J. Korean Math. Soc.
Vol. 11, No. 2, 1974

ON THE EVALUATION OF INFORMATION VALUE AND AMOUNTS IN VIEW OF DECISION THEORY

BY JAE-JOO KIM

0. Introduction.

Consider a situation in which it is desired to gain knowledge 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 K. Miyasawa [11], 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. Information amount of an information is defined by [4], [7] and [8], 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 information will be evaluated by its information value which is defined depending on [8], [11] and [4], as the expected difference between the Bayes risk of the prior distribution over S and Bayes risk of the posterior distribution. The loss function in the specific decision problem is at hand. In studying information value, it is true that information cost should also be taken into consideration. However, in this paper, we shall confine our study to the gross information value neglecting information cost.

The problems which will be discussed in what follows are:

1. 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)?
2. Can the information amount of an information provide any guide post to its information value by giving a certain lower and upper bounds to it?
3. It is possible to evaluate marginal information value [11] by giving certain lower and upper bounds to it?

The main object of this paper is to try to answer the above questions in the case $S=\{s_1, s_2, \cdots, s_m\}$.

In section 1 we shall discuss the following problem from the point of view of the question 1. Let be given an infinite sequence $\{X_n\}$, $(n=1, 2, \cdots)$ of observations.

We suppose that the distributions of the random variables depend on a state S, whose set of possible value is finite. We suppose further that for each fixed value of S the random variables are not independent in general. We shall consider the amount of information on S which is still missing after having observed the values $X_1, X_2, \cdots, X_n$ 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. Matsusita [10] evaluated the error of the "standard decision" by means of "*affinity*".

We shall evaluate the success rate of the "*standard decision*" by means of "*affinity*".

Received by the editors Nov. 9, 1974.

In section 2 we shall discuss the question 1 by means of the results in section 1.

In section 3 and 4 we shall discuss the question 2 and 3 respectively using the results in section 1.

In section 5 we shall show that our results in section 1 cannot be generalized to the case where the set of possible value of S is infinite.

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

Let $(S, \boldsymbol{S})$ be a measurable state space, ξ be a prior probability measure on $(S, \boldsymbol{S})$. Let ξ be 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 $s \epsilon S$ is defined by

$$H(\xi) = -\int \xi(s) \log \xi(s) d\lambda. \tag{1.1}$$

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

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

Let us denote by $X(n)$ the random n-dimensional vector with coponents $(X_1, X_2, \cdots, X_n)$. Let $X(n)$ be a random variable on a measurable space $(X(n), \boldsymbol{X}(n))$ whose probability density function $f(x(n)|s)$ given $s \epsilon S$ with respect to a measure μ on $(X(n), \boldsymbol{X}(n))$ is assumed to be known. We suppose that the random variables $X_i (i=1, \cdots, n)$ are not independent in general under the condition $S=s_i$ $(i=1, \cdots, m)$ is given. Let $f_1(x(n))$, $f_2(x(n)), \cdots$ and $f_m(x(n))$ denote the density function for $S=s_1$, $S=s_2, \cdots$ and $S=s_m$, respectively.

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

Let $(A, \mathcal{A})$ be a measurable action space and $W(.,.)$ be a measurable loss function defined on $(S \times A, \ \mathbf{S} \times \mathcal{A})$. Then we shall say that these elements $(S, \boldsymbol{S}), \xi, (A, \mathcal{A})$ and W specify a *basic decision problem* D_0 and denote it as $D_0 = \{S, \xi, A, W\}$. If a decision maker can know the realized value $x(n)$ of $X(n)$, then we shall say an information $e(X(n))$ is available to a decision maker for solving a decision problem D_0. And then we shall say that he has a decision problem $D \equiv \{D_0 : e(X(n))\}$.

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

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

For simplicity, we shall denote the posterior probability measure given $X(n) = x(n)$ by $\xi(x(n))$, when the prior probability measure is ξ.

Specifically, if $S = \{s_1, s_2, , \cdots, s_m\}$ and $\xi = (\xi_1, \xi_2, \cdots, \xi_m) \in S^{m-1}$, 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 $$S^{m-1} \equiv \{(r_1, r_2, \cdots, r_m); r_i \geqq 0,\ \sum_{i=1}^{m} r_i = 1\}.$$

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

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

is called *the amount of missing information after observing* $e(X(n))$ by Rényi [13], [14] *the equivocation of* $e(X(n))$.

The information amount $I(X(n)\mid\xi)$ which an information $e(X(n))$ provides is defined by

(1.5) $$I(X(n)\mid\xi)\equiv H(\xi)-M(X(n)\mid\xi).$$

It is well known that $H(\xi)$ and $M(X(n)\mid\xi)$ are concave function of $\xi\in S^{m-1}$and $\xi(x(n))\in S^{m-1}$ respectively, and that $I(X(n)\mid\xi)\geqq 0$ [4].

Let us introduce the following decision rule. The most natural decision after having observed $X(n)=x(n)$ is essentially same as the reasoning applied in the case $S=\{s_1, s_2\}$, the case of two simple hypotheses [14].

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 the greatest, as 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 random variable $\Phi_n=\Phi(X(n))$ as follows:

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

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

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

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

(1.8) $$\begin{aligned}\varepsilon_n=\Big\{&\xi_1\Big[\int_{X(2)}f_1(x(n))d\mu+\int_{X(3)}f_1(x(n))d\mu+\cdots+\int_{X(m)}f_1(x(n))d\mu\Big]\\ &+\xi_2\Big[\int_{X(1)}f_2(x(n))d\mu+\int_{X(3)}f_2(x(n))d\mu+\cdots+\int_{X(m)}f_2(x(n))d\mu\Big]\\ &+\cdots\cdots\cdots\cdots\\ &+\xi_m\Big[\int_{X(1)}f_m(x(n))d\mu+\int_{X(2)}f_m(x(n))d\mu+\cdots+\int_{X(m-1)}f_m(x(n))d\mu\Big]\Big\}\end{aligned}$$

where

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

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

and

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

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

(1.12) $X_{(1)}=\left\{x(n):\frac{\xi_1}{\xi_2}\cdot\frac{f_1(x(n))}{f_2(x(n))}\geqslant 1, \frac{\xi_1}{\xi_3}\cdot\frac{f_1(x(n))}{f_3(x(n))}\geqslant 1, \cdots and\ \frac{\xi_1}{\xi_m}\cdot\frac{f_1(x(n))}{f_m(x(n))}\geqslant 1\right\},$

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

$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$

and

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

Matusita [10] evaluated ε_n by means of "*affinity*".

