log b \,\varepsilon\, L^1(A)$, and by Corollary 2.1

$$E[a \mid R(b)] \log b = E[a \log b \mid R(b)] \,\varepsilon\, L^1(A)$$

and

$$a \prec E[a \mid R(b)] \prec b$$

hold (in this relation, $a \prec E[a \mid R(b)]$ is a special case of Lemma 7.1, below). Similarily $b \prec a$ is obtained and hence $a \sim b$ holds. Therefore by Corollary 4.1,

$$0 \leqq I(E[a \mid b], b) \leqq I(a, b) = 0$$

and

$$0 \leqq I(E[b \mid a], a) \leqq I(b, a) = 0,$$

where denote $E[\cdot \mid a] = E[\cdot \mid R(a)]$ and $E[\cdot \mid b] = E[\cdot \mid R(b)]$. Since

$$E[a \mid b]b = bE[a \mid b] \quad \text{and} \quad E[a \mid b] \prec b,$$

the pair $E[a \mid b]$, b satisfies (ii) and hence (iii), that is, $E[a \mid b] = b$ holds and similarily $E[b \mid a] = a$. Therefore

$$E[E[a \mid b] \mid a] = E[b \mid a] = a \quad \text{and} \quad E[E[b \mid a] \mid b] = b \tag{5.3}$$

hold. Consequently, for every $x \,\varepsilon\, A$, taking a sequence $\{x_n\} \subset A$ such as

$$\begin{cases} x_1 = E[x \mid a], \quad x_2 = E[x_1 \mid b], \quad \cdots, \\ x_{2n} = E[x_{2n-1} \mid b], \quad x_{2n+1} = E[x_{2n} \mid a], \quad \cdots, \end{cases} \tag{5.4}$$

then $\{x_n\}$ converges strongly (and hence L^2 and L^1-means) to an operator $x' \,\varepsilon\, A$. The operation $x \to x'$ is the conditional expectation conditioned by $R(a) \cap R(b)$, that is, $x' = E[x \mid R(a) \cap R(b)]$.

For the given operator $a \,\varepsilon\, L^1(A)$ and for any fixed $\varepsilon > 0$, taking an operator $x \,\varepsilon\, A$ such as $\| a - x \|_1 < \varepsilon/2$, then

$$a - x' = \lim_{n \to \infty} (a - x)_{2n-1}, \qquad b - x' = \lim_{n \to \infty} (b - x)_{2n}, \tag{5.5}$$

where $(\cdot)_{2n}$ and $(\cdot)_{2n-1}$ are the notations given in (5.4) and the limits in (5.5) are L^1-means. Therefore we obtain that

$$\begin{aligned} \| a - b \|_1 &\leqq \| a - x' \|_1 + \| b - x' \|_1 \\ &= \lim_{n \to \infty} \| (a - x)_{2n-1} \|_1 + \lim_{n \to \infty} \| (b - x)_{2n} \|_1 \\ &\leqq \| a - x \|_1 + \| b - x_2 \|_1 \\ &= 2 \| a - x \|_1 < \varepsilon \end{aligned}$$

and $a = b$. Q.E.D.

Applying Theorem 2 for normal states σ, ρ of A, we have immediately

THEOREM 2′. *For any pair of normal states σ, ρ of A with finite entropy,*

$J(\sigma, \rho) = 0$ if and only if $\sigma = \rho$.

REMARK 5.1. If σ and ρ are probability measures, $\sigma \prec \rho$, over a measurable space $(\Omega, \mathfrak{A})$, then the information $I(\sigma, \rho)$ is defined by

$$I(\sigma, \rho) = \int \left(\frac{d\sigma}{d\tau} \log \frac{d\sigma}{d\tau} - \frac{d\sigma}{d\tau} \log \frac{d\rho}{d\tau} \right) d\tau, \tag{5.6}$$

