Natural Science Report, Ochanomizu University, Vol. 4, No. 2, 1953
お茶の水女子大学 自然科学報告 第4巻 第2号

Dependent Experiments and Sufficient Statistics

Hirokichi Kudō (工藤弘吉)

Department of Mathematics, Faculty of Science,
Ochanomizu University, Tokyo

It seems that any sufficient statistic contains all the informations in a sample. This is a justification for use of sufficient statistics (see [1]). E. L. Lehmann showed, in this respect, that for any test function $\phi(x)$ there exists a test function $\psi(t)$ of a given sufficient statistic $t(x)$, such that $E_\theta\phi(x)=E_\theta\psi(t)$ holds for every θ in the parameter space Ω of distributions (Theorem 4. 1, [2], *cf*. [3]). Is the converse of the above statement true? Blackwell's paper [5] answered this question in the case of Ω consisting of n simple hypotheses. According to his results [4], [5], the sufficiency of a statistic $t(x)$ is equivalent to the condition that $t(x)$ contains all the informations of a sample in *every* decision problems, if we restrict ourselves to the case of finite simple hypotheses.

The principal aim of this paper is to prove the converse proposition of the above Lehmann's result in Blackwell's case. To attain this aim, we shall study the properties of the joint "experiment" of two dependent ones. The other aim of this paper is these studies.

The term "experiment" was definded, by Blackwell, as a finite ordered set of distributions. For two independent experiments α and β, Blackwell [4] (and the author [6] in the case $n=2$) defined the product $\alpha\times\beta$ as the set of joint distributions of corresponding components of α and β. And in the case when every component of β depends on the corresponding component of α, a similar definition can be adopted. In this paper we shall give an integral representation to this joint experiment.

The principal tool in the following is the theory of convex bodies. In the first section we shall outline the relations between the supporting function and the convex body. The details of this may be found in [7]. In section 2 we shall define an integral of a convex body valued function $\boldsymbol{A}(x)$ with respect to an experiment, and learn properties of this integral (Theorem I and II). It will be clear, in Theorem III of section 3, that $\boldsymbol{A}(x)$ can be regarded as a conditional experiment of a dependent variable, and then the first aim will be attacked in Theorem IV.

1. Supporting function. Denote the n-space by R^n, and its points by $\xi=(\xi_1, \cdots, \xi_n)$. Suppose that $\boldsymbol{A}\subset R^n$ be a convex body (=bounded, closed, convex and non-empty subset), and denote the supporting function of $\boldsymbol{A}$ by $H(\xi, \boldsymbol{A})$, that is, the function $H(\xi, \boldsymbol{A})$ is defined by

$$H(\xi, A)=\sup_{a=(a_1,\dots,a_n)\in A} \sum_{i=1}^n \xi_i a_i .$$

Thus defined $H(\xi, A)$ is a continuous function of ξ, such that

i) $H(0, A)=0$,

ii) $H(k\xi, A)=kH(\xi, A)$ for all positive number k.

iii) $H(\xi+\eta, A)\leqq H(\xi, A)+H(\eta, A)$,

and therefore

iv) all differentials

$$H'(\xi, \eta, A)=\lim_{h\to 0+} \frac{H(\xi+h\eta, A)-H(\xi, A)}{h}$$

of $H(\xi, A)$ exist. (see [7])

Conversely, a real valued function $H(\xi)$ of $\xi\in R^n$ satisfying the conditions i), ii) and iii) determines a convex body.

The differential $H'(\xi, \eta, A)$ has the following meaning: The intersection A_ξ of the body A and the hyperplane $\sum\xi_i a_i=H(\xi, A)$ is an at most $(n-1)$-dimensional convex body supported by hyperplanes $\sum\eta_i a_i= H'(\xi, \eta, A)$. This set A_ξ is a part of the surface of A. Each point of this set A_ξ is called the *supporting point* of A in direction ξ.

These differentials $H'(\xi, \eta, A)$ being also the supporting function of A_ξ, each of them has a differential $H''(\xi, \eta, \zeta, A)$, and so on. Thus we have, for n vectors $\xi^1, \xi^2, \dots, \xi^n$, a set of supporting functions $H(\xi^1, A)$, $H'(\xi^1, \xi^2, A), \dots, H^{(n-1)}(\xi^1, \xi^2, \dots, \xi^n, A)$, such that $H^{(i)}(\xi^1, \dots, \xi^{i+1}, A)$ is the differential of $H^{(i-1)}(\xi^1, \dots, \xi^i, A)$ in direction $\xi^{i+1}=(\xi_1^{i+1}, \xi_2^{i+1}, \dots, \xi_n^{i+1})$, and that $H^{(i)}(\xi^1, \dots, \xi^{i+1}, A)$ is a supporting function of the at most $(n-i)$-dimensional convex body $A_{\xi^1,\dots,\xi^i}$, the intersection of $A_{\xi^1,\dots,\xi^{i-1}}$ and the hyperplane $\sum\xi_j^i a_j=H^{(i-1)}(\xi^1, \dots, \xi^i, A)$, $i=1, 2, \dots, n-1$.