The "*affinity*" is defined as follows. Let $F_1, F_2, \cdots, F_m$ be distributions defined in the same space $(X(n), \boldsymbol{X}(n))$ with measure μ (Lebesque or counting or mixed), and let $f_1(x(n)), f_2(x(n)), \cdots, f_m(x(n))$ be respectively their density functions with respect to μ. Then, we call the quantity

(1.15) $$\rho_m(F_1, F_2, \cdots, F_m)=\int_{X(n)}(f_1(x(n))\cdots f_m(x(n)))^{\frac{1}{m}}d\mu$$

the affinity of $F_1, F_2, \cdots, F_m$.

Hinted by [9] and [10], we shall evaluate the success rate φ_n which is defined as follows:

(1.16) $$\varphi_n=\int_{X(1)}\xi_1 f_1(x(n))d\mu+\int_{X(2)}\xi_2 f_2(x(n))d\mu+\cdots+\int_{X(m)}\xi_m f_m(x(n))d\mu.$$

THEOREM 1.1. *Under the decision rule Φ_n, the success rate φ_n is bounded from above by*

(1.17) $$1-(m-1)\xi_1\xi_2\cdots\xi_m\frac{\rho_m{}^m(F_1, F_2, \cdots, F_m)}{m^{m-1}}$$

and from below by $\frac{1}{m}$.

Proof. We can easily know that $\varphi_n=1-\varepsilon_n$. By [10], ε_n is bounded from below by

(1.18) $$(m-1)\xi_1\xi_2\cdots\xi_m\frac{\rho_m{}^m(F_1, F_2, \cdots, F_m)}{m^{m-1}}.$$

Since the quantity of (1.18) is nonnegative [9], we have

(1.19) $$\varphi_n\leqslant 1-(m-1)\xi_1\xi_2\cdots\xi_m\frac{\rho_m{}^m(F_1, F_2, \cdots, F_m)}{m^{m-1}}.$$

On the other hand, we have

$$(m-1)\int_{X(1)}\xi_1 f_1(x(n))d\mu\geqslant\int_{X(1)}(\xi_2 f_2(x(n))+\xi_3 f_3(x(n))+\cdots+\xi_m f_m(x(n)))d\mu$$

(1.20) $$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$(m-1)\int_{X(m)}\xi_m f_m(x(n))d\mu\geqslant\int_{X(m)}(\xi_1 f_1(x(n))+\xi_2 f_2(x(n))+\cdots+\xi_{m-1}f_{m-1}(x(n)))d\mu.$$

The sum of right hand side of (1.20) is not greater than the sum of left hand side of (1.20). Therefore, we have

$$(m-1)\varphi_n \geqq \varepsilon_n, \tag{1.21}$$

By $\varphi_n = 1-\varepsilon_n$, we obtain $\varphi_n \geqq \frac{1}{m}$.

THEOREM 1.2. *One has*

$$\varepsilon_n \leqq E[H_2(\xi(X(n)))] = M_2(X(n)\,|\,\xi) \tag{1.22}$$

where the subscript 2 *on* H_2 *stands for "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:

(1.23) $f_1 \equiv f_1(x(n))$, $f_2 \equiv f_2(x(n)), \cdots$, $f_m \equiv f_m(x(n))$ and $f \equiv f(x(n)) = \xi_1 f_1 + \xi_2 f_2 + \cdots + \xi_m f_m$.

One has clearly

$$\begin{aligned} E[H(\xi(X(n)))] &= \int_{X(n)} H(\xi(x(n)))(\xi_1 f_1 + \xi_2 f_2 + \cdots + \xi_m f_m)d\mu \\ &= \xi_1 \int_{X(n)} H(\xi(x(n)))f_1 d\mu + \xi_2 \int_{X(n)} H(\xi(x(n)))f_2 d\mu \\ &\quad + \cdots + \xi_m \int_{X(n)} H(\xi(x(n)))f_m d\mu. \end{aligned} \tag{1.24}$$

By the definition of $H(\cdot)$, the r.h.s. term in (1.24) is

$$\begin{aligned} &\xi_1 \int_{X(n)} \Bigg[\frac{\xi_1 f_1}{\sum_{i=1}^m \xi_i f_i} \log\left(1 + \frac{\xi_2 f_2}{\xi_1 f_1} + \frac{\xi_3 f_3}{\xi_1 f_1} + \cdots + \frac{\xi_m f_m}{\xi_1 f_1}\right) \\ &\quad + \frac{\xi_2 f_2}{\sum_{i=1}^m \xi_i f_i} \log\left(1 + \frac{\xi_1 f_1}{\xi_2 f_2} + \frac{\xi_3 f_3}{\xi_2 f_2} + \cdots + \frac{\xi_m f_m}{\xi_2 f_2}\right) \\ &\quad + \cdots\cdots\cdots\cdots\cdots\cdots \\ &\quad + \frac{\xi_m f_m}{\sum_{i=1}^m \xi_i f_i} \log\left(1 + \frac{\xi_1 f_1}{\xi_m f_m} + \frac{\xi_2 f_2}{\xi_m f_m} + \cdots + \frac{\xi_{m-1} f_{m-1}}{\xi_m f_m}\right)\Bigg] f_1 d\mu \\ &\quad + \xi_2 \int_{X(n)} [*] f_2 d\mu + \cdots + \xi_m \int_{X(n)} [*] f_m d\mu \end{aligned} \tag{1.25}$$

$$\begin{aligned} &= \xi_1 \int_{X(n)} f_1 \log\left(1 + \frac{\xi_2 f_2}{\xi_1 f_1} + \frac{\xi_3 f_3}{\xi_1 f_1} + \cdots + \frac{\xi_m f_m}{\xi_1 f_1}\right) d\mu \\ &\quad + \xi_2 \int_{X(n)} f_2 \log\left(1 + \frac{\xi_1 f_1}{\xi_2 f_2} + \frac{\xi_3 f_3}{\xi_2 f_2} + \cdots + \frac{\xi_m f_m}{\xi_2 f_2}\right) d\mu \\ &\quad + \cdots\cdots\cdots\cdots\cdots\cdots \\ &\quad + \xi_m \int_{X(n)} f_m \log\left(1 + \frac{\xi_1 f_1}{\xi_m f_m} + \frac{\xi_2 f_2}{\xi_m f_m} + \cdots + \frac{\xi_{m-1} f_{m-1}}{\xi_m f_m}\right) d\mu. \end{aligned} \tag{1.26}$$

By (1.12), (1.13) and (1.14), each terms in (1.26) is not less than each term in the following (1.27),

---

*The notation [*] denote the first term in [ ] of (1.25)

$$\begin{aligned}&\Big\{\xi_1\Big[\int_{X(2)}f_1d\mu+\int_{X(3)}f_1d\mu+\cdots+\int_{X(m)}f_1d\mu\Big]\\&+\xi_2\Big[\int_{X(1)}f_2d\mu+\int_{X(3)}f_2d\mu+\cdots+\int_{X(m)}f_2d\mu\Big]\\&+\cdots+\xi_m\Big[\int_{X(1)}f_md\mu+\int_{X(2)}f_md\mu+\cdots+\int_{X(m-1)}f_md\mu\Big]\Big\}=\varepsilon_n.\end{aligned}\tag{1.27}$$

Thus the theorem is proved.