$\tau = (\sigma + \rho)/2$, and the divergence J by

$$J(\sigma, \rho) = I(\sigma, \rho) + I(\rho, \sigma).$$

Then it is proved that

$$I(\sigma, \rho) = \int \left(\frac{d\sigma}{d\mu} \log \frac{d\sigma}{d\mu} - \frac{d\sigma}{d\mu} \log \frac{d\rho}{d\mu} \right) d\mu$$

for every probability measure μ satisfying $\sigma \prec \mu$ and $\rho \prec \mu$. Indeed, let $\mathfrak{s}$ be the support set of the measure τ $(= (\sigma + \rho)/2)$ and put $\mu'(\mathfrak{a}) = \mu(\mathfrak{a} \cap \mathfrak{s})/\mu(\mathfrak{s})$. Then μ' is a probability measure over $(\Omega, \mathfrak{A})$ with $\tau \sim \mu'$,

$$\int \left(\frac{d\sigma}{d\mu} \log \frac{d\sigma}{d\mu} - \frac{d\sigma}{d\mu} \log \frac{d\rho}{d\mu} \right) d\mu = \int \left(\frac{d\sigma}{d\mu'} \log \frac{d\sigma}{d\mu'} - \frac{d\sigma}{d\mu'} \log \frac{d\rho}{d\mu'} \right) d\mu'$$

and by Proposition 4.2, this equals to the right side of (5.6). Under theses notations, *the conditions*

$$\text{(i)}\ J(\sigma, \rho) = 0, \qquad \text{(ii)}\ I(\sigma, \rho) = 0 \qquad \text{and} \qquad \text{(iii)}\ \sigma = \rho$$

are equivalent each others. This is the Kullback-Leibler's Theorem, cf. [11], and reduces to a special case of Theorems 2 and 2', because putting $a = d\sigma/d\tau$ and $b = d\rho/d\tau$ $(\tau = (\sigma + \rho)/2)$, then $ab = ba$ holds.

6. Information of direct product operators.

Let A_i be a von Neumann algebras of finite class with faithful normal traces τ_i $(i = 1, 2)$, respectively. Let $A = A_1 \otimes A_2$ be the direct product von Neumann algebra (cf. Dixmier [4], Misonou [12]) and let $\tau = \tau_1 \otimes \tau_2$ be the direct product trace. Then A is of finite class and τ is a faithful normal trace. Under these notations we prove the following:

THEOREM 3. *For every pair of operators $a_i, b_i \in \mathcal{E}_i$ $(= \mathcal{E}(A_i))$ if $a_i \prec b_i$ $(i = 1, 2)$, then $a_1 \otimes a_2 \prec b_1 \otimes b_2$ which are belonging to $\mathcal{E}$ $(= \mathcal{E}(A))$, and the following equality holds:*

$$I(a_1 \otimes a_2,\ b_1 \otimes b_2) = I(a_1, b_1) + I(a_2, b_2) \tag{6.1}$$

and if $a_i \sim b_i$ $(i = 1, 2)$, then $a_1 \otimes a_2 \sim b_1 \otimes b_2$ and the following holds:

$$J(a_1 \otimes a_2,\ b_1 \otimes b_2) = J(a_1, b_1) + J(a_2, b_2). \tag{6.2}$$

Proof. If $a_i \prec b_i$ $(i=1,2)$, then $a_1 \otimes a_2 \prec b_1 \otimes b_2$ is obvious and the both sides belong to $\mathcal{E}$, and hence the left-sides of the equalities (6.1) and (6.2) are well-defined under the respective conditions. Therefore, under the contract (4.2) in Definition 1, the following computation can be carried out:

$$
\begin{aligned}
& I(a_1 \otimes a_2,\ b_1 \otimes b_2) \\
&= \tau((a_1 \otimes a_2)[\log (a_1 \otimes a_2) - \log (b_1 \otimes b_2)]) \\
&= \tau((a_1 \otimes a_2)[\{\log (a_1 \otimes 1) + \log (1 \otimes a_2)\} - \{(\log (b_1 \otimes 1) + \log (1 \otimes b_2)\}]) \\
&= \tau((a_1 \otimes a_2)[\{\log (a_1 \otimes 1) - \log (b_1 \otimes 1)\} + \{\log (1 \otimes a_2) - \log (1 \otimes b_2)\}]) \\
&= \tau(\{a_1[\log a_1 - \log b_1]\} \otimes a_2) + \tau(a_1 \otimes \{a_2[\log a_2 - \log b_2]\}) \\
&= \tau(a_1[\log a_1 - \log b_1])\tau(a_2) + \tau(a_1)\tau(a_2[\log a_2 - \log b_2]) \\
&= I(a_1, b_1) + I(a_2, b_2)
\end{aligned}
$$

and (6.1) is obtained, similarily (6.2) holds. Q.E.D.

