J. Korean Math. Soc.
Vol. 9, No. 2, 1972

ON A CRITERION FOR OBTAINING FULL INFORMATION ABOUT THE UNKNOWN STATE OF NATURE

BY HYUN-CHUN SHIN AND JAE-JOO KIM

§0. Introduction

Consider a situation in which it is desired to gain knowlege about the true value of the unknown state of the nature by means of observations. Information concerning the unknown state of the nature is defined by as a random variable whose (objective) probability law is known given any state of the nature which is an element of a fixed state space S. [6], [7].

Information amount of an information is defined by [2], [4], [5], as the expected difference between the entropy of the prior distribution over S and the entropy of the posterior distribution. If an information becomes available to a decision maker for solving a specific decision problem, then the loss function in the specific decision problem at hand.

The problem which will be discussed in what follows is: On what conditions can one decide whether or not a certain sequence of observations contains all the information which is needed (for example, to find the true value of the state or the parameter)?

The objective of the paper to try to answer the above question in the case $S=\{s_1, s_2, \cdots, s_m\}$.

In section 1 we shall discuss from the point of view of the question the following problem. Let us be given an infinite sequence $\{\tilde{X}_n\}$, $(n=1, 2, \cdots)$ of observations. We suppose that the distributions of the random variables $\tilde{X}_i$ $(i=1, 2, \cdots)$ depend on a state $\tilde{S}$, whose set of possible values is finite. We suppose further that for each fixed value of $\tilde{S}$ the random variables are independent. We shall consider the amount of information $\tilde{S}$ which is still missing after having observed the values $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n$ and compare it with the error of the "standard" decision, consisting in deciding always in favor of the hypothesis which has the largest posterior probability. And we give an upper bound for the amount of missing information

In section 2 we give a neccessary and sufficient condition for obtaining full information.

§1. The amount of missing information and the error of standard decision

Let $(S, \boldsymbol{S})$ be a measurable state space, ξ be a prior probability measure on $(S, \boldsymbol{S})$. Let ξ be a absolutely continuous with respect to a measure λ on $(S, \boldsymbol{S})$ and let $d\xi=\xi(s)\,d\lambda$. Then the uncertainty measure $H(\xi)$ of the unknown state $s\in S$ is defined by

Received by the editors Oct. 31, 1972

(1.1) $$H(\xi) = -\int \xi(s) \log \xi(s)\, d\lambda.$$

Specifically, if $S = \{s_1, s_2, \cdots, s_m\}$ and $\xi = (\xi(s_1), \xi(s_2), \cdots, \xi(s_m)) = (\xi_1, \xi_2, \cdots, \xi_m)$, where $\xi_i \geqq 0$, $\sum_{i=1}^{m} \xi_i = 1$, then it is

(1.2) $$H(\xi) = -\sum_{i=1}^{m} \xi_i \log \xi_i.$$

Let us denote by $\tilde{X}(n)$ the random n-dimensional vector with components $(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n)$. Let $\tilde{X}(n)$ be a random variable on a measurable space $(X(n), \boldsymbol{X}(n))$ whose probability density function $f(x(n)\,|\,s)$ given $s \in S$ with respect to a measure μ on $(X(n), \boldsymbol{X}(n))$ is assumed to be known. We suppose that the random variables $\tilde{X}_i$ $(i=1, \cdots, n)$ are independent under the condition that $\tilde{S} = s_i$ $(i=1, \cdots, m)$ is given. Let $f_1(x(n))$, $f_2(x(n))$, $\cdots$ and $f_m(x(n))$ denote the density function for $\tilde{S} = s_1$, $\tilde{S} = s_2$, $\cdots$ and $\tilde{S} = s_m$, respectively.

We suppose further (this restriction is made only to simplify notations) that all the distributions in question are absolutely continuous.

Let $(A, \boldsymbol{A})$ be a measurable action space and $W(\cdot, \cdot)$ be a measurable loss function defined on $(S \times A, \boldsymbol{S} \times \boldsymbol{A})$. Then we shall say that these elemets $(S, \boldsymbol{S})$, ξ, $(A, \boldsymbol{A})$ and W specify a ***basic decision problem*** D_0 and denote it as $D_0 \equiv \{S, \xi, A, W\}$.