For n linearly independent vectors $\xi^1, \xi^2, \dots, \xi^n$ in R^n, the solution $a=(a_1, a_2, \dots, a_n)$ of a system of linear equations

$$(1)\qquad \begin{cases} \sum \xi_i^1 a_i=H(\xi^1, A) , \\ \sum \xi_i^2 a_i=H'(\xi^1, \xi^2, A) , \\ \cdots \quad \cdots \\ \sum \xi_1^n a_i=H^{(n-1)}(\xi^1, \xi^2, \dots, \xi^n, A) , \end{cases}$$

exists and is unique. This solution $a=(a_1, \dots, a_n)$ is obviously an extreme point of A. And every extreme point is given by the solution of such system.

2. Integral of convex bodies. Let $(X, \mathfrak{B})$ be a measurable space, $\mu(E)=(\mu_1(E), \dots, \mu_n(E))$ $(0\leqq\mu_i\leqq 1, i=1, \dots, n)$ be an n-dimensional vector valued measure or an experiment defined on $(X, \mathfrak{B})$. The component measures μ_i $(i=1, \dots, n)$ of μ being absolutely continuous with respect to a measure $\lambda(E)=(\sum_{i=1}^n\mu_i(E))/n$, we can represent μ_i as an integral

$$\mu_i(E)=\int_E f_i(x)d\lambda(x)\ ,\quad E\in\mathfrak{B}\ ,$$

where $f_i(x)$ is a $\mathfrak{B}$-measurable function ($i=1, \ldots, n$).

A real valued function $H(\xi, x)$ of a vector $\xi\in R^n$ and a point x of the space X is called a *measurable supporting function*, if $H(\xi, x)$ satisfies i), ii), iii) of section 1 and is bounded as a function of ξ for almost (λ) all fixed x, and is $\mathfrak{B}$-measurable as a function of x for any fixed ξ. This measurable supporting function $H(\xi, x)$ defines a convex body $\boldsymbol{A}(x)$ for almost (λ) every $x\in X$. We call this mapping $\boldsymbol{A}(x)$ of X into a set of convex bodies a *measurable convex body valued function* on X.

Lemma 1. *Let $H(\xi, x)$ be a supporting function for almost (λ) every fixed $x\in X$. It is a necessary and sufficient condition for $H(\xi, x)$ to be a measurable supporting function that, for any $\xi\in R^n$, there exists a $\mathfrak{B}$-measurable vector valued function $a(x, \xi)=(a_1(x, \xi), \ldots, a_n(x, \xi))$ belonging to $\boldsymbol{A}_\xi(x)$.*

Proof. Sufficiency. $H(\xi, x)=\sum\xi_i a_i(x, \xi)$ is evidently $\mathfrak{B}$-measurable, since all $a_i(x, \xi)$ are $\mathfrak{B}$-measurable.

Necessity. $H(\xi^1, x)$ being $\mathfrak{B}$-measurable, $H'(\xi^1, \xi^2, x), \ldots, H^{(n-1)}(\xi^1, \ldots, \xi^n, x)$ are all $\mathfrak{B}$-measurable. Hence the solution $a(x; \xi^1, \ldots, \xi^n)$ of the system (1) is $\mathfrak{B}$-measurable function for any fixed $\xi^1, \ldots, \xi^n$. Putting $\xi=\xi^1$ and $a(x, \xi)=a(x, \xi, \xi^2, \ldots, \xi^n)$, our Lemma is proved.

In the rest of this section, we shall assume that $\boldsymbol{A}(x)$ is measurable. By $\Delta(\boldsymbol{A}(x))$ we denote the family of all $\mathfrak{B}$-measurable vector functions $a(x)=(a_1(x), \ldots, a_n(x))$, whose values belong to $\boldsymbol{A}(x)$ for almost (λ) every $x\in X$, and by $v(a)$ the vector in R^n whose i-th component is

$$\int_X a_i(x)d\mu_i(x)=\int_X a_i(x)f_i(x)d\lambda(x),\quad i=1, \ldots. n\ .$$

Further, by $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$ we mean the set of all vectors $v(a)$ for $a\in\Delta(\boldsymbol{A}(x))$.

Now we shall try to seek for the form of the supporting function $K(\xi)$ and a supporting point $b(\xi)=(b_1(\xi), \ldots, b_n(\xi))$ of the smallest convex body $\boldsymbol{K}$, containing $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$, in direction ξ. Denote a vector $(\xi_1 f_1(x), \ldots, \xi_n f_n(x))\in R^n$ by $\xi\times f(x)$, and a supporting point of $\boldsymbol{A}(x)$ in direction ξ by $a(x, \xi)=(a_1(x, \xi), \ldots, a_n(x, \xi))$. Under these notations, the vector $b(\xi)=(b_1(\xi), \ldots, b_n(\xi))$:

$$b_i(\xi)=\int_X a_i(x, \xi\times f(x))d\mu_i(x)\ ,$$ [1]

is a supporting point of $\boldsymbol{K}$ in direction ξ, i.e.