This theorem is an operator generalization of Kullback-Leibler's form of Shannon-Wiener Theorem, which is fundamental for information theory and is described as following: *If f_1 and f_2, and also g_1 and g_2, respectively, are independent random events defined over a probability space $(\Omega, \mathfrak{A}, \mu)$, then*

$$I(f_1 f_2, g_1 g_2) = I(f_1, g_1) + I(f_2, g_2). \tag{6.3}$$

Indeed, the events $f_1 f_2$ and $g_1 g_2$ are considered as functions in $L^1(\Omega \times \Omega) = L^1(A(\Omega) \otimes A(\Omega))$, and represented by $(f_1 f_2)(\cdot) = (f_1 \otimes f_2)(\cdot)$ and $(g_1 g_2)(\cdot) = (g_1 \otimes g_2)(\cdot)$, and therefore the equality (6.3) reduces to (6.1), where $A(\Omega)$ is the multiplication algebra (cf. §2).

In general, $H(a) = -I(a, I)$ for every $a \varepsilon \mathcal{E}$ holds, consequently as a corollary of Theorem 3.

COROLLARY 3.1. *For every $a_i \varepsilon L^1(A_i)$, $a_i \geqq 0$, $i=1,2$,*

$$H(a_1 \otimes a_2) = H(a_1) + H(a_2). \tag{6.4}$$

Theorem 3 and Corollary 3.1 are also described for normal states σ_i, ρ_i of A_i $(i=1,2)$ and for the direct product states $\sigma_1 \otimes \sigma_2$, $\rho_1 \otimes \rho_2$,

7. Information on subalgebra.

Let σ and ρ be a pair of probability measures over a measurable space $(\Omega, \mathfrak{A})$, and let $\mathfrak{B}$ be a Borel subfield of $\mathfrak{A}$. Denote $I_{\mathfrak{B}}(\sigma, \rho)$ the information between σ and ρ over $(\Omega, \mathfrak{B})$. Then

$$I_{\mathfrak{B}}(\sigma, \rho) \leqq I(\sigma, \rho).$$

In this section, we shall generalize this for von Neumann algebras. *Let B be an arbitary but fixed von Neumann subalgebra of A and let $e = E[\cdot \mid B]$*

be the conditional expectation conditioned by B. We shall begin with the following:

LEMMA 7.1. *For $a, b \varepsilon L^1(A)$, $a, b \geqq 0$, it holds that*

$$a \prec a^e \quad and \quad b \prec b^e.$$

If $a \prec b$ or $a \sim b$, then

$$a^e \prec b^e \quad or \quad a^e \sim b^e,$$

respectively.

Proof. Put $p = s(a)$, $p' = s(a^e)$, $q = s(b)$ and $q' = s(b^e)$ for their support-projections. Then, by $q' \varepsilon R(b^e)$,

$$\tau((1-q')b(1-q')) = \tau(b^e(1-q')) = \tau(b) - \tau(b) = 0$$

and

$$[(1-q')b^{1/2}]\cdot[(1-q')b^{1/2}]^* = (1-q')b(1-q') = 0.$$

Hence

$$(7.1) \qquad (1-q')b = b(1-q') = 0, \qquad q \leqq q' \text{ and } b \prec b^e.$$

Similarily $p \leqq p'$ and $a \prec a^e$.

Suppose $a \prec b$, then (7.1) implies

$$0 \leqq (1-q')r(1-q') \leqq (1-q')q(1-q') = 0$$

for every projection $r \varepsilon A$, $r \leqq q$, and hence

$$(1-q')a(1-q') = 0,$$
$$(1-q')a^e(1-q') = E[(1-q')a(1-q')|B] = 0$$

and $(1-q')a^e = 0$ or $a^e = q'a^e$. Therefore

$$(2.2) \qquad p \leqq p', \quad q \leqq q', \quad p \leqq q \quad \text{and} \quad p' \leqq q',$$

or $a^e \prec b^e$. Using this fact it is obvious that $a \sim b$ implies $a^e \sim b^e$.