REMARK 1.1. Theorem 1.2 also holds when the random variables $X_i(i=1,\cdots,n)$ are independent under the condition that $S=s_i(i=1,2,\cdots,m)$ is given.

REMARK1.2. Rényi's theorem 3 in [13] is special case of our theorem 1.2 where $X_i(i=1,\cdots,n)$ are independent and identically distributed under the condition $S=s_i(i=1,2)$ is given.

THEOREM 1.3 *Let us write*

$$\lambda_{ij}=\int_{X(n)}\sqrt{f_i(x(n))f_j(x(n))}\,d\mu \quad \textit{for} \quad i\neq j=1,2,\cdots,\tag{1.28}$$

then the following inequality holds:

$$0\leqslant E[H(\xi(X(n)))]\leqslant C\sum_{i\neq j=1}^{m}\sqrt{\xi_i\cdot\xi_j}\lambda_{ij}\tag{1.29}$$

where

$$C=\max_{0\le x\le 1}\frac{h(x)}{\sqrt{x}}\left(=\max_{0\le x\le 1}\frac{h(x)}{\sqrt{1-x}}\right)\tag{1.30}$$

and

$$h(x)=-x\log x-(1-x)\log(1-x) \quad \textit{for } 0<x<1, \quad h(0)=h(1)=0.\tag{1.31}$$

Proof. Since $f=\xi_1f_1+\xi_2f_2+\cdots+\xi_mf_m$ by (1.23), we have

$$\begin{aligned}E[H(\xi(X(n)))]&=\int_{X(n)}H(\xi(x(n)))(\xi_1f_1+\cdots+\xi_mf_m)d\mu\\&=\xi_1\int_{X(n)}H(\xi(x(n)))f_1d\mu+\xi_2\int_{X(n)}H(\xi(x(n)))f_2d\mu+\cdots\\&\quad+\xi_m\int_{X(n)}H(\xi(x(n)))f_md\mu.\end{aligned}\tag{1.32}$$

By the grouping axiom of entropy [3], the first r.h.s. term in (1.32) is

$$\begin{aligned}&\xi_1\int_{X(n)}H(\xi(x(n)))f_1d\mu\\&=\xi_1\int_{X(n)}\Big[H(\xi_1(x(n))+\xi_2(x(n))+\cdots+\xi_{m-1}(x(n)),\xi_m(x(n)))\\&\quad+\Big(\sum_{i=1}^{m-1}\xi_i(x(n))\Big)H\Big(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))},\cdots,\frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_1(x(n))}\Big)\\&\quad+\xi_m(xn))H(1)\Big]f_1d\mu\\&=\xi_1\int_{X(n)}\Big[H\Big(\sum_{i=1}^{m-1}\xi_1(x(n)),\ \xi_m(x(n))\Big)\Big]f_1d\mu\end{aligned}\tag{1.33}$$

$$(1.34)\quad +\xi_1\int_{X(n)}\left[\left(\sum_{i=1}^{m-1}\xi_i(x(n))\right) H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \cdots, \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_1(x(n))}\right)\right] f_1 d\mu$$

Since $\sum_{i=1}^{m-1}\xi_i(x(n))\leqq 1$, the r. h. s. terms in (1.34) are not greater than the following:

$$(1.35)\quad \xi_1\int_{X(n)}\left[H\left(\sum_{i=1}^{m-1}\xi_i(x(n)),\ \xi_m(x(n))\right)\right] f_1 d\mu$$
$$+\xi_1\int_{X(n)}\left[H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \cdots, \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)\right] f_1 d\mu.$$

From the definition of C it follows that

$$H\left(\sum_{i=1}^{m-1}\xi_i(x(n)),\ \xi_m(x(n))\right)\leqq C[\xi_m(x(n))]^{\frac{1}{2}}$$
$$=C\left(\frac{\xi_m f_m}{\sum_{i=1}^{m}\xi_i f_i}\right)^{\frac{1}{2}}\leqq C\left(\frac{\xi_m}{\xi_1}\right)^{\frac{1}{2}}\left(\frac{f_m}{f_1}\right)^{\frac{1}{2}}.$$

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

$$H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \frac{\xi_2(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \cdots, \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)$$
$$=H\left(\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)$$
$$(1.37)\quad +\sum\left(\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right) H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}, \frac{\xi_2(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}, \cdots, \frac{\xi_{m-2}(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}\right)$$
$$+\frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}H(1).$$

Since $\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\leqq 1$ and $H(1)=0$, the r. h. s. in (1.37) is not greater than the following:

$$H\left(\frac{\sum_{i=1}^{m-2}\xi_i(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}, \frac{\xi_{m-1}(x(n))}{\sum_{i=1}^{m-1}\xi_i(x(n))}\right)$$
$$(1.38)\quad +H\left(\frac{\xi_1(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}, \frac{\xi_2(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}, \cdots, \frac{\xi_{m-2}(x(n))}{\sum_{i=1}^{m-2}\xi_i(x(n))}\right).$$

From the definition of C, it also follows that

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

$$(1.39)\qquad =C\left(\frac{\xi_{m-1}f_{m-1}}{\sum_{i=1}^{m-1}\xi_i f_i}\right)^{\frac{1}{2}}$$

$$\leqslant C\left(\frac{\xi_{m-1}}{\xi_1}\right)^{\frac{1}{2}}\left(\frac{f_{m-1}}{f_1}\right)^{\frac{1}{2}}.$$

Proceeding this process succesively, then[*] it follows that

$$(1.40)\qquad \xi_1\int_{X(n)}H(\xi(x(n)))f_1 d\mu\leqslant C\sum_{j=2}^{m}\sqrt{\xi_1\cdot\xi_j}\int_{X(n)}\sqrt{f_1\cdot f_j}d\mu$$

$$=C\sum_{j=2}^{m}\sqrt{\xi_1\cdot\xi_j}\lambda_{1j}.$$

By a similar method, it follows that

$$(1.41)\qquad \xi_2\int_{X(n)}H(\xi(x(n)))f_2 d\mu\leqslant C\sqrt{\xi_1\cdot\xi_2}\lambda_{12}+C\sum_{j=3}^{m}\sqrt{\xi_2\cdot\xi_j}\lambda_{2j}$$

and

$$(1.42)\qquad \xi_m\int_{X(n)}H(\xi(x(n)))f_m d\mu\leqslant C\sum_{i\neq j=1}^{m-1}\sqrt{\xi_m\cdot\xi_j}\lambda_{mj}.$$

Thus we have

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

The proof of the lower inequality of the theorem follows from the theorem 1.2. This proves the theorem.