[1] $H(\xi\times f(x), x)$ is $\mathfrak{B}$-measurable, because $H(\xi, x)$ is a continuous function of ξ and $f(x)$ is $\mathfrak{B}$-measurable. Similarly all $H^{(i)}(\xi^1\times f(x), \cdots, \xi^{i+1}\times f(x), x)$ are $\mathfrak{B}$-measurable. From this fact, there exists a supporting point $a(x, \eta)$ such that $a_i(x, \xi\times f(x))$ are $\mathfrak{B}$-measurable, $i=1, \cdots, n$.

$$\sum \xi_i b_i(\xi) \geqq \sum \xi_i \int_X a_i(x) d\mu_i \text{ for every } a(x) \in \Delta(A(x)) .$$

In fact, since $a(x, \eta)$ is a supporting point of $\boldsymbol{A}(x)$, it holds for any $a(x) \in \Delta(\boldsymbol{A}(x))$ that

$$\begin{aligned}\sum \xi_i \int_X a_i(x) d\mu_i &= \int_X \{\sum a_i(x) \xi_i f_i(x)\} d\lambda \\ &\leqq \int_X \{\sum a_i(x, \xi \times f(x)) \xi_i f_i(x)\} d\lambda \\ &= \sum \xi_i \int_X a_i(x, \xi \times f(x)) d\mu_i = \sum \xi_i b_i(\xi) .\end{aligned}$$

Therefore $K(\xi) = \sum \xi_i b_i(\xi)$ is a supporting function of $\boldsymbol{K}$, provided that $a_i(x, \xi \times f(x))$ is μ_i-integrable $(i=1, \dots, n)$. Noting that $\sum \xi_i f_i(x) a_i(x, \xi \times f) = H(\xi \times f, x)$, we have

$$K(\xi) = \int_X H(\xi \times f, x) d\lambda(x) . \tag{2}$$ [1]

However, since the μ_i-integrability of $a_i(x, \xi \times f(x))$ $(i=1, \dots, n)$ is equivalent to the λ-integrability of $H(\xi \times f(x), x)$, we can define the *integrability* of a convex body valued function $\boldsymbol{A}(x)$ by the λ-integrability of $H(\xi \times f(x), x)$.

Theorem 1. *For every integrable* $\boldsymbol{A}(x)$, $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$ *is a convex body.*

Proof. The convexity of $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$ follows evidently from that of $\Delta(\boldsymbol{A}(x))$. Therefore it is sufficient to prove that every extreme points of $\boldsymbol{K}$ belongs to $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$. Any extreme point $b^0 = b^0(\xi^1, \dots, \xi^n) = (b_1^0, \dots, b_n^0)$ of $\boldsymbol{K}$ is the solution of the system of equations :

$$\left\{\begin{aligned} &\sum b_i^0 \xi_i^1 = K(\xi^1) , \\ &\sum b_i^0 \xi_i^2 = K'(\xi^1, \xi^2) , \\ &\cdots \quad \cdots \\ &\sum b_i^0 \xi_i^n = K^{(n-1)}(\xi^1, \dots, \xi^n), \end{aligned}\right. \tag{3}$$

where $\xi^1, \dots, \xi^n$ are linearly independent vectors.

On the other hand, from the convexity of $H(\xi, x)$, we have

$$\begin{aligned} -H(-\xi^2 \times f(x), x) &\leqq \frac{H((\xi^1 + k\xi^2) \times f(x), x) - H(\xi^1 \times f(x), x)}{k} \\ &\leqq H(\xi^2 \times f(x), x) , \end{aligned}$$

and hence it follows from Lebesgue's theorem that

$$K'(\xi^1, \xi^2) = \int_X H'(\xi^1 \times f(x), \xi^2 \times f(x), x) d\lambda(x) .$$

Similarly we can show the following equations

1) See the footnote on page 153.

$$K^{(i)}(\xi^1, \ldots, \xi^{i+1}) = \int_X H^{(i)}(\xi^1 \times f(x), \ldots, \xi^{i+1} \times f(x), x) d\lambda(x) ,$$
$$i=1, \ldots, n-1.$$

From the uniqueness of the solution of (3), we have

$$b_i^0(\xi^1, \ldots, \xi^n) = \int_X a_i^0(\xi^1, \ldots, \xi^n, x) d\mu_i(x) , \quad i=1, \ldots, n ,$$

where $a_i^0 = a_i^0(\xi^1, \ldots, \xi^n, x)$ $(i=1, 2, \ldots, n)$ satisfy

$$\begin{aligned} &\sum a_i^0 \xi_i^1 f_i(x) = H(\xi^1 \times f(x), x) , \\ &\sum a_i^0 \xi_i^2 f_i(x) = H'(\xi^1 \times f(x), \xi^2 \times f(x), x) , \\ &\cdots \quad \cdots \\ &\sum a_i^0 \xi_i^n f_i(x) = H^{(n-1)}(\xi^1 \times f(x), \ldots, \xi^n \times f(x), x) , \end{aligned}$$

Hence $a^0 = (a_1^0, \ldots, a_n^0) \in \Delta(\boldsymbol{A}(x))$. This shows $b^0 \in \boldsymbol{R}(\mu, \boldsymbol{A}(x))$.

Since $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$ is of the form $\left\{\int_X a(x) d\mu(x); a \in \Delta(\boldsymbol{A}(x))\right\}$, in the following we shall use the notation

$$\int \boldsymbol{A}(x) d\mu(x)$$

instead of $\boldsymbol{R}(\mu, \boldsymbol{A}(x))$, analogically to the notation $f(E) = \{f(x); x \in E\}$ in the functional theory. We call this *the integral of* $\boldsymbol{A}(x)$ with respect to the experiment $\mu = \{\mu^1, \cdots, \mu^n\}$.