By this lemma, *$I(a^e, b^e)$ for $a \prec b$ or $J(a^e, b^e)$ for $a \sim b$ ($a, b \varepsilon \mathcal{E}$) are well-defined.*

LEMMA 7.2. *For $a, b \varepsilon L^1(B' \cap A)$, $a, b \geqq 0$, the triples of the operators a^e, b^e and a, or a^e, b^e and b commute each others, respectively.*

Proof. The equalities $a^e x = x a^e$ and $b^e x = x b^e$ are obvious for $x = a, b$. To see $a^e b^e = b^e a^e$, it can be assumed a and b being bounded without loss of generality. Then for any $c \varepsilon A$

$$\tau(a^e b^e c) = \tau(a^e b^e c^e) = \tau(c^e a^e b) = \tau(c^e b a^e) = \tau(b^e a^e c)$$

and hence $a^e b^e = b^e a^e$.

By these lemmas, we can prove the following

THEOREM 4. *Let a and b be operators in $\mathcal{E}$ affiliated with B'. Then the following inequalities hold*

$$I(a^e, b^e) \leqq I(a, b) \qquad \text{if } a \prec b, \tag{7.3}$$

and

$$J(a^e, b^e) \leqq J(a, b) \qquad \text{if } a \sim b. \tag{7.4}$$

Proof. Firstly, we prove (7.3) when $I(a, b)$ and $I(a^e, b^e)$ are finite. Let p, p', q and q' be the projections defined in the proof of Lemma 7.1. Then

$$\begin{aligned}
&I(a, b) - I(a^e, b^e)\\
&= \tau(a[p \log a - q \log b]) - \tau(a^e[p' \log a^e - q' \log b^e])\\
&= \tau(a[p \log a - q \log b]) - \tau(a[p' \log a^e - q' \log b^e]) && \text{(by Corollary 2.1)}\\
&= \tau(a[p \log a - (q \log b + p' \log a^e - q' \log b^e)]) \qquad (7.5)\\
&= \tau(a[p \log a - p'qq' \log \{ba^e(b^e + (I - q'))^{-1}\}]) && \text{(by Lemma 7.2)}\\
&= \tau(a \log a - a \log \{ba^e(b^e + (I - q'))^{-1}\}) && \text{(by (7.1))}\\
&= I(a, ba^e(b^e + (I - q'))^{-1})
\end{aligned}$$

because $a \prec ba^e(b^e + (I - q'))^{-1} \geqq 0$ and

$$\tau(ba^e(b^e + (I - q'))^{-1}) = \tau(b^e a^e(b^e + (I - q'))^{-1}) = \tau(a^e) = \tau(a) = 1$$

and hence the last side of (7.5) is defined as non-negative, by Theorem 1.

If $I(a, b) = +\infty$, it is trivial.

If $I(a^e, b^e) = +\infty$, then since $\tau(a^e \log a^e)$ is finite,

$$\tau(a^e \log b^e) = \tau(a \log b^e) = -\infty.$$

By this fact and by the expression

$$a^e \log b^e = a^e \log^- b^e + a^e \log^+ b^e,$$

we obtain that

$$\tau(a \log^- b^e) = \tau(a^e \log^- b^e) = -\infty \quad \text{and} \quad \tau(a^e \log^+ b^e) = \text{finite} \geqq 0.$$

For the operator b, taking the operator b'' and the sequence of operators b_n' as the (3.5) and (3.6) in §3 (take b in the place of a), and putting $b_n = (b_n' + b'')/\tau(b_n' + b'')$, since

$$(b_n' + b'') \downarrow b, \qquad (b_n' + b'')^e \downarrow b^e \qquad \text{(a. e.)} \tag{7.6}$$

and $a^e b_n^e = b_n^e a^e$ by Lemma 7.2, it holds that

$$a^e \log b_n^e = a^e \log (b_n' + b'')^e - a^e \log \tau(b_n + b'') \rightarrow a^e \log b^e \qquad \text{(a. e.)}.$$

These facts and (7.6) imply

$$\tau(a^e \log b_n^e) \to \tau(a^e \log b^e) = -\infty.$$

The finiteness of $I(a^e, b_n^e)$ implies by the first part of this proof

$$I(a, b_n) \geqq I(a^e, b_n^e) \to +\infty.$$

While by (7.6)

$$\begin{aligned} I(a, b_n) &= \tau(a \log a) - \tau(a \log b_n) \\ &= \tau(a \log a) - \tau(a^{1/2}(\log b_n)a^{1/2}) \to \tau(a \log a) - \tau(a \log b) \\ &= I(a, b). \end{aligned}$$

Therefore $I(a, b) = +\infty$ and the inequality (7.3) is obtained. The rest part of Theorem 4 follows immediately from the above part. Q.E.D.