If a decision maker can know the realized value $x(n)$ of $\tilde{X}(n)$, then we shall say an *information* $e(\tilde{X}(n))$ is available to a decision maker for solving a decision problem D_0. And then we shall say that he has a decision problem $D \equiv \{D_0 : e(\tilde{X}(n))\}$.

It is needless to say that an information $e(\tilde{X}(n))$ for $s \in S$ is defined independently of ξ and W. After observing $\tilde{X}(\mathrm{n}) = x(n)$, by Bayes' theorem, the posterior probility law $\xi(s\,|\,x(n))$ is given by

(1.3) $$\xi(s\,|\,x(n)) = \xi(s) f(x(n)\,|\,s) / f(x(n)), \quad \text{where } f(x(n)) = \int \xi(s) f(x(n)\,|\,s)\, d\lambda.$$

For simplicity, we shall denote the posterior probability measure given $\tilde{X}(\mathrm{n}) = x(n)$ by $\xi(x(n))$, when the prior probability meature is ξ. Specifically, if $S = \{s_1, s_2, \cdots, s_m\}$ and $\xi = (\xi_1, \xi_2 .., \xi_m) \in S^{m-1}$, then $\xi(x(n)) = (\xi(s_1\,|\,x(n)), \xi(s_2\,|\,x(n)), \cdots, \xi(s_m\,|\,x(n)) = (\xi_1(x(n)), \xi_2(x(n)), \cdots, \xi_m(x(n))) \in S^{m-1}$ where $S^{m-1} \equiv \{(r_1, r_2, \cdots, r_m) ; r_i \geqq 0, \sum_{i=1}^{s} r_i = 1\}$.

Then the still remaining uncertainty measure after observing $\tilde{X}(n) = x(n)$ is given dy $H(\xi(x(n))$ and its expected value

(1.4) $$M(\tilde{X}(n)\,|\,\xi) \equiv E[H(\xi(\tilde{X}(n)))] = \int_{X(n)} H(\xi(x(n)) f(x(n))\, d\mu$$

is called the ***amount of missing information after observing*** $e(\tilde{X}(n))$ ***by*** Rényi [8], [9], or the ***equivocation of*** $e(\tilde{X}(n))$.

The information amount $I(\tilde{X}(n)\,|\,\xi)$ which an information $e(\tilde{X}(n))$ provides is defined by

(1.5) $$I(\tilde{X}(n)\,|\,\xi) \equiv H(\xi) - M(\tilde{X}(n)\,|\,\xi).$$

It is well known that $H(\xi)$ and $M(\tilde{X}(n)\,|\,\xi)$ are concave function $\xi \in S^{m-1}$ and $\xi(x(n))$, $\in S^{m-1}$ respectively, and that $I(\tilde{X}(n)\,|\,\xi) \geqq 0$ [2].

Let us introduce the following decision rule. The most natural decision after having observed $\tilde{X}(n)=x(n)$ is essentially the same as the reasonning applied in the case $S=\{s_1, s_2\}$, the case of two simple hypothesis [9].

It states that $s_1, s_2, \cdots$ or s_m is accepted according as $\xi_1 f_1(x(n))$, $\xi_2 f_2(x(n)), \cdots$ or $\xi_m f_m(x(n))$ is greatest, the greatest posterior probability, and if $\xi_1 f_1(x(n))=\xi_2 f_2(x(n))=\cdots=\xi_m f_m(x(n))$, one makes a random choice among $s_1, s_2 \cdots$ or s_m with probabilities $\xi_1, \xi_2, \cdots$ or ξ_m, respectively. We shall call this the *standard decision*. Let us define the random variable $\Phi_n=\Phi(\tilde{X}(n))$ as follows:

(1.6) $\Phi_n=s_i$ if the standard decision means acceptance of s_i $(i=1,.,m)$.

We adopt the common convention of defining $W(\tilde{S}, \Phi_n)$ such that $W=0$ when the correct decision is made, and 1 otherwise. *The error ε_n of the standard decision* after taking n observations is defined as the probability of the standard decision being false. We have clearly

$$
\begin{aligned}
\varepsilon_n = P_r(\Phi_n \neq \tilde{S}) &= \xi_1 [P_r(\Phi_n=s_2 | \tilde{S}=s_1) + P_r(\Phi_n=s_3 | \tilde{S}=s_1) + \cdots \\
&+ P_r(\Phi_n=s_m | \tilde{S}=s_1)] + \xi_2 [P_r(\Phi_n=s_1 | \tilde{S}=s_2) + P_r(\Phi_n=s_3 | \tilde{S}=s_2) + \cdots \\
&+ P_r(\Phi_n=s_m | \tilde{S}=s_2)] + \cdots + \xi_m [P_r(\Phi_n=s_1 | \tilde{S}=s_m) + P_r(\Phi_n=s_2 | \tilde{S}=s_m) + \cdots \\
&+ P_r(\Phi_n=s_{m-1} | \tilde{S}=s_m)].
\end{aligned} \tag{1.7}
$$