COROLLARY 2.1. *Suppose that the random variables* $X_i(i=1,2,\cdots,n)$ *are independent under the condition* $S=s_i(i=1,2,\cdots m)$ *is given. Let us write*

$$(1.43)\qquad \lambda_{ij}^{(r)}=\int_{-\infty}^{\infty}\sqrt{f_i(x_r)f_j(x_r)}dx_r \text{ for } r=1,2,\cdots,n \text{ and } i\neq j=1,2,\cdots,m.$$

Then the following inequality holds:

$$(1.44)\qquad 0\leqslant E\Big[H(\xi(X(n)))\Big]\leqslant C\sum_{i\neq j=1}^{m}\sqrt{\xi_i\cdot\xi_j}\prod_{r=1}^{n}\lambda_{ij}^{(r)}$$

Where C is defined by (1.30).

REMARK 1.3. The inequalities of (1.44) is extension of Rényi's result [14]:

$$(1.45)\qquad 0\leqslant E\Big[[H(\xi(X(n)))\Big]\leqslant 2C\sqrt{\xi_1\cdot\xi_2}\prod_{r=1}^{n}\lambda_{12}^{(r)}$$

The inequalities in (1.45) are special case of $m=2$ in (1.44).

Proof of corollary 2.1. Under the hypotheses we have

$$(1.46)\qquad {}^*\lambda_{ij}=\prod_{r=1}^{n}\int_{-\infty}^{\infty}\sqrt{f_i(x_r)dx_r}=\prod_{r=1}^{n}\lambda_{ij}^{(r)}$$

where λ_{ij} is defined by (1.28). From (1.46) and (1.29), we have (1.44).

REMARK 1.4. Suppose that the random variables $X_i(i=1,2,\cdots,n)$ are independent and identically distributed under the condition $S=s_i(i=1,\cdots,m)$ is given. Let us write

$$(1.47)\qquad \lambda_{ij}^{(1)}=\lambda_{ij}^{(2)}=\cdots=\lambda_{ij}^{(n)}=\beta_{ij}$$

Then from (1.44), we have

*By Schwarz inequality, we have $0\leqslant\lambda_{ij}\leqslant 1$.

$$(1.48) \qquad 0\leqslant E[H(\xi(X(n)))]\leqslant C\sum_{i\neq j=1}^{m}\sqrt{\xi_i\cdot\xi_j}(\beta_{ij})^n.$$

If $0\leqslant\beta_{ij}<1$ for all i and j, then the relation (1.48) shows how fast the equivocation $E[H(\xi(X(n)))]$ of an information $e(X(n))$ approaches 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 obtainining full information.

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

Let us denote by $X(n)$ the random n-dimensional vector with components $(X_1, X_2, \cdots, X_n)$. Let I_n denote the amount of information contained in $e(X(n))$ concerning S. Then we have

$$(2.1) \qquad I_n=H(\xi)-E[H(\xi(X(n)))]$$

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

It is easy to see that I_n is nondecreasing for $n=1, 2, \cdots$ and $I_n\leqslant H(\xi)$. Thus $\lim_{n\to+\infty} I_n=I^*$ always exists. If $I^*=H(\xi)$ we shall say that the sequence of observations $\{X_i\}$ $(i=1, 2, \cdots)$ gives us *full information on* S, whereas in the case $I^*<H(\xi)$ we shall say that the observations $\{X_i\}$, $(i=1, 2, \cdots)$ do not give full information on S. Rényi [14] introduced a criterion for obtaining full information

THEOREM 2.1. *If* $\lambda_{12}^{(r)}>0$ *for* $r=1, 2, \cdots$ *where*

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

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

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

is divergent.

REMARK 2.1. For each fixed value of S if the random variables $X_i (i=1, 2, \cdots)$ are not independent, then author can not give any criterion for obtaining full information.

For generalization of Theorem 2.1 to the case $S=\{s_1, s_2, \cdots, s_m\}$ we get the following Lemma 2.1.

LEMMA 2.1. *One has*

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

where $\lambda_{ij}^{(r)}$ *and* ε_n *is defined by* (1.43) *and* (1.8) *respectively.*

REMARK 2.2. This lemma also holds in the case of fixed value of S where the random variables $X_i (i=1, 2, \cdots, n)$ are not independent.

Proof of lemma 2.1. Clearly,

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

where $X(n)$ is the n-dimensional Euclidian space and $dx(n)$ stands for $dx_1, dx_2, \cdots, dx_n$. 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

$$(2.6) \qquad \int_{X_{(i)}} \sqrt{f_i(x(n)) f_j(x(n))}\, dx(n) \leqslant \left(\int_{X_{(i)}} f_j(x(n))\, dx(n)\right)^{\frac{1}{2}}$$

and

$$(2.7) \qquad \int_{\bar{X}_{(i)}} \sqrt{f_i(x(n)) f_j(x(n))}\, dx(n) \leqslant \left(\int_{\bar{X}_{(i)}} f_i(x(n))\, dx(n)\right)^{1}.$$

Since

$$(2.8) \qquad \varepsilon_n = \sum_{i \neq j = 1}^{m} \xi_j \int_{X_{(i)}} f_j(x(n))\, dx(n) \ [15],$$

we have

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

Therefore, we have

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

This proves Lemma.

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* (1.43), *the sequence of observations $X_i (i = 1, 2, \cdots)$ contains full information on S if and only if*

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

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 by corollary 2.1 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

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

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 sequence $\{\lambda_{ij}^{(r)}\}$ $(r = 1, 2, \cdots)$ has a positive lower bound for some fixed i and j, $\lambda_{ij}^{(r)} \geqslant k > 0$ for $r = 1, 2, \cdots$. It 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 1.2 the sequence $H(\xi)-I_n$ has a positive lower bound too. This proves the "only if" part of the theorem 2.2.