This theorem can be applied to proving the inequality for the measure of the informations between normal states.

Let σ, ρ be a pair of normal states of A and let σ_B, ρ_B be the restrictions of σ, ρ onto B. Then σ_B, ρ_B are also normal states of B and the information and the divergence between them are defined, which are denoted by $I_B(\sigma, \rho)$ and $J_B(\sigma, \rho)$. The assersion is following

THEOREM 4′. *If the states σ, ρ with finite entropy belong to the tracelet spaces T_B, i.e. satisfy the equality* (2.1), *then the following inequalities hold*:

$$I_B(\sigma, \rho) \leqq I(\sigma, \rho) \qquad \text{if } \sigma \prec \rho, \tag{7.8}$$

and

$$J_B(\sigma, \rho) \leqq J(\sigma, \rho) \qquad \text{if } \sigma \sim \rho. \tag{7.9}$$

Proof. The relation $\sigma \prec \rho$ is equivalent to $(d\sigma/d\tau) \prec (d\rho/d\tau)$ and this implies $(d\sigma/d\tau)^e \prec (d\rho/d\tau)^e$ by Lemma 7.1, where $e = E[\cdot \mid B]$. Hence the information $I((d\sigma/d\tau)^e, (d\rho/d\tau)^e)$ is uniquely determined and by Theorem 4

$$I\left(\left(\frac{d\sigma}{d\tau}\right)^e, \left(\frac{d\rho}{d\tau}\right)^e\right) \leqq I\left(\left(\frac{d\sigma}{d\tau}\right), \left(\frac{d\rho}{d\tau}\right)\right)$$

While $\tau((d\sigma/d\tau)^e b) = \sigma_B(b)$ for every $b \varepsilon B$ and $(d\sigma/d\tau)^e \varepsilon L^1(B)$ and similarily $(d\rho/d\tau)^e \varepsilon L^1(B)$. Consequently, these operators $(d\sigma/d\tau)^e$ and $(d\rho/d\tau)^e$ can be identified with the derivatives $d\sigma_B/d\tau_B$ and $d\rho_B/d\tau_B$ respectively. Therefore

$$I\left(\left(\frac{d\sigma}{d\tau}\right)^e, \left(\frac{d\rho}{d\tau}\right)^e\right) \leqq I(\sigma, \rho)$$

and (7.8) is obtained, and similarily (7.9).

8. Characterization of sufficient subalgebra.

The investigation of sufficiency in abstract form was initiated by Halmos and Savage [7] and they established a measure theoretic characterization of the sufficient statistics. While, under the notion of the information, another characterization was given by Kullback and Leibler [11]. In the previous paper [27], the author gave a non-commutative extension of this concept by introducing the notion of the sufficiency of subalgebra which is corresponding to the sufficiency of subfield, and he extended the Halmos-Savage's Theorem to the von Neumann algebra A. In the present section, it will be extended the Kullback-Leibler's Theorem to the algebra A.

Let S be a set of faithful normal states of A, and let M be a set of all operators $b \varepsilon A$ such that

(8.1) $$\rho(ab) = \rho(ba) \qquad \text{for every } a \varepsilon A \text{ and } \rho \varepsilon S.$$

Then M is a von Neumann subalgebra of A, and S is contained in the tracelet space T_M (cf. §2). Let B be a fixed von Neumann subalgebra of M, then

$$S \subset T_M \subset T_B.$$

Hence for a states $\rho \varepsilon S$ there corresponds uniquely a B-expectation $\epsilon = \epsilon(\rho)$ satisfying (2.2). Whence, if

(8.2) $$\epsilon(\sigma) = \epsilon(\rho) \qquad \text{for every pair of states } \sigma, \rho \varepsilon S$$

then B is called to be *sufficient for S or sufficient subalgebra for S* (cf. [27]).