In a decision problem, if D is available to a decision maker, then we devide the sample space into the disjoint acceptance regions, $X_{(1)}, X_{(2)}, \cdots$, and $X_{(m)}$ such that $\Phi_n=s_j$ is accepted when $x(n)\in X_{(j)}, j=1,2,\cdots,m$. With this specification we have (see [10])

$$
\begin{aligned}
\varepsilon_n = &\xi_1 \left(\int_{X_{(2)}} f_1(x(n))\, d\mu + \int_{X_{(3)}} f_1(x(n))\, d\mu + \cdots + \int_{X_{(m)}} f_1(x(n))\, d\mu\right) \\
&+ \xi_2 \left(\int_{X_{(1)}} f_2(x(n))\, d\mu + \int_{X_{(3)}} f_2(x(n))\, d\mu + \cdots + \int_{X_{(m)}} f_2(x(n))\, d\mu\right) + \cdots \\
&+ \xi_m \left(\int_{X_{(1)}} f_m(x(n))\, d\mu + \int_{X_{(2)}} f_m(x(n))\, d\mu + \cdots + \int_{X_{(m-1)}} f_m(x(n))\, d\mu\right)
\end{aligned} \tag{1.8}
$$

where

$$
X_{(1)} = \left\{x(n) : \frac{f_1(x(n))}{f_2(x(n))} \geq \frac{\xi_2}{\xi_1},\ \frac{f_1(x(n))}{f_3(x(n))} \geq \frac{\xi_3}{\xi_1}, \cdots \text{ and } \frac{f_1(x(n))}{f_m(x(n))} \geq \frac{\xi_m}{\xi_1}\right\}, \tag{1.9}
$$

$$
X_{(2)} = \left\{x(n) : \frac{f_2(x(n))}{f_1(x(n))} > \frac{\xi_1}{\xi_2},\ \frac{f_2(x(n))}{f_3(x(n))} \geq \frac{\xi_3}{\xi_2}, \cdots \text{ and } \frac{f_2(x(n))}{f_m(x(n))} \geq \frac{\xi_m}{\xi_2}\right\}, \tag{1.10}
$$

$\cdots$

and

$$
X_{(m)} = \left\{x(n) : \frac{f_m(x(n))}{f_1(x(n))} > \frac{\xi_1}{\xi_m},\ \frac{f_m(x(n))}{f_2(x(n))} > \frac{\xi_2}{\xi_m}, \cdots \text{ and } \frac{f_m(x(n))}{f_{m-1}(x(n))} > \frac{\xi_{m-1}}{\xi_m}\right\}. \tag{1.11}
$$

The equations (1.9), (1.10) and (1.11) correspond to

$$
X_{(1)} = \left\{x(n) : \frac{\xi_1}{\xi_2} \cdot \frac{f_1(x(n))}{f_2(x(n))} \geq 1,\ \frac{\xi_1}{\xi_3} \cdot \frac{f_1(x(n))}{f_3(x(n))} \geq 1, \cdots \text{and} \right. \tag{1.12}
$$

$$X_{(2)}=\left\{x(n):\frac{\xi_1}{\xi_2}\cdot\frac{f_1(x(n))}{f_2(x(n))}<1,\ \frac{\xi_2}{\xi_3}\cdot\frac{f_2(x(n))}{f_3(x(n))}\geq 1,\cdots \text{ and } \frac{\xi_2}{\xi_m}\cdot\frac{f_2(x(n))}{f_m(x(n))}\geq 1\right\}, \tag{1.13}$$

$$\frac{\xi_1}{\xi_m}\cdot\frac{f_1(x(n))}{f_m(x(n))}\geq 1\Big\},$$

and

$$X_{(m)}=\left\{x(n):\frac{\xi_1}{\xi_m}\cdot\frac{f_1(x(n))}{f_m(x(n))}<1,\ \frac{\xi_2}{\xi_m}\cdot\frac{f_2(x(n))}{f_m(x(n))}<1,\cdots \text{ and } \frac{\xi_{m-1}}{\xi_m}\cdot\frac{f_{m-1}(x(n))}{f_m(x(n))}<1\right\}. \tag{1.14}$$

We can prove the following theorem.