From this definition of the integral, it follows directly that $\boldsymbol{A}(x) \supseteqq \boldsymbol{B}(x)$ for almost (λ) all $x \in X$ implies

$$\int \boldsymbol{A}(x) d\mu \supseteqq \int \boldsymbol{B}(x) d\mu .$$

If $\boldsymbol{A}(x) \equiv \boldsymbol{A}$, that is, if every $\boldsymbol{A}(x)$ are always a convex body $\boldsymbol{A}$, we write

$$\int \boldsymbol{A}(x) d\mu = \boldsymbol{A} \cdot \mu .$$

This set $\boldsymbol{A} \cdot \mu$ coincides with $R(\mu, \boldsymbol{A})$ in [4] and $B(\mu, \boldsymbol{A})$ in [5]. If $\boldsymbol{A}(x)$ is contained in a convex body $\boldsymbol{A}$ for almost (λ) all $x \in X$, we have

$$\int \boldsymbol{A}(x) d\mu \subseteqq \boldsymbol{A} \cdot \mu .$$

Let $R(x)$ be the linear subspace consisting of all vectors $\xi \in R^n$ whose i-th component ξ_i vanishes for every $i : f_i(x) = 0$.

Theorem II. *Let us denote by* $\boldsymbol{A}_1(x)$, $\boldsymbol{B}_1(x)$ *the projections of* $\boldsymbol{A}(x)$ *and* $\boldsymbol{B}(x)$ *on* $R(x)$, *respectively. If* $\boldsymbol{A}(x) \supseteqq \boldsymbol{B}(x)$ *holds almost* (λ) *everywhere on* X, *and if the set* $E = \{x; \boldsymbol{A}_1(x) \supsetneqq \boldsymbol{B}_1(x)\}$ *has a positive* λ*-measure, then*

$$\int A(x)d\mu \geqq\!\!\!\!\!\!\not\;\; \int B(x)d\mu \ .$$

Proof. Let $H_A(\xi, x)$ and $H_B(\xi, x)$ be the supporting functions of $A(x)$ and $B(x)$, respectively. Under the assumption of our theorem, we have $H_A(\xi, x) \geqq H_B(\xi, x)$ for almost (λ) all $x \in X$ and all $\xi \in R^n$, and for every $x \in E$ the set $Q_x = \{\xi; \xi \in R(x) \text{ and } H_A(\xi, x) > H_B(\xi, x)\}$ is not empty, and hence, by the continuity of the supporting function, Q_x is open and has the property :

(*) *If* $\xi \in Q_x$, *then* $k\xi \in Q_x$ *for every* $k>0$.

Therefore the set $Q'_x = \{\xi; H_A(\xi \times f(x); x) > H_B(\xi \times f(x); x)\}$ is also open and has the above properties (*) for every $x \in E$. The Lebesgue measure of the intersection of Q'_x and the unit sphere C of R^n being non-zero, for every $x \in E$ holds

$$\int_C \{H_A(\xi \times f(x), x) - H_B(\xi \times f(x), x)\} d\xi > 0 \ .$$

By the equation (2), the supporting functions of $\int A(x)d\mu$ and $\int B(x)d\mu$ are respectively

$$K_A(\xi) = \int_X H_A(\xi \times f(x), x) d\lambda(x) \ ,$$

and

$$K_B(\xi) = \int_X H_B(\xi \times f(x), x) d\lambda(x) \ .$$

Hence by Fubini's theorem

$$\begin{aligned}\int_C \{K_A(\xi) - K_B(\xi)\} d\xi &= \int_C d\xi \int_X \{H_A(\xi \times f(x), x) - H_B(\xi \times f(x), x)\} d\lambda(x) \\ &= \int_X d\lambda(x) \int_C \{H_A(\xi \times f(x), x) - H_B(\xi \times f(x), x)\} d\xi > 0 \ .\end{aligned}$$

Since $K_A(\xi) \geqq K_B(\xi)$ holds for every $\xi \in R^n$, there exists a $\xi \in R^n$ at which $K_A(\xi) > K_B(\xi)$ holds, and the proof is complete.

In the next section we shall denote by I the segment (=a convex body) joining the origin $(0, \cdots, 0)$ with the point $e = (1, 1, \cdots, 1)$ in R^n, and write

$$L_\mu = I \cdot \mu \ .$$

3. Conditional Experiment and Sufficiency. Let $T(x)$ be a measurable mapping of X onto another measurable space $(Y, \mathfrak{C})$, that is to say, $\mathfrak{B} \supset T^{-1}(\mathfrak{C})$. Let $\nu(F) = (\nu_1(F), \cdots, \nu_n(F))$ be an n-dimensional vector measure on $(Y, \mathfrak{C})$, such that $\nu_i(F) = \mu_i T^{-1}(F)$ for $i = 1, \cdots, n$, and let be $\tau(F) = \sum \nu_i(F)/n$. A vector valued function $g(t) = (g_1(t), \cdots, g_n(t))$ is such that

$$\nu_i(F) = \int_F g_i(t) d\tau(t) \ , \quad i = 1, \cdots, n \ .$$

This measure $\tau(F)$ satisfies $\lambda(T^{-1}(F))=\tau(F)$. Further, for every $\mathfrak{B}$-measurable function $\varphi(x)$ $(0\leqq\varphi\leqq 1)$, there exist n integrable functions $Q_i(\varphi;t)$ defined on $(Y, \mathfrak{C}, \nu)$, such that

$$\int_{T^{-1}F}\varphi(x)d\mu_i=\int_F Q_i(\varphi;t)d\nu_i \qquad \text{for every } F\in\mathfrak{C}. \tag{4}$$

(see Kolmogoroff [8]). In the following discussions we do not assume that these functions $Q_i(\varphi;t)$ are conditional expectations of φ, in spite of the fact that they can be regarded as conditional expectations under some regular conditions. However, since $Q_i(\varphi;t)$ can be nomalized without any condition, we shall assume that the values of $Q_i(\varphi;t)$, $i=1,\cdots,n$, belong to the interval $[0,1]$ at every t on Y.