Suppose, that *the set S consists of faithful normal states σ, ρ, $\cdots$ with finite entropy such that $I(\sigma, \rho)$ are finite for each pair $\sigma, \rho \varepsilon S$ and the set of the derivatives $\{d\sigma/d\tau;\ \sigma \varepsilon S\}$ is a commutative system, and that M and B be a von Neumann subalgebras, $B \subset M$, described above.* Then we prove

THEOREM 5. *Let $e = E[\cdot | B]$ be the conditional expectation. Then the following conditions are equivalent each others:*

(i) *B is sufficient for S,*

(ii) $I_B(\sigma, \rho) = I(\sigma, \rho)$ *for every pair* $\sigma, \rho \varepsilon S$

and

(iii) $J_B(\sigma, \rho) = J(\sigma, \rho)$ *for every pair* $\sigma, \rho \varepsilon S$.

Proof. Put $d(\sigma) = d\sigma/d\tau$ for every normal state σ. It was proved (in the proof of Theorem 5 in the previous paper [27]) that B is sufficient for S if and only if

(8.3) $$d(\sigma)\, d(\rho)^{-1} = d(\sigma)^e\, d(\rho)^{e^{-1}} \qquad \text{for every pair } \sigma, \rho \varepsilon S.$$ [7]

7) Even if A is semi-finite and τ is a semi-trace, this characteristic condition (8.3) for sufficiency is satisfied (cf. the proof of Theorem 5 in the preceding paper [27]).

Since $d(\sigma), d(\rho) \varepsilon L^1(B' \cap A)$, by Lemma 7.2, $d(\sigma)$, $d(\rho)$, $d(\sigma)^e$ and $d(\rho)^e$ are commuting each others. Therefore

$$
\begin{aligned}
I(\sigma, \rho) - I_B(\sigma, \rho) &= I(d(\sigma), d(\rho)) - I(d(\sigma)^e, d(\rho)^e) \\
&= \tau(d(\sigma)[\log d(\sigma) - \log (d(\rho)\, d(\sigma)^e\, d(\rho)^{e^{-1}})]) \\
&= I(d(\sigma), d(\rho)\, d(\sigma)^e\, d(\rho)^{e^{-1}}).
\end{aligned} \tag{8.4}
$$

Since $d(\rho)\, d(\sigma)^e\, d(\rho)^{e^{-1}}$ commutes $d(\sigma)$, by Theorem 2, (ii) is equivalent to $d(\sigma) = d(\rho)\, d(\sigma)^e\, d(\rho)^{e^{-1}}$ and so is to (8.3). Hence (ii) is equivalent to (i).

While, by (8.4)

$$
\begin{aligned}
& J(d(\sigma), d(\rho)) - J(d(\sigma)^e, d(\rho)^e) \\
&= I(d(\sigma), d(\rho)) - I(d(\sigma)^e, d(\rho)^e) + I(d(\rho), d(\sigma)) - I(d(\rho)^e, d(\sigma)^e) \\
&= I(d(\sigma), d(\rho)\, d(\sigma)^e\, d(\rho)^{e^{-1}}) + I(d(\rho), d(\sigma)\, d(\rho)^e\, d(\sigma)^{e^{-1}}).
\end{aligned}
$$

Therefore (iii) is equivalent to

$$
d(\sigma) = d(\rho)\, d(\sigma)^e\, d(\rho)^{e^{-1}} \quad \text{and} \quad d(\rho) = d(\sigma)\, d(\rho)^e\, d(\sigma)^{e^{-1}}. \tag{8.5}
$$

Since both equalities in (8.5) are obviously equivalent each other, (ii) and (iii) are equivalent each other.

9. Information with respect to a von Neumann subalgebra.

In the paper of Nakamura-Umegaki [16], it was defined a measure of information $I(a:B)$ of a with respect to a von Neumann subalgebra B such that

$$
I(a:B) = H(E[a \mid B]) - H(a). \tag{9.1}
$$

Then it is immediately proved that

THEOREM 6. *$I(a:B) = I(a, E[a \mid B])$ for every $a \varepsilon \mathcal{E}$.*

Proof. Putting $a^e = E[a \mid B]$, then

$$
\begin{aligned}
I(a:B) &= H(a^e) - H(a) \\
&= -\tau(a^e \log a^e) + \tau(a \log a),
\end{aligned}
$$

by $a \prec a^e$ (cf. Lemma 7.1) and by Corollary 2.1

$$
\begin{aligned}
&= \tau(a(\log a - \log a^e)) \\
&= I(a, a^e).
\end{aligned}
$$