THEOREM 1.1. *One has*

$$\varepsilon_n\leq E[H_2(\xi(\tilde{X}(n)))]=M_2(\tilde{X}(n)\,|\,\xi) \tag{1.15}$$

where $_2$ *denotes logarithm with base* 2.

Here and in what follows log always denotes logarithm with base 2.

Proof. For simplicity, we shall denote $f_1(x(n)), f_2(x(n)), \cdots, f_m(x(n))$ and $f(x(n))$ as follows:

$$f_1\equiv f_1(x(n)),\ f_2\equiv f_2(x(n)),\ \cdots,\ f_m\equiv f_m(x(n)) \text{ and } f\equiv f(x(n))=\xi_1 f_1+\xi_2 f_2+\cdots+\xi_m f_m. \tag{1.16}$$

One has clearly

$$\begin{aligned} E(H(\xi\tilde{X}(n)))&=\int_{X(n)}H(\xi(x(n)))\,(\xi_1 f_1+\xi_2 f_2+\cdots+\xi_m f_m)\,d\mu \\ &=\xi_1\int_{X(n)}H(\xi(x(n)))f_1 d\mu+\xi_2\int_{X(n)}H(\xi(x(n)))f_2 d\mu \\ &\quad+\cdots+\xi_m\int_{X(n)}H(\xi(x(n)))f_m d\mu. \end{aligned} \tag{1.17}$$

By the definition of $H(\xi)$, the r.h.s. term in (1.17) is

$$\begin{aligned} &\xi_1\int_{X(n)}\Big[\frac{\xi_1 f_1}{\sum_{i=1}^m \xi_i f_i}\log\Big(1+\frac{\xi_2 f_2}{\xi_1 f_1}+\frac{\xi_3 f_3}{\xi_1 f_1}+\cdots+\frac{\xi_m f_m}{\xi_1 f_1}\Big) \\ &+\frac{\xi_2 f_2}{\sum_{i=1}^m \xi_i f_i}\log\Big(1+\frac{\xi_1 f_1}{\xi_2 f_2}+\frac{\xi_3 f_3}{\xi_2 f_2}+\cdots+\frac{\xi_m f_m}{\xi_2 f_2}\Big)+\cdots \\ &+\frac{\xi_m f_m}{\sum_{i=1}^m \xi_i f_i}\log\Big(1+\frac{\xi_1 f_1}{\xi_m f_m}+\frac{\xi_2 f_2}{\xi_m f_m}+\cdots+\frac{\xi_{m-1} f_{m-1}}{\xi_m f_m}\Big)\Big]f_1 d\mu \end{aligned} \tag{1.18}$$

$$+\xi_2\int_{X(n)}[*]f_2d\mu+\cdots+\xi_m\int_{X(n)}[*]f_md\mu$$

$$=\xi_1\int_{X(n)}f_1\log\left(1+\frac{\xi_2f_2}{\xi_1f_1}+\cdots+\frac{\xi_mf_m}{\xi_1f_1}\right)d\mu$$

(1.19)
$$+\xi_2\int_{X(n)}f_2\log\left(1+\frac{\xi_1f_1}{\xi_2f_2}+\cdots+\frac{\xi_mf_m}{\xi_2f_2}\right)d\mu+\cdots$$

$$+\xi_m\int_{X(n)}f_m\log\left(1+\frac{\xi_1f_1}{\xi_mf_m}+\frac{\xi_2f_2}{\xi_mf_m}+\cdots+\frac{\xi_{m-1}f_{m-1}}{\xi_mf_m}\right)d\mu.$$

By (1.12), (1.13) and (1.14), the terms in (1.19) are not less than followings:

$$\geqslant\xi_1\left(\int_{X(2)}f_1d\mu+\int_{X(3)}f_1d\mu+\cdots+\int_{X(m)}f_1d\mu\right)$$

(1.20)
$$+\xi_2\left(\int_{X(2)}f_2d\mu+\int_{X(3)}f_2d\mu+\cdots+\int_{X(m)}d\mu\right)+\cdots$$

$$+\xi_m\left(\int_{X(1)}f_md\mu+\int_{X(2)}f_md\mu+\cdots+\int_{X(m-1)}f_md\mu\right)=\varepsilon_n.$$

Thus the theorem is proved.