From the definition of the supporting function $K(\xi)$ of $\boldsymbol{L}_\mu=\boldsymbol{I}\cdot\mu$, there exists a $\mathfrak{B}$-measurable function $\varphi^\xi(x)$ for any vector ξ, which satisfies $0\leqq\varphi^\xi(x)\leqq 1$ and

$$\begin{aligned} K(\xi)&=\sum\xi_i\int_X\varphi^\xi(x)d\mu_i(x) \\ &=\int_Y\{\sum\xi_iQ_i(\varphi^\xi,t)g_i(t)\}d\tau(t)\,. \end{aligned} \tag{5}$$

Let us now write

$$H_1(\xi,t)=\sum\xi_ig_i(t)Q_i(\varphi^\xi;t)\,. \tag{6}$$

Lemma 2. *Let Ξ be the set of all the vectors in R^n whose components are all rational. For every $\mathfrak{B}$-measurable function $\varphi(x)$ $(0\leqq\varphi(x)\leqq 1)$, there exists a set $N_\varphi(\in\mathfrak{C})$ of τ-measure zero, such that*

$$\sum\xi_ig_i(t)Q_i(\varphi;t)\leqq H_1(\xi,t)$$

holds whenever $\xi\in\Xi$, and $t\notin N_\varphi$.

Proof. Suppose that the set

$$N_\xi=\{t;\sum\xi_ig_i(t)Q_i(\varphi;t)>H_1(\xi,t)\}$$

has a positive τ-measure for $\xi\in R^n$. For a $\mathfrak{B}$-measurable function

$$\varphi^*(x)=\begin{cases}\varphi(x), & \text{when } x\in T^{-1}(N_\xi)\,,\\ \varphi^\xi(x), & \text{when } x\notin T^{-1}(N_\xi)\,,\end{cases}$$

we have

$$\begin{aligned}\sum\xi_i\int_X\varphi^*(x)d\mu_i&=\sum\xi_i\int_{T^{-1}(N_\xi)}\varphi(x)d\mu_i+\sum\xi_i\int_{X-T^{-1}(N_\xi)}\varphi^\xi(x)d\mu_i\\ &=\sum\xi_i\int_{N_\xi}Q_i(\varphi;t)d\nu_i+\int_{Y-N_\xi}H_1(\xi,t)d\tau\\ &>\int_YH_1(\xi,t)d\tau=K(\xi)\,.\end{aligned}$$

This contradicts with the definition of the supporting function $K(\xi)$, and hence $\tau(N_\xi)=0$. Writting $N_\varphi=\bigcup_{\xi\in\Xi} N_\xi$, the proof is complete.

Lemma 3. *There is a set N of τ-measure zero such that, if $t\notin N$, then $H_1(\xi, t)$ has the following three conditions:*

i) *$H_1(k\xi, t)=kH_1(\xi, t)$ whenever $\xi\in\Xi$, and $k>0$ is a rational number,*

ii) *$H_1(\xi, t)+H_1(\eta, t)\geqq H_1(\xi+\eta, t)$ whenever $\xi, \eta\in\Xi$,*

iii) *$H_1(\xi, t)$ is bounded in the intersection of Ξ and the unit sphere of R^n.*

Proof. Let

$$N=\bigcup_{\xi\in\Xi} N_{\varphi^\xi}\ .$$

The condition i) holds true for $t\notin N$, because by Lemma 2 the inequalities

$$k\sum\xi_i g_i(t)Q_i(\varphi^{k\xi}; t)\leqq kH_1(\xi, t)\ ,$$

and

$$\sum k\xi_i g_i(t)Q_i(\varphi^\xi; t)\leqq H_1(k\xi, t)$$

hold for $\xi\in\Xi$ and any positive rational number k.

re ii). If $t\notin N_{\varphi^{\xi+\eta}}\subset N$, then we have

$$\begin{aligned}H_1(\xi, t)+H_1(\eta, t)-H_1(\xi+\eta, t)=&\{H_1(\xi, t)-\sum\xi_i g_i(t)Q(\varphi^{\xi+\eta}; t)\}\\ &+\{H_1(\eta, t)-\sum\eta_i g_i(t)Q(\varphi^{\xi+\eta}; t)\}\geqq 0.\end{aligned}$$

The condition iii) is clear from the following inequalities

$$\begin{aligned}|H_1(\xi, t)|&\leqq\sum|\xi_i g_i(t)|\cdot|Q_i(\varphi^\xi; t)|\\ &\leqq\sum|\xi_i g_i(t)|\leqq K\sqrt{\sum\xi_i^2}\ ,\end{aligned}$$

where K is a constant independent of t and ξ.[1]

By the condition i), ii) and iii) of Lemma 3, the function $H_1(\xi, t)$ is continuous on Ξ for every fixed $t\notin N$, and hence any continuous extension of $H_1(\xi, t)$ onto R^n defines a convex body $\boldsymbol{A}_1(t)$ whenever $t\notin N$.