If the operator $a \varepsilon \mathcal{E}$ is affiliated with B, then it holds always

$$
I(a:B) = 0. \tag{9.2}
$$

But the converse is not plain. However, if we introduce $J(\cdot:B)$, say *the divergence with respect to B*, such that

$$J(a:B) = J(a, E[a \mid B]) \quad \text{for every } a \in \mathcal{E},\ a \gg 0, \tag{9.3}$$

where, since $E[a \mid B] \gg 0$, the right-side of (9.3) is well-defined, then we obtain that

THEOREM 7. *For every $a \in \mathcal{E}$, $a \gg 0$,*

$$J(a:B) = 0 \tag{9.4}$$

if and only of a is affiliated with B.

Proof. If a is affiliated with B, then $a = E[a \mid B]$ and (9.4) is obvious. Conversely, if $J(a:B) = 0$, then

$$J(a, E[a \mid B]) = 0.$$

Hence, by Theorem 3, $a = E[a \mid B]$ and a is affiliated with B.

The measure of the informations can be applied to a characterization of the maximality of commutative von Neumann subalgebra:

THEOREM 8. *For a von Neumann subalgebra B, the following conditions are equivalent each others:*

(i) *B is a maximally alelian,*

(ii) *B is sufficient for T_B,*

(iii) *For every $a, b \in \mathcal{E}$, $a \prec b$, affiliated with B',*

$$I(a, b) = I(E[a \mid B],\ E[b \mid B])$$

and

(iv) $I(a:B) = 0$ *for every $a \in \mathcal{E}$ affiliated with B'.*

Proof. The equivalence between (i) and (ii) was proved in the previous paper (cf. Theorem 5 in [27]).

(i) implies (iii): (i) implies that $B' \cap A \subset B$ and $ab = ba$ for the operators a, b in (iii). Hence, putting $\sigma(\cdot) = \tau(a\cdot)$ and $\rho(\cdot) = \tau(b\cdot)$, the normal states σ, ρ belong to T_B and therefore by Theorem 5

$$I(E[a \mid B],\ E[b \mid B]) = I_B(\sigma, \rho) = I(\sigma, \rho) = I(a, b).$$

(iii) implies (iv): This is clear by putting $b = I$ in (iii).

(iv) implies (i): For any $a \in \mathcal{E} \cap B'$, $a \geqq 0$

$$I(a, E[a \mid B]) = I(a:B) = 0$$

by Theorem 6 and

$$aE[a \mid B] = E[a \mid B]a$$

hold. Hence by Theorem 2,

$$a = E[a \mid B]$$

which belongs to B, and (i) is obtained.

When A is commutative, (9.4) is equivalent to (9.2) for $a \gg 0$ by Theorem 2. This fact is applicable, in ordinary sample space, to finding the measurability of a random event with respect to a subfield as the following: The information $I(a(\cdot):\mathfrak{B})$ of a random event $a(\cdot)$ with respect to a Borel subfied $\mathfrak{B}$ is defined by

$$I(a(\cdot):\mathfrak{B}) = H(E(a(\cdot)\mid\mathfrak{B})) - H(a(\cdot)),$$

then

$$I(a(\cdot):\mathfrak{B}) = 0 \tag{9.2'}$$

is equivalent to $a(\cdot)$ being measurable with respect to $\mathfrak{B}$.

In quantum mechanical system, von Neumann [19] proved that a statistical operator U is changed to another statistical operator U' by a measurement, and each measurement increases the entropy:

$$H(U) \leqq H(U'). \tag{9.5}$$

The equality of this formula holds when and only when U is simultaneously observable with physical quantity corresponding to the measurement. This physical theory is described by the present mathematical formulation for the von Neumann algebra A being total ring of bounded operators over a Hilber space. In the paper of Nakamura-Umegaki [18], it was proved that the statistical development

$$U \to U' \tag{9.6}$$

by the measurements (with the operator S) is nothing but the conditional expectation $E[U\mid R(S)]$ and the inequality (9.5) is formulated by

$$H(U) \leqq H(E[U\mid R(S)]) \tag{9.5'}$$

and the equality in (9.5′) holds when and only when U belongs to $R(S)$, where $R(S)$ is the von Neumann algebra generated by S. Under the notion of the pinching operation, the inequality (9.5′) was generalized by Davis [3] for arbitary concave real function in the place of the function $h(\lambda)$ $(= -\lambda \log \lambda)$, in which he also gave the exact form in case of equality for arbitrary operator-concave functions.