REMARK 1.1. Rényi's Theorem 3 in [8] is special case of our Theorem 1.1 where $\tilde{X}_i\,(i=1,\cdots,n)$ are independent and identically distributed under the condition that $\tilde{S}=s_i\,(i=1,2)$ is given.

THEOREM 1.2. *Let us write*

(1.21)
$$\lambda_{ij}^{(r)}=\int_{-\infty}^{\infty}\sqrt{f_i(x_r)f_j(x_r)}\,dx_r,\quad r=1,2,\cdots n,\quad i\neq j=1,2,\cdots,m.$$

Then the following inequality holds:

(1.22)
$$0\leqslant E[H(\xi(\tilde{X}(n)))]\leqslant C\sum_{i\neq j=1}^{m}\sqrt{\xi_i\cdot\xi_j}\prod_{r=1}^{n}\lambda_{ij}^{(r)}$$

where

(1.23)
$$C=\max_{0\leqslant x\leqslant 1}\frac{h(x)}{\sqrt{x}}\left(=\max_{0\leqslant x\leqslant 1}\frac{h(x)}{\sqrt{1-x}}\right)$$

and

(1.24)
$$h(x)=-x\log x-(1-x)\log(1-x)\qquad \textit{for } 0\leqslant x\leqslant 1.$$

Proof. Since $f=\xi_1f_1+\xi_2f_2+\cdots+\xi_mf_m$ by (1.16), we have

$$E[H(\xi(\tilde{X}(n)))]=\int_{X(n)}H(\xi(x(n)))\,(\xi_1f_1+\cdots+\xi_mf_m)\,d\mu$$

(1.25)
$$=\xi_1\int_{X(n)}H(\xi(x(n)))f_1d\mu+\xi_2\int_{X(n)}H(\xi(x(n)))f_2d\mu$$

The notation [*] denotes the first term in [ ] of (1.18).

$$+\cdots+\xi_m\int_{X(n)} H(\xi(x)n)))f_m d\mu.$$

By the grouping axiom of entropy [1], the first r.h.s. term in (1.25) is

$$\xi_1\int_{X(n)} H(\xi(x(n)))f_1 d\mu=\xi_1\int_{X(n)} [H(\xi_1(x(n))+\xi_2(x)n))+\cdots+\xi_{m-1}(x(n)),\ \xi_m(x(n)))$$

$$+\left(\sum_{i=1}^{m-1}\xi_i(x(n))\right)H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \cdots,\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right) \tag{1.26}$$

$$+\xi_m(x(n))H(1)]f_1 d\mu$$

$$=\xi_1\int_{X(n)}[H(\sum_{i=1}^{m-1}\xi_i(x(n)),\ \xi_m(x(n))]f_1 d\mu$$

$$+\xi_1\int_{X(n)}[(\sum_{i=1}^{m-1}\xi_i(x(n)))H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \cdots,\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)]f_1 d\mu \tag{1.27}$$

Since $\sum_{i=1}^{m-1}\xi_i(x(n))\leqslant 1$, the r.h.s. terms in (1.27) are less than following:

$$\leqslant\xi_1\int_{X(n)}[H(\sum_{i=1}^{m-1}\xi_i(x(n)),\ \xi_m(x(n))]f_1 d\mu$$

$$+\xi_1\int_{X(n)}[H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \cdots,\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)]f_1 d\mu. \tag{1.28}$$

From the definition of C, it follows that

$$H(\sum_{i=1}^{m-1}\xi_i(x(n)),\ \xi_m(x(n))\leqslant C[\xi_m(x(n))]^{\frac{1}{2}}$$

$$=C\left(\frac{\xi_m f_m}{\sum_{i=1}^{m}\xi_i f_i}\right)^{\frac{1}{2}}\leqslant C\left(\frac{\xi_m}{\xi_1}\right)^{\frac{1}{2}}\left(\frac{f_m}{f_1}\right)^{\frac{1}{2}}. \tag{1.29}$$