REMARK 2.3. In view of the theorem 2.1 and theorem 2.2, we have the following result; if the numbers of components of the state space S increase, then the obtaining of full information on S is relatively difficult. This coincides with one's intuitive sense.

THEOREM 2.3. *A sequence of statistics $T_n=t_n\ (X_1, X_2, \cdots, X_n)$ converging in probability to the true value of the parameter (real valued) can exist only if the sequence $\{X_i\}$ $(i=1, 2, \cdots)$ contains full information with respect to s, that is, if $\lim_{n\to+\infty} I_n=H(\xi)$ holds. Conversely if $\lim_{n\to+\infty} I_n=H(\xi)$ holds, there exists a sequence of statistics T_n which converges with probability* 1 *to s.*

Proof. If $\lim_{n\to\infty} P(|T_n-s|>\epsilon)=0$ for every $\epsilon>0$, then we construct the following decision function:

$$(2.13)\quad \begin{cases} \text{if } |t_n-s_1|<|t_n-s_j| \text{ where } j=2, \cdots, m \text{ accept } s_1; \\ \text{if } |t_n-s_2|<|t_n-s_j| \text{ where } j=1, 3, \cdots, m \text{ accept } s_2; \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \text{if } |t_n-s_m|<|t_n-s_j| \text{ where } j=1, 2, \cdots, m-1 \text{ accept } s_m; \\ \text{if } |t_n-s_i|=|t_n=s_j| \text{ where } i\neq j=1, \cdots, m \text{ choose at random} \\ \text{between } s_i \text{ and } s_j \text{ with probability } \xi_i \text{ and } \xi_j. \end{cases}$$

Let us suppose that $s_i>s_j,\ j=2, \cdots,$ m. Clearly,

$$(2.14)\qquad P(|T_n-s_j|\leq|T_n-s_1|\,\big|\,s=s_1)=P(T_n\leq(s_j+s_1)/2|s=s_1)$$

and thus by assumption,

$$\lim_{n\to\infty} P(|T_n-s_j|\leq|T_n-s_1|\,\big|\,s=s_1)=0.$$

Let us suppose that $s_1<s_j,\ j=2, \cdots, m$. Clearly,

$$(2.15)\qquad P(|T_n-s_j|\leq|T_n-s_1|\,\big|\,s=s_1)=P(T_n\geq(s_1+s_j)/2|s=s_1),$$

and thus by assumption,

$$(2.16)\qquad \lim_{n\to\infty} P(|T_n-s_j|\leq|T_n-s_1|\,\big|\,s=s_1)=0.$$

Similary,

$$(2.17)\qquad \lim_{n\to\infty} P(|T_n-s_1|\leq|T_n-s_j|\,\big|\,s=s_j,\ j=2, \cdots, m)=0.$$

By the same method, we have the followings:

$$(2.18)\qquad \lim_{n\to 8} P(|T_n-s_j|\leq|T_n-s_2|\,\big|\,s=s_2)=0, \text{ where } j=1, 3, \cdots, m,$$

$$(2.19)\qquad \lim_{n\to\infty} P(|T_n-s_2|\leq|T_n-s_j|\,\big|\,s=s_j,\ j=1, 3, \cdots, m)=0,$$

....

(2.20) $\lim_{n\to\infty} P(|T_n-s_j|\leq|T_n-s_m|\,|s=s_m)=0,$ where $j=1, 2, \cdots, m-1,$

and

(2.21) $\lim_{n\to\infty} P(|T_n-s_m|\leq|T_n-s_j|\,|\,s=s_j,\ j=1, 2, \cdots, m-1)=0.$

The sum of all terms of (2.16), (2.17), (2.18), (2.19), (2.20) and (2.21) is equal to the error of the decision for $n\longrightarrow\infty$. Thus the error of the standard decision tends to zero which implies by theorem 1.2 and lemma 2.1 that $\lim_{n\to\infty} I_n=H(\xi)$. This proves the "only if" part of theorem 2.3. Conversely, if $\lim_{n\to\infty} I_n=H(\xi)$ holds and $s_1, s_2, \cdots$ and s_m are distinct real numbers, let us choose a sequence n_r such that the series $\sum_{r=1}^{\infty}\varepsilon_{n_r}$ is convergent. Then by the Borel-Cantelli lemma [6] the standard decision ϕ_{n_r} tends with probability 1 to s. More exactly, $\phi_{n_r}=s$ for all but a finite numbers of value of r. Now if we put $T_n=\phi_{n_r}$, for $n_r\leq n<n_{r+1}$, then the sequence of statistics $T_n=t_n(X_1, X_2, \cdots, X_n)$ tends with probability 1 to s, which implies our theorem.

3. Evaluation of information value.

In this section, for one information, we shall study if its information amount could give some guide post to its information value.

We shall study in the case of $S=\{s_1, s_2, \cdots, s_m\}$ and $A=\{a_1, a_2, \cdots, a_m\}$. Then without loss of generality, the loss matrix can be assumed to be given by Diagram I:

$$\begin{array}{c|ccccccc}
 & a_1 & a_2 & \cdot & \cdot & \cdot & & a_m \\
\hline
s_1 & 0 & W_{12} & \cdot & \cdot & \cdot & & W_{1m} \\
s_2 & W_{21} & 0 & \cdot & & & & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & W_{m-1m} \\
s_m & W_{m1} & W_{m2} & \cdot & & \cdot & W_{mm-1} & 0
\end{array}$$

Diagram I.

Let $\xi=(\xi_1, \xi_2, \cdots, \xi_m)$ be the prior probability law on S and $D_0=\{S, \xi, A, W\}$ be a basic decision problem.

Let $e(X(n))$ be an information for $s\in S$ defined by a random variable $X(n)$ on $(X(n), \boldsymbol{X}(n))$ whose probability density function given $S=s_i$ with respect to a measure μ on

$(X(n),\ X(n))$ is $f_i(x(n))$, $i=1,2,\cdots,m$.