Lemma 4. i) *The function $H_1(\xi, t)$ is the measurable supporting function defining a measurable convex body valued function $\boldsymbol{A}_1(t)$ for almost (τ) all $t\in Y$.*

ii) *For every $\mathfrak{B}$-measurable function φ $(0\leqq\varphi\leqq 1)$, there exists a vector valued function $Q(\varphi; t)$ belonging to the convex body $A_1(t)$ almost (τ) everywhere on Y, and satisfying* (4).

Proof. Let $\xi^1, \xi^2, \cdots, \xi^m, \cdots$ be a sequence of vectors of Ξ tending to a vector ξ outside of Ξ. By Lemma 2, we have

$$\sum\xi_i^m g_i(t)Q_i(\varphi; t)\leqq H_1(\xi^m, t)\ ,$$

and, by letting $m\to\infty$,

---

[1] We shall assume in this paper without any loss of generality that $g_i(t)$ is bounded on Y for every $i=1, 2, \cdots, n$.

$$\sum \xi_i g_i(t) Q_i(\varphi; t) \leqq \lim_{m\to\infty} H_1(\xi^m, t) \tag{7}$$

for every $t \notin N$. Suppose that there exist a sequence $\{\xi^m\}$ $(\xi^m \in \Xi)$ and a vector $\xi \notin \Xi$, such that

$$\lim_{m\to\infty} \xi^m = \xi ,$$

and that

$$M_\xi = \{t; H_1(\xi, t) < \lim_{m\to\infty} H_1(\xi^m, t)\}$$

has a positive τ-measure. Let ε be a positive number, such that

$$M_\varepsilon = \{t; H_1(\xi, t) + \varepsilon < \lim_{m\to\infty} H_1(\xi^m, t)\}$$

is also of positive τ-measure. By Egoroff's theorem, there exists a subset $M' \subset M_\varepsilon$ such that $\tau(M') > 0$ and the convergence of the sequence $\{H_1(\xi^m, t)\}$ is uniform on M'. Therefore, since the sequence $\{\sum \xi_i g_i(t) Q_i(\varphi \xi^m, t) - H_1(\xi^m, t)\}$ tends uniformly to zero as $\xi - \xi^m \to 0$,[2] it holds

$$\sum \xi_i g_i(t) Q_i(\varphi^{\xi^m}; t) > H_1(\xi^m, t) - \frac{\varepsilon}{2} > \lim_{m\to\infty} H_1(\xi^m, t) - \varepsilon > H_1(\xi, t)$$

for every large m on a set of positive τ-measure. This is a contradiction, which shows, combining with (7), that

$$H_1(\xi, t) = \lim_{m\to\infty} H_1(\xi^m, t)$$

almost (τ) everywhere on Y, that is, that the function $H_1(\xi, t)$ is the supporting function of an $\boldsymbol{A}_1(t)$. The measurability of $H_1(\xi, t)$ is evident from its definition (6). The condition ii) follows from (7).

Let $S(t)$ be the set $\{i; g_i(t) \neq 0\}$, $R_s(t)$ be the linear space consisting of all vectors $\xi (\in R^n)$ with their i-th components zero for $i \notin S(t)$, and $R_c(t)$ be the linear space whose vectors are all rectangular to the vector $e = (1, 1, \cdots, 1)$ and the linear space $R_s(t) \cap \{e\}^{\perp}$. By these two linear spaces $R_s(t)$ and $R_c(t)$, we have the unique decomposition of $\xi \in R^n$, such that

$$\xi = \xi(t) \times g(t) + \xi'(t) , \quad \xi(t) \in R_s(t) \quad \text{and} \quad \xi'(t) \in R_c(t) .$$

Using this decomposition, we shall define the function

$$H(\xi, t) = H_1(\xi(t), t) .$$

This function is a measurable supporting function. In fact the measurability follows from the continuity of $H_1(\xi, t)$ with respect to ξ, and the $\mathfrak{S}$-measurability of $\xi(t)$. And the convexity of $H_1(\xi, t)$ and the equalities $(\xi + \eta)(t) = \xi(t) + \eta(t)$ and $(k\xi)(t) = k\xi(t)$ $(\xi, \eta \in R^n, k \geqq 0)$ imply the fact that $H(\xi, t)$ is a supporting function.

On the other hand, we shall conventionally define the values of

---

[2] See the footnote on page 158.

$Q_i(\varphi; t)$ $(i \notin S(t))$ as follows:

$$Q_i(\varphi; t) = [\textit{the arithmetic mean of } \{Q_j(\varphi; t);\ j \in S(t)\}]\ ,$$

since the values of $Q_i(\varphi; t)$ for $i \notin S(t)$ are undetermined and $S(t)$ is not empty for almost (τ) all $t \in Y$.

Under these definitions of $H(\xi, t)$ and $Q(\varphi; t)$, Lemma 4 implies directly the following

Corollary. $H(\xi, t) = H_1(\xi(t), t)$ determines a measurable convex body valued function $\boldsymbol{A}(t)$, and for every $\mathfrak{B}$-measurable φ $(0 \leqq \varphi \leqq 1)$ the vector $Q(\varphi; t)$ runs almost (τ) always inside of $\boldsymbol{A}(t)$ when t runs on Y.