And again by the grouping axiom of entropy, we have following:

$$H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \cdots,\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)$$

$$=H\left(\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right) \tag{1.30}$$

$$+\left(\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))},\ \frac{\xi_2(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))},\ \cdots,\ \frac{\xi_{m-2}(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}\right)$$

$$+\frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}H(1).$$

Since $\dfrac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\leqslant 1$ and $H(1)=0$, the r. h. s. terms in (1.30) are less than following:

$$\begin{aligned}&\leqslant H\left(\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)\\ &+H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))},\ \frac{\xi_2(x(n))}{\sum_{i=1}^{m-2}\xi_1(x(n))},\cdots,\ \frac{\xi_{m-2}(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}\right).\end{aligned}\tag{1.31}$$

From the definition of C, it also follows that

$$\begin{aligned}H\left(\frac{\sum_{i=1}^{m-2}\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\ \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)&\leqslant C\left(\frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)^{\frac{1}{2}}\\ &=C\left(\frac{\xi_{m-1}f_{m-1}}{\sum_{i=1}^{m-1}\xi_if_i}\right)^{\frac{1}{2}}\\ &\leqslant C\left(\frac{\xi_{m-1}}{\xi_1}\right)^{\frac{1}{2}}\left(\frac{f_{m-1}}{f_1}\right)^{\frac{1}{2}}.\end{aligned}\tag{1.32}$$

We procede this process succesively, it follows that

$$\begin{aligned}\xi_1\int_{X(n)}H(\xi(x(n)))f_1d\mu&\leqslant C\sum_{j=2}^{m}\sqrt{\xi_1\cdot\xi_j}\int_{X(n)}\sqrt{f_1\cdot f_j}\,d\mu\\ &=C\sum_{j=2}^{m}\sqrt{\xi_1\cdot\xi_j}\prod_{r=1}^{n}\lambda_{1j}{}^{(r)}\end{aligned}\tag{1.33}$$

By a similar method, it follows that

$$\xi_2\int_{X(n)}H(\xi(x(n)))f_2d\mu\leqslant C\sqrt{\xi_1\cdot\xi_2}\prod_{r=1}^{n}\lambda_{12}{}^{(r)}+C\sum_{j=3}^{m}\sqrt{\xi_2\cdot\xi_j}\prod_{r=1}^{n}\lambda_{2j}{}^{(r)}\tag{1.34}$$

and

$$\xi_m\int_{X(n)}H(\xi(x(n))f_md\mu\leqslant C\sum_{j=1}^{m-1}\sqrt{\xi_m\cdot\xi_j}\prod_{r=1}^{n}\lambda_{mj}{}^{(r)}\tag{1.35}$$

Thus we have

$$E[H(\xi(\tilde{X}(n)))]\leqslant C\sum_{i\neq j=1}^{m}\sqrt{\xi_i\cdot\xi_j}\prod_{r=1}^{n}\lambda_{ij}{}^{(r)}\tag{1.36}$$

The proof of the lower inequality of the theorem is follows from the theorem 1.1. Q. E. D.

REMARK 1.2. Suppose that the random variables $\tilde{X}_i\,(i=1,2,\cdots,n)$ are independent and identically distributed under the condition $\tilde{S}=s_i\,(i=1,2,\cdots,m)$ is given. Let us write

$$\lambda_{ij}{}^{(1)}=\lambda_{ij}{}^{(2)}=\cdots=\lambda_{ij}{}^{(n)}=\beta_{ij}.\tag{1.37}$$

Then from (1.36), we have

$$0\leqslant E[H(\xi(\tilde{X}(n)))]\leqslant C\sum_{i\neq j=1}^{m}\sqrt{\xi_1\cdot\xi_j}\,(\beta_{ij})^n.\tag{1.38}$$

If $0\leqslant\beta_{ij}<1$ for all i and j, then the relation (1.38) shows how fast the equivocation

$E[H(\xi(\tilde{X}(n))]$ of an information $e(\tilde{X}(n))$ approach to zero as n increases, irrespective of what a prior probability law is. We discuss this point in detail in the next section.

§2. A criterion for obtaining full information.

It this section we shall suppose that for each fixed value of $\tilde{S}$ the random variables $\tilde{X}_i$, $(i=1,2,\cdots)$ are independent, but in general do not have the same distribution.

Let us deneote by $\tilde{X}(n)$ the random n-dimensional vector with components $(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n)$. Let I_n denote the amount of information contained $\tilde{X}(n)$ concerning $\tilde{S}$.