Now being given the above basic decision problem $D_0=\{S,\xi,A,W\}$, we shall construct a modified, so to speak, a decision oriented prior probability law $\eta=(\eta_1,\eta_2,\cdots,\eta_m)$ on S as follows:

$$\eta_j=\frac{\sum_{i=1}^{m}\xi_i W_{ij}}{\bar{w}} \tag{3.2}$$

where

$$\bar{w}=\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\xi_j W_{ij}\right). \tag{3, 3}$$

The information value $V(X(n)\,|\,\xi,W)$ of $e(X(n))$ for D_0 is defined as follows:

$$V(X(n)\,|\,\xi,W)=R(D_0)-R(D), \tag{3.4}$$

where

$$R(D_0)\equiv R_0(\xi\,|\,W)\equiv\inf_{a\in A}W(s,a)d\xi(s) \tag{3.5}$$

and

$$R(D)\equiv R(X(n)\,|\,\xi,W)\equiv E[R_0(\xi(X(n))\,|\,W)]. \tag{3.6}$$

$R(D_0)$ and $R(D)$ are concave functions of $\xi\in S^{m-1}$ and $\xi(x(n))\in S^{m-1}$ respectively which depend on W [5]. Then the value $V(X(n)\,|\,\xi,W)$ of $e(X(n))$ for D_0 can be evaluated by the following:

THEOREM 3.1.

$$R(D_0)-\bar{W}M(X(n)\,|\,\eta)\leqslant V(X(n)\,|\,\xi,W)\leqslant R(D_0)-\frac{\bar{W}M^2(X(n)\,|\,\eta)}{C^2\left[\sum_{i\neq j=1}^{m}(\sqrt{\eta_i}+\sqrt{\eta_j})\right]^2} \tag{3.7}$$

where $M(X(n)\,|\,\eta)$ is the equivocation of $e(X(n))$ defined by (1.4) *under the assumption that the prior probability law on S is not the given ξ but the modified η by* (3.2) *and* (3.3), *and C is a constant defined by* (1.30).

REMARK 3.1. Since the information amount of $e(X(n))$ under the assumption that the prior probability law on S is η is given by $I(X(n)\,|\,\eta)=H(\eta)-M(X(n)\,|\,\eta)$, it is clear that the inequalities of theorem 3.1 could be expressed in terms of a modified information amount $I(X(n)\,|\,\eta)$.

REMARK 3.2. If the prior probability law ξ on S is objectively agreed upon in D_0, then we feel uneasy with theorem 3.1, because of the reason that the bounds to $V(X(n)\,|\,\xi,W)$ given by the theorem 3.1 are expressed in terms of a modified information amount $I(X(n)\,|\,\eta)$ of $e(X(n))$ but not in terms of the information amount $I(X(n)\,|\,\xi)$. This point is relieved in Miyasawa [11] on the simple case;

$$S=\{s_1,s_2\},\ A=\{a_1,a_2\}\ \text{and loss matrix:}\ \begin{array}{c|cc} & a_1 & a_2 \\ \hline s_1 & 0 & W_1 \\ s_2 & W_2 & 0 \end{array}\ . \tag{3.8}$$

REMARK 3.3. In spite of remark 3.2 the theorem 3.1 gives the following result. For each fixed value of S if the random variables $X_i(i=1,2,\cdots,n)$ are independent, then by corollary 2.1, we have

$$(3.9)\qquad M(X(n)\mid\eta)\leqq C\sum_{i\neq j=1}^{m}\sqrt{\eta_i\cdot\eta_j}\prod_{r=1}^{n}\lambda_{ij}^{(r)}.$$

Then from theorem 3.1 and (3.9), we have

$$(3.10)\qquad R(D_0)-\bar{W}C\sum_{i\neq j=1}^{m}\sqrt{\eta_i\cdot\eta_j}\prod_{r=1}^{n}\lambda_{ij}^{(r)}\leqq V(X(n)\mid\xi,w)\leqq R(D_0)-\frac{\bar{W}\left(\sum_{i\neq j=1}^{m}\sqrt{\eta_i\cdot\eta_j}\prod_{r=1}^{n}\lambda_{ij}^{(r)}\right)^2}{\sum_{i\neq j=1}^{m}(\sqrt{\eta_i}+\sqrt{\eta_j})^2}$$

By the theorem 2.2 and (3.10), we know that if $\lambda_{ij}^{(r)}>0$ for $r=1,2,\cdots$ and $j>i=1,2,\cdots,m$, and $\sum_{r=1}^{\infty}(1-\lambda_{ij}^{(r)})$ $(j>i=1,2,\cdots,m)$ are divergent for all i and j, then we have that the value $V(X(n)\mid\xi,W)$ of an information $e(X(n))$ approaches to its maximum possible value $R(D_0)$ as n increases, irrespective of what a modified prior η is.

REMARK 3.4. For each fixed value of S if the random variables $X_i(i=1,2,\cdots,n)$ are independent and identically distributed, then by remark 1.4, we have

$$(3.11)\qquad M(X(n)\mid\eta)\leqq C\sum_{i\neq j=1}^{m}\sqrt{\eta_i\cdot\eta_j}(\beta_{ij})^n,$$

where $0\leqq\beta_{ij}<1$, assuming $f_i(x)\neq f_j(x)$ for all i and $j(j>i=1,2,\cdots,m)$ with respect to μ. Then from theorem 3.1 and (3.11), we have

$$(3.12)\qquad R(D_0)-\bar{W}C\sum_{i\neq j=1}^{m}\sqrt{\eta_i\cdot\eta_j}(\beta_{ij})^n\leqq V(X(n))\mid\xi,W)\leqq R(D_0).$$

Since $0\leqq\beta_{ij}<1$ for all i and j, the relation (3.12) shows how fast the value $v(X(n)\mid\xi,W)$ of an information $e(X(n))$ approaches to its maximum possible value $R(D_0)$ as the sample size n increases, irrespective of what a modified prior probability law η is.

Proof of the theorem 3.1. The Bayes risks $R(D_0)$ and $R(D)$ in the decision problems D_0 and $D=\{D_0\,;\,e(X(n))\}$ are given by

$$(3.13)\qquad R(D_0)=\bar{W}\min(\eta_1,\eta_2\cdots,\eta_m)$$

and

$$(3.14)\qquad R(D)=\bar{W}\int_{X(n)}\min(\eta_1 f_1(x(n)),\ \eta_2 f_2(x(n)),\cdots,\ \eta_m f_m(x(n)))d\mu$$

respectively. Here let us define

$$(3.15)\qquad r(D_0)\equiv R(D_0)/\bar{W},\ \ r(D)\equiv R(D)/\bar{W}$$

and follow the same reasoning as in Miyasawa [11].