REFERENCES

[1] BIRKHOFF, G., Moyennes des functions bornées. Algèbre et théorie des Collocq. Intern. CNRS. Paris, 1950, pp. 143–153.

[2] DAVIS, C., Various averaging operations onto subalgebras. Illinois J. Math. 3 (1959), 538–553.

[3] DAVIS, C., Operator-valued entropy of a quantum mechanical measurement. Proc. Japan Acad. 37 (1961), 533–538.

[4] DIXMIER, J., Les algèbres d'opérateurs dans l'espace Hilbertien. Paris (1957).
[5] DOOB, J. L., Stochastic processes. John Wiley, New York (1953).
[6] DYE, H. A., The Radon-Nikodym theorem for finite rings of operators. Trans. Amer. Math. Soc. 72 (1952), 243-280.
[7] HALMOS, P. R., AND L. J. SAVAGE, Application of the Radon-Nikodym theorem to the theory of sufficient statistics. Ann. Math. Satist. 20 (1949), 225-241.
[8] KHINCHIN, A. I., Mathematical foundations of information theory. Dover, New York (1957).
[9] KELLEY, J. L., Averaging operators on $C_\infty(X)$. Illinois J. Math. 2 (1958).
[10] KULLBACK, S., Information theory and statistics. John Wiley, New York (1959).
[11] KULLBACK, S., AND R. A. LEIBLER, On information and sufficiency. Ann. Math. Statist. 22 (1951), 79-86.
[12] MISONOU, Y., On the direct product of W^*-algebras. Tôhoku Math. J. 6 (1954), 189-204.
[13] NAKAMURA, M., AND T. TURUMARU, Expectations in an operator algebra. Tôhoku Math. J. 6 (1954), 182-188.
[14] NAKAMURA, M., AND H. UMEGAKI, On a proposition of von Neumann. Kōdai Math. Sem. Rep. 8 (1956), 142-144.
[15] NAKAMURA, M., M. TAKESAKI AND H. UMEGAKI, A remark on the expectations of operator algebras. Kōdai Math. Sem. Rep. 12 (1960), 82-90.
[16] NAKAMURA, M., AND H. UMEGAKI, A note on the entropy for operator algebras. Proc. Japan Acad. 37 (1961), 149-154.
[17] NAKAMURA, M., AND H. UMEGAKI, On the Blackwell theorem in operator algebra. Proc. Japan Acad. 37 (1961), 312-315.
[18] NAKAMURA, M., AND H. UMEGAKI, On von Neumann's theory of measurements in quantum statistics. Math. Jap. 7 (1962), 151-157.
[19] NEUMANN, J. VON, Mathematical foundations of quantum mechanics. Princeton Univ. Press, Princeton (1955).
[20] ROTA, G. C., On the representation of averaging operators. Rend. Semin. Math. 30 (1960), 52-64.
[21] SEGAL, I. E., A non-commutative extension of abstract integration. Ann. Math. 57 (1953), 401-457.
[22] SEGAL, I. E., A note on the concept of entropy. J. Math. Mech. 9 (1960), 623-629.
[23] SHANNON, G. R., A mathematical theory of communication. Bell System Tech. J. 27 (1948), 379-423; 623-656.
[24] TOMIYAMA, J., On the projection of norm one in W^*-algebra. Proc. Japan Acad. 33 (1957), 608-612.
[25] UMEGAKI, H., Conditional expectation in an operator algebra. Tôhoku Math. J. 6 (1954), 177-181.
[26] UMEGAKI, H., Conditional expectation in an operator algebra, II. Tôhoku Math. J. 8 (1956), 86-100.
[27] UMEGAKI, H., Conditional expectation in an operator algebra, III. Kōdai Math. Sem. Rep. 11 (1959), 51-64.
[28] UMEGAKI, H., On information in operator algebras. Proc. Japan Acad. 37 (1961), 459-461.

DEPARTMENT OF MATHEMATICS,
TOKYO INSTITUTE OF TECHNOLOGY.