Then we have

$$I_n = H(\xi) - E[H(\xi(\tilde{X}(n)))] \tag{2.1}$$

where $H(\xi)$ is defined by (1.2).

It is easy to see that I_n is nondecreasing for $n=1,2,\cdots$ and $I_n \leqq H(\xi)$. Thus $\lim_{n\to+\infty} I_n = I^*$ always exists. If $I^*=H(\xi)$, we shall say that the sequence of observations $\{\tilde{X}_i\}$, $(i=1,2,\cdots)$ give us *full information on* $\tilde{S}$, where as in the case $I^*<H(\xi)$ we shall say that the observations $\{\tilde{X}_i\}$, $(i=1,2,\cdots)$ do not give full information on $\tilde{S}$. Rényi [9] introduced a criterion for obtaining full information in the case $S=\{s_1, s_2\}$:

THEOREM. 2.1. *If* $\lambda_{12}^{(r)}>0$ *for* $r=1.2,\cdots$, *where*

$$\lambda_{12}^{(r)} = \int_{-\infty}^{\infty} \sqrt{f_1(x_r) f_2(x_r)}\, dx_r, \tag{2.2}$$

then the sequence of observations $\tilde{X}_i$, $(i=1,2,\cdots)$ *contains full information on* $\tilde{S}$ *if and only if the series*

$$\sum_{r=1}^{\infty} (1-\lambda_{12}^{(r)}) \tag{2.3}$$

is divergent.

LEMMA 2.1. *One has*

$$\prod_{r=1}^{n} \lambda_{ij}^{(r)} \leqq (\varepsilon_n)^{\frac{1}{2}} \frac{\sqrt{\xi_i}+\sqrt{\xi_j}}{\sqrt{\xi_i \cdot \xi_j}} \quad \textit{for all } i \textit{ and } j \ (i\neq j=1,2,\cdots,m) \tag{2.4}$$

where $\lambda_{ij}^{(r)}$ *and* ε_n *is defined by* (1.21) *and* (1.8) *respectively.*

Proof of Lemma 2.1. Clearly,

$$\prod_{r=1}^{n} \lambda_{ij}^{(r)} = \int_{X(n)} \sqrt{f_i(x(n)) f_j(x(n))}\ dx(n) \quad \textit{for } i\neq j=1,2,\cdots,m \tag{2.5}$$

where $X(n)$ is the n-dimensiooal Euclidian space and $dx(n)$ stands for $dx_1 \cdot dx_2 \cdots dx_n$. Let us denote again by $X_{(i)}$ the subset of $X(n)$ on which $\phi_n=s_i$ and put $\bar{X}_{(i)}=X(n)-X_{(i)}$. Taking into account $f_i(x(n))$ is a density function, the Cauchy-Schwarz inequality gives

$$\int_{X_{(i)}} \sqrt{f_i(x(n)) . f_j(x(n))}\ \ dx(n) \leqq \left(\int_{X_{(i)}} f_j(x(n))\, dx(n)\right)^{\frac{1}{2}} \tag{2, 6}$$

and

$$\int_{\bar{X}_{(i)}} \sqrt{f_i(x(n)) f_j(x(n))}\, dx(n) \leqq \left(\int_{\bar{X}_{(i)}} f_i(x(n)\, dx(n)\right)^{\frac{1}{2}} \tag{2. 7}$$

since

$$\varepsilon_n = \sum_{i \neq j=1}^{m} \xi_j \int_{X(i)} f_j(x(n))\,dx(n). \tag{2.8}$$

we have

$$\frac{\varepsilon_n}{\xi_j} \geqslant \int_{X(i)} f_j(x(n))\,dx(n) \text{ and } \frac{\varepsilon_n}{\xi_i} \geqslant \int_{X(j)} f_i(x(n))\,dx(n). \tag{2.9}$$

Therefore we have

$$\int_{X(n)} \sqrt{f_i(x(n)) f_j(x(n))}\,dx(n) \leqslant \left(\frac{\varepsilon_n}{\xi_j}\right)^{\frac{1}{2}} + \left(\frac{\varepsilon_n}{\xi_i}\right)^{\frac{1}{2}} = \frac{(\varepsilon_n)^{\frac{1}{2}}(\sqrt{\xi_i}+\sqrt{\xi_j})}{\sqrt{\xi_i \cdot \xi_j}} \tag{2.10}$$

This proves Lemma 2.1.