By noting (5) and the relation

$$H(\xi \times g(t), t) = H_1(\xi, t) = \sum \xi_i g_i(t) Q_i(\varphi^\xi; t),$$

we have

$$\boldsymbol{L}_\mu = \int \boldsymbol{A}(t) d\nu\ .$$

Therefore from the above Corollary follows

Theorem III. *For an experiment (=a vector valued measure)* $\mu = (\mu_1, \cdots. \mu_n)$ *defined on* $(X, \mathfrak{B})$ *and for a statistic (=a $\mathfrak{B}$-measurable mapping)* $T(x)$ *with values in* Y, *there corresponds a measurable convex valued function* $\boldsymbol{A}(t)$, *such that*

i) $\boldsymbol{A}(t) \supset \boldsymbol{I}$ *almost* (τ) *everywhere on* Y, $\tau = \sum \mu_i T^{-1}/n$,

ii) $\boldsymbol{L}_\mu = \boldsymbol{I} \cdot \mu = \int \boldsymbol{A}(t) d\mu T^{-1}(t)$,

iii) *for every test function (=$\mathfrak{B}$-measurable non-negative function bounded by 1)* $\varphi(x)$ *on* X, *there exists a vector valued function* $Q(\varphi; t) = (Q_1(\varphi; t), \cdots, Q_n(\varphi; t))$, *which belongs to* $\boldsymbol{A}(t)$ *almost* (τ) *everywhere on* Y, *and each component of which satisfies the equation*

$$\int_{T^{-1}F} \varphi(x) d\mu_i = \int_F Q_i(\varphi; t) d\mu_i T^{-1} \quad \textit{for every} \quad F \in \mathfrak{C}, \quad \textit{and} \quad i = 1, \cdots, n.$$

The following theorem is our principal aim.

Theorem IV. *For every experiment* $\mu = (\mu_1, \cdots, \mu_n)$ *and every statistic* $T(x)$, *we have*

$$\boldsymbol{L}_{\mu T^{-1}} \subseteqq \boldsymbol{L}_\mu\ .$$

The equality holds when and only when $T(x)$ *is a sufficient statistic, that is, every vector* $Q(\varphi; t)$ *for test function* $\varphi(x)$ *is such that* $Q_1(\varphi; t) = Q_2(\varphi; t) = \cdots = Q_n(\varphi; t)$ *almost* (τ) *everywhere on* Y.

Proof. The first part of our theorem is obvious from the fact that $\mathfrak{B} \supset T^{-1}\mathfrak{C}$.

Theorem III and $\boldsymbol{L}_{\mu T^{-1}} = \boldsymbol{L}_\mu$ show that

$$\boldsymbol{I}\cdot\mu T^{-1}=\boldsymbol{L}_{\mu T^{-1}}=\boldsymbol{L}_{\mu}=\int \boldsymbol{A}(t)d\mu T^{-1}\supseteqq \boldsymbol{I}\cdot\mu T^{-1}\ ,$$

that is,

$$\int \boldsymbol{A}(t)d\mu T^{-1}=\boldsymbol{I}\cdot\mu T^{-1}\ .$$

Therefore, by Theorem II, we have

$$\boldsymbol{A}(t)=\boldsymbol{I}$$

almost (τ) everywhere on Y, and hence $Q(\varphi;t)\in \boldsymbol{A}(t)=\boldsymbol{I}$.

Conversely if $Q(\varphi;t)\in \boldsymbol{I}$ almost (τ) everywhere on Y, it holds from the definition of $\boldsymbol{A}(t)$ that $\boldsymbol{A}(t)=\boldsymbol{I}$. Therefore

$$\boldsymbol{I}\cdot\mu=\int \boldsymbol{A}(t)d\mu T^{-1}=\boldsymbol{I}\cdot\mu T^{-1}=\boldsymbol{L}_{\mu T^{-1}}\ .$$

This completes the proof.

Example 1. $(X,\mathfrak{B})$ is the direct product space of two measurable spaces $(Y,\mathfrak{C})$, and $(Z,\mathfrak{D})$, and $\mu=(\mu_1,\cdots,\mu_n)$ is an experiment on $(X,\mathfrak{B})$ such that

$$\mu_i(F\times G)=\int_F \sigma_i(G;t)d\nu_i(t)\quad \text{for}\quad F\in\mathfrak{C}\quad\text{and}\quad G\in\mathfrak{D}\ ,$$

where $\sigma_i(G;t)$ is a function such that (1) for fixed t, σ_i is a probability measure over $(Z;\mathfrak{D})$ and (2) for fixed G, σ_i is a $\mathfrak{C}$-measurable function. If the measures ν_i are absolutely continuous each other, the set

$$\begin{aligned}\boldsymbol{L}_{\sigma(\cdot,t)}&=\left\{\left(\int\varphi(s,t)\sigma_1(ds;t),\cdots,\int\varphi(s,t)\sigma_n(ds;t)\right);\ \varphi\in\Delta(\boldsymbol{I})\right\}\\&=\left\{\left(\int\varphi(s)\sigma_1(ds;t),\cdots,\int\varphi(s)\sigma_n(ds;t)\right);\ \varphi\in\Delta(\boldsymbol{I})\right\}\end{aligned}$$

coinsides with the convex body valued fuction $\boldsymbol{A}(t)$ almost (τ) everywhere on Y.