Let the marginal probability law of $X(n)$ under the assumption that the prior probability law on S is a modified $\eta=(\eta_1,\eta_2,\cdots,\eta_m)$ be

$$(3.16)\qquad g(x(n))=\eta_1 f_1(x(n))+\eta_2 f_2(x(n))+\cdots+\eta_m f_m(x(n)).$$

Then the posterior probability of s_i after observing $X(n)=x(n)$ is given by

(3.17) $$\eta_i(x(n))=\eta_i f_i(x(n))/g(x(n)), \quad i=1,2,\cdots,m.$$

From (3.14), (3.15) and (3.17), we have

(3.18) $$r(D)=\int_{X(n)} g(x(n))\min(\eta_1(x(n)),\eta_2(x(n)),\cdots,\ \eta_m(x(n)))d\mu.$$

Define a division of the sample space $X(n)$ by

(3.19)
$$\begin{aligned}
X_{(1)}=\{x(n) :\ &\eta_1(x(n))\leqslant\eta_2(x(n)),\ \eta_1(x(n))\leqslant\eta_3(x(n)),\cdots \text{ and}\\
&\eta_1(x(n))\leqslant\eta_m(x(n))\}\\
X_{(2)}=\{x(n) :\ &\eta_2(x(n))<\eta_1(x(n)),\ \eta_2(x(n))\leqslant\eta_3(x(n)),\cdots \text{ and}\\
&\eta_2(x(n))\leqslant\eta_m(x(n))\}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
X_{(m)}=\{x(n) :\ &\eta_n(x(n))<\eta_1(x(n)),\ \eta_m(x(n))<\eta_2(x(n)),\cdots \text{ and}\\
&\eta_m(x(n))<\eta_{m-1}(x(n))\}.
\end{aligned}$$

Then from (3.18) and (3.19), we have

(3.20) $$r(D)=\eta_1\int_{X(1)} f_1(x(n))d\mu+\eta_2\int_{X(2)} f_2(x(n))d\mu+\cdots+\eta_m\int_{X(m)} f_m(x(n))d\mu.$$

On the other hand, under the assumption that the prior probability law on S is η, the equivocation $M(X(n)\mid\eta)$ of $e(X(n))$ is given by

(3.21)
$$\begin{aligned}
M(X(n)\mid\eta)&=E[H(\eta(X(n)))]\\
&=\eta_1\int_{X(n)} f_1(x(n))H(\eta(x(n)))d\mu+\eta_2\int_{X(n)} f_2(x(n))H(\eta(x(n)))d\mu\\
&\quad+\cdots+\eta_m\int_{X(n)} f_m(x(n))H(\eta(x(n)))d\mu.
\end{aligned}$$

Calculating the right hand side of (3.21), we have

(3.22)
$$\begin{aligned}
M(X(n)\mid\eta)=\eta_1\int_{X(n)} f_1(x(n))\ \log\frac{1}{\eta_1(x(n))}d\mu+\eta_2\int_{X(n)} f_2(x(n))\log\frac{1}{\eta_2(x(n))}d\mu\\
+\cdots+\eta_m\int_{X(n)} f_m(x(n))\log\frac{1}{\eta_m(x(n))}d\mu.
\end{aligned}$$

Now since $1/\eta_i(x(n))\geqslant1$ $(i=1,2,\cdots,m)$ for all $x(n)\in X(n)$, from (3.22), we have

$$\begin{aligned}
M(X(n)\mid\eta)&\geqslant\eta_1\int_{X(1)} f_1(x(n))\log\frac{1}{\eta_1(x(n))}d\mu+\eta_2\int_{X(2)} f_2(x(n))\log\frac{1}{\eta_2(x(n))}d\mu\\
&\quad+\cdots+\eta_m\int_{X(m)} f_m(x(n))\log\frac{1}{\eta_m(x(n))}d\mu\\
&=\eta_1\int_{X(1)} f_1(x(n))\log\left(1+\frac{\eta_2 f_2(x(n))}{\eta_1 f_1(x(n))}+\cdots+\frac{\eta_m f_m(x(n))}{\eta_1 f_1(x(n))}\right)d\mu
\end{aligned}$$

$$(3.23)\qquad +\eta_2\int_{X(2)} f_2(x(n))\log\left(1+\frac{\eta_1 f_1(x(n))}{\eta_2 f_2(x(n))}+\cdots+\frac{\eta_m f_m(x(n))}{\eta_2 f_2(x(n))}\right)d\mu$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+\eta_m\int_{X(m)} f_m(x(n))\log\left(1+\frac{\eta_1 f_1(x(n))}{\eta_m f_m(x(n))}+\cdots+\frac{\eta_{m-1} f_{m-1}(x(n))}{\eta_m f_m(x(n))}\right)d\mu.$$

From the definitions of $X_{(1)}, X_{(2)}, \cdots$ and $X_{(m)}$ given by (3.19), we have $\eta_2 f_2(x(n))/\eta_1 f_1(x(n))\geqq 1$ for $x(n)\in X_{(1)}$, $\eta_1 f_1(x(n))/\eta_2 f_2(x(n))>1$ for $x(n)\in X_{(2)}, \cdots$ and $\eta_1 f_1(x(n))/\eta_m f_m(x(n))>1$ for $x(n)\in X_{(m)}$. Therefore from (3.23) we have

$$(3.24)\qquad M(X(n)\,|\,\eta)\geq\eta_1\int_{X(1)} f_1(x(n))\,d\mu+\eta_2\int_{X(2)} f_2(x(n))\,d\mu+\cdots+\eta_m\int_{X(m)} f_m(x(n))\,d\mu.$$

Accordingly, from (3.20) and (3.24), we have

$$(3.25)\qquad r(D)\leqq M(X(n)\,|\,\eta).$$

Since

$$(3.26)\qquad V(X(n)\,|\,\xi, W)=R(D_0)-R(D),$$

therefore from (3.15), (3.25) and (3.26), we have the lower inequality of (3.7). Again by the same reasoning as in theorem 1.3 and lemma 2.1, we can show that

$$(3.27)\qquad M(X(n)\,|\,\eta)\leqq C\sum_{i\neq j=1}^{m}\sqrt{\eta_i\cdot\eta_j}\,(r(D))^{\frac{1}{2}}\frac{(\sqrt{\eta_i}+\sqrt{\eta_j})}{\sqrt{\eta_i\cdot\eta_j}}$$

$$=C\sum_{i\neq j=1}^{m}\left(\frac{R(D)}{\bar{W}}\right)^{\frac{1}{2}}(\sqrt{\eta_i}+\sqrt{\eta_j}).$$

Therefore, from (3.26) and (3.27), we have upper inequality of (3.7). Thus the theorem is proved.