THEOREM 2.2. *If* $\lambda_{ij}^{(r)} > 0$ *for* $r=1,2,\cdots$ *and* $j>i=1,\cdots,m$, *where* $\lambda_{ij}^{(r)}$ *is defined by* (1.21), *the sequence of observations* $\tilde{X}_i$ $(i=1,2\cdots)$ *contains full information on* $\tilde{S}$ *if and only if*

$$\sum_{r=1}^{\infty} (1-\lambda_{ij}^{(r)}) \quad (j>i=1,2,\cdots,m) \tag{2.11}$$

are divergent for all i *and* j.

Proof. since $1-x \leqslant e^{-x}$, if the series $\sum_{r=1}^{\infty} (1-\lambda_{ij}^{(r)})$ are divergent for all i and j, one has $\lim_{n\to\infty} \prod_{r=1}^{n} \lambda_{ij}^{(r)} = 0$ for all i and j. And thus by theorem 1,2 it follows that $\lim_{n\to\infty} I_n = H(\xi)$ for all i and j $(j>i=1,\cdots,m)$. This proves the "*if*" part of the theorem. On the other hand, using the inequality $1-x \geqslant e^{-x/(1-x)}$, $(0\leqslant x \leqslant 1)$, we obtain

$$\prod_{r=1}^{n} \lambda_{ij}^{(r)} \geqslant \exp\left\{-\sum_{r=1}^{n}\left(\frac{1-\lambda_{ij}^{(r)}}{\lambda_{ij}^{(r)}}\right)\right\} \tag{2.12}$$

Now if $\sum_{r=1}^{\infty}(1-\lambda_{ij}^{(r)})$ is convergent for some fixed i and j, then $\lim_{r\to\infty} \lambda_{ij}^{(r)} = 1$ for some fixed i and j, and since by assumption $\lambda_{ij}^{(r)} > 0$ for $r=1,2,\cdots$ and $j>i=1,2,\cdots,m$ it follows that the sequence $\lambda_{ij}^{(r)}$ has a positive lower bound for some fixed i and j:

$$\lambda_{ij}^{(r)} \geqslant K > 0 \text{ for } r=1,2,\cdots \text{ and some fixed } i \text{ and } j.$$

If follows that the series $\sum_{r=1}^{\infty} (1-\lambda_{ij}^{(r)}/\lambda_{ij}^{(r)})$ is also convergent for some fixed i and j. By lemma 2.1 this implies ε_n has a positive lower bound. Therefore, by Theorem 1.1 the sequence $H(\xi)-I_n$ has a positive lower bound too. This proves the "only if" part of Theorem2.2.

REMARK 2.1. In view of the Theorem 2.1 and theorem 2.2, we have following result: If the number of components of the state space S increase, then the obtaining full information on S is relatively difucult. This coincides with one's intuitive sence.

References

[1] Ash, R. B., *Information Theory*, Johhn Wiley and sons, New York. London Sydney, (1967), 1-26.

[2] DeGroot, M. H., *Uncertainty, Information and Sequential Experiments*, Ann. Math. Stat Vol. 33 (1962), 404-419.

[3] DeGroot, M. H., *Optimal Statistical decision.* McGraw-Hill (1970), 121-154.

[4] Iindley, D. V., *On the measure of the information provided by an experiment*, Ann. Math. Stat. Vol 27 (1956), 986-1005.

[5] Marschak, J and Miyasawa, K., *Economic Comparability of Information Systems*, Intern. Econ. Rev. 9 (1968), 137-174.

[6] Miyasawa, K., *Information value and the entropy formulars*, (To be presented at the Second World Congress of the Economic Society in Cambridge, England: September) (1970), 1-40.

[7] Miyasawa, K., *Theory of Information and Statistical Decision*, Ewanami. (1971), 266-436,

[8] Rényi, A., *On the amount of missing information and the Neyman-Peason lemma*, Research Papers in Statistics (ed by David), John Wiley and Sons, Inc., New York (1966), 281-288,

[9] Rényi, A., *On some basic problems of statistics from the point of view of information theory.* Proc. 5th Berkley Symp. Math. Statist. Probability 1. University of California Press, Berkeley, Calif. (1967), 531-543.

[10] Maritz, J. S., *Empirical Bayes Methods*, Methuen and Co ltd (1970), 1-151.

Seoul National University