Example 2. The statistic

$$T_0(x)=f(x)=(f_1(x),\cdots,f_n(x))$$

is sufficient with respect to $\mu=(\mu_1,\cdots.\mu_n)$: $\mu_i=\int f_i d\lambda$, $i=1,\cdots,n$, because the experiment of $T_0(x)$ has the components

$$\mu_i T_0^{-1}(E)=\int_E f_i d\lambda T_0^{-1}(f)\quad\text{for every Borel set}\quad E\subset R^n,\quad i=1,\cdots,n\ .$$

Therefore the supporting function of $\boldsymbol{L}_{\mu T^{-1}}$ is, for the supporting function H_0 of $\boldsymbol{I}$,

$$\int H_0(\xi\times f)d\mu T_0^{-1}(f)=\int H_0(\xi\times f(x))d\mu(x)\ ,$$

which shows that $\boldsymbol{L}_{\mu}=\boldsymbol{L}_{\mu T_0^{-1}}$.

This experiment of T_0 is called, by Blackwell [4] and [5], the *standard experiment* of μ.

Remark 1. If it is true that $L_\mu = L_\nu$ implies $A\cdot\mu = A\cdot\nu$ for all convex bodies A, and if both μ and $\mu^{T^{-1}}$ are standard experiments, then our Theorem IV is a direct implication of Blackwell's theorem (Theorem 3 and 8, [5]).

Remark 2. We shall restrict ourselves to the case where $n=2$ and μ_1, μ_2 are absolutely continuous one another. In our previous paper [9], we have introduced a function, named the information generating function,

$$\rho(a; L_\mu) = \int \left(\frac{f_2}{f_1}\right)^a d\mu_1 , \qquad 0 \leq a \leq 1 ,$$

and investigated the relations of the experiment and the mean information

$$I(L_\mu) = -\frac{d}{da}\rho(a; L_\mu)|_{a=0} \quad \text{(if exists)}$$

defined by Kullback and Leibler [10], where $\rho(a; L_\mu)$ and hence $I(L_\mu)$ do not depend on the measures μ_1 and μ_2, but only on the convex set L_μ. If

$$X = Y \times Z, \quad \mu_i(E) = \int \sigma_i(E; t) d\nu_i \ \text{ for } E \in \mathfrak{B}, \quad \sigma_2(E; t) = \int h(x; t) d\sigma_1(x; t),$$

and $\nu_2(F) = \int_F g(t) d\nu_1$ for $F \in \mathfrak{C}$, then we have, by Example 1,

1) $L_\mu = \int A(t) d\nu, \quad L_{\sigma(\cdot\, t)} = A(t)$,

2) $\rho\left(a; \int A(t) d\nu\right) = \int \rho(a; A(t))\, (g(t))^a d\nu_1(t)$,

3) $I\left(\int A(t) d\nu\right) = \int I(A(t)) d\nu_1(t) + I(L_\nu)$.

Therefore we have

$$\int I(A(t)) d\nu_1(t) = I\left(\int A(t)\, d\nu\right) - I(L_\nu) .$$

This is an amount of information of the experiment μ, which is left after the experiment ν has accomplished. For example, if Y and Z are finite sets and if $\mu_1(t, s) = P_{ts} > 0$, $\mu_2(t, s) = \frac{1}{kl}$, $t = 1, \cdots, k$, $s = 1, \cdots, l$, then we have

$$\int I(A(t)) d\nu(t) - \log l = -\sum_{t,s} P_{ts} \log \frac{P_{ts}}{\sum_s P_{ts}} .$$

The right hand side of the above equation is called the *conditional entropy* $H_t(s)$ of s in Shannon's book [11].

References

[1] P. R. Halmos and L. J. Savage, Application of the Radon-Nikodym theorem to the theory of sufficient statistics, *Annals of Math. Stat.*, **20** (1949).

[2] E. L. Lehmann, Some principles of the theory of testing hypotheses, *Annals of Math. Stat.*, **21** (1950)

[3] E. L. Lehmann and H. Scheffé, Completeness, similar regions, and unbiased estimation, Part I, *Sankya*, **10** (1950).

[4] D. Blackwell, Comparison of experiments, *Proceedings of the Second Berkley Symposium on Mathematical Statistics and Probability*, Univ. of California Press, 1951.

[5] D. Blackwell, Equivalent comparisons of experiments, *Annals of Math. Stat.*, **24** (1953).

[6] H. Kudō, A theorem of Kakutani on infinite product measures, *Natural Science Report of the Ochanomizu University*, **3** (1952).

[7] T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Ergebnisse der Mathematik und ihre Grenzgebiete, **1**, Julius Springer, 1934.

[8] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und ihre Grenzgebiete, **3**, Julius Springer, 1934.

[9] H. Kudō, Experiments and their informations, *Theories and Applications of Time Series and Informations*, Union of the Japanese Scientists and Engineers Press, 1953 (in Japanese).

[10] S. Kullback and B. A. Leibler, On Information and Sufficiency, *Annals of Math. Stat.*, **22** (1951).

[11] C. E. Shannon and W. Weaver, *The mathematical theory of communication*, Univ. Illinois Press, 1949.

(*Received November 26, 1953*)