4. Evaluaion of marginal information values.

In this section we evaluate the increment of information value which results from addition of one more observation to a given information, i.e., the marginal information value for the case of $S=\{s_1, \cdots, s_m\}$ and $A=\{a_1, \cdots, a_m\}$. Without loss of generality, the loss matrix can be assumed to be given by Diagram I in section 3. Hereafter we shall assume that the loss function W is bounded by K, i.e.,

$$0\leqq W(s, a)\leqq K,\ \text{for all } s\in S \text{ and } a\in A.$$

Now let us assume that the probability measure P_s, $s\in S$, on $(X, \boldsymbol{X})$ which specifies the random variable X of the information $e(X)$ has a density function $f(x\,|\,s)$ with respect to a measure μ on $(X, \boldsymbol{X})$. Let $X_1, \cdots, X_n, X_{n+1}$ be the random variables which, given $s\in S$, are distributed independently and identically with X. Then we shall call an information $e(X(n))$ which is defined by these random variables $X(n)=(X_1, \cdots, X_n)$ *a sample information of size n generated from X.*

Given $s\in S$, the probability density function of $X(n)$ on $(X(n), \boldsymbol{X}(n))$ is given by

(4.2)
$$f(x(n)\mid s)=f(x_1\mid s)\cdots f(x_n\mid s)$$

where $X(n)=X\cdot\cdots\cdot X$ and $\boldsymbol{X}(n)=\boldsymbol{X}\cdot\cdots\cdot\boldsymbol{X}$ (both multiplies n times X or $\boldsymbol{X}$). Similary, letting $e(X(n+1))$ be a sample information of size $(n+1)$ generated from X, on $(X(n+1),\ \boldsymbol{X}(n+1))$, its density function, given $s\in S$, is given by

(4.3)
$$f(x(n+1)\mid s)=f(x(n)\mid s)f(x_{n+1}\mid s).$$

We shall consider the decision problems D_n and D_{n+1} which are defined by $D_n=\{D_0\,;\,e(X(n))\}$ and $D_{n+1}=\{D_0\,;\,e(X(n+1))\}$ respectively. Thus if we can evaluate the marginal information amount $I(X(n+1)\mid\xi)-I(X(n)\mid\xi)=M(X(n)\mid\xi)-M(X(n+1)\mid\xi)$, then we can evaluate the marginal information value $V(X(n+1)\mid\xi,W)-V(X(n)\mid\xi,W)=R(D_n)-R(D_{n+1})$ by means of the result [1], [2] and [11]:

THEOREM 4.1.

(4.4)
$$0\leqslant R(D_n)-R(D_{n+1})\leqslant[2KR(D_n)]^{\frac{1}{2}}[I(X(n+1)\mid\xi)-I(X(n)\mid\xi)]^{\frac{1}{2}}.$$

THEOREM 4.2. *Let the state space be $S=\{s_1,\cdots,s_m\}$ and the prior probability law on S be $\xi=(\xi_1,\xi_2,\cdots,\xi_m)>0$ and the basic decision problem D_0 be given by $D_0=\{S,\xi,A,W\}$, where $0\leqslant W(s,a)\leqslant K$ for all $s\in S$ and $a\in A$. Let X be a random variable on $(X,\boldsymbol{X})$ whose probability density function given $S=s_i$ is $f_i(x)$ with respect to a measure μ on $(X,\boldsymbol{X})$, $i=1,2,\cdots,m$. Hereafter the following condition is assumed:*

(4.5) *$f_i(x)$ and $f_j(x)$ are not equal almost everywhere for $j>i=1,2,\cdots,m$ with respect to μ.*

Let $e(X(n))$ and $e(X(n+1))$ be the sample information of size n and $n+1$ respectively generated from X, and let the two decision problems D_n and D_{n+1} be defined by $D_n=\{D_0\,;\,e(X(n))\}$ and $D_{n+1}=\{D_0\,;\,e(X(n+1))\}$ respectively. Then we have

(4.6)
$$0\leqslant V(X(n+1)\mid\xi,W)-V(X(n)\mid\xi,W)$$
$$\leqslant\Big[2CKR(D)_n\Big]^{\frac{1}{2}}\alpha^{\frac{1}{2}}\Big[\sum_{i\neq j=1}^{m}\sqrt{\xi_1\cdot\xi_j}(\beta_{ij})^n\Big]^{\frac{1}{2}},$$

where C and K are absolute constants defined by (1.30) *and*

(4.7)
$$\beta_{ij}=\int_X[f_i(x)f_j(x)]^{\frac{1}{2}}d\mu,$$

where X is one dimensional Euclidian space and

(4.8)
$$\alpha=\max(\alpha_1,\alpha_2,\cdots,\alpha_m)$$

where

(4.9)
$$\alpha_1=\int_{X(1)}f_1(x)d\mu,\quad \alpha_2=\int_{X(2)}f_2(x)d\mu,\cdots,\quad \alpha_m=\int_{X(m)}f_m(x)d\mu,$$

and

$$X_{(1)}=\{x:f_1(x)\geqslant f_2(x),\ f_1(x)\geqslant f_3(x),\cdots \text{ and } f_1(x)\geqslant f_m(x)\},$$

(4.10)
$$X_{(2)}=\{x:f_2(x)>f_1(x),\ f_2(x)\geqslant f_3(x),\cdots \text{ and } f_2(x)\geqslant f_m(x)\},$$

....

$$X_{(m)}=\{x: f_m(x)>f_1(x),\ f_m(x)>f_2(x), \cdots \textit{ and } f_m(x)>f_{m-1}(x)\}.$$

REMARK 4.1. Since under the condition (4.5), by Schwarz inequality, we have $0\leqq \beta_{ij}<1$ for $j>i=1, 2, \cdots, m$. The theorem 4.2 shows that the marginal information value of the sample information decreases to zero as the sample size n increases.

Proof of theorem 4.2. From the definition of information amount, we have

$$\begin{aligned} I_d &\equiv I(X(n+1)\,|\,\xi)-I(X(n)\,|\,\xi) \\ &= E[H(\xi(X(n))]-E[H(\xi(X(n+1))]. \end{aligned} \tag{4.11}$$

Let

$$\xi_1 f_1(x(n))+\xi_2 f_2(x(n))+\cdots+\xi_m f_m(x(n))=f(x(n)) \tag{4.12}$$

and

$$\begin{aligned} &\xi_1 f_1(x(n+1))+\xi_2 f_2(x(n+1))+\cdots+\xi_m f_m(x(n+1)) \\ &=\xi_1 f_1(x(n))f_1(x_{n+1})+\xi_2 f_2(x(n))f_2(x_{n+1})+\cdots+\xi_m f_m(x(n))f_m(x_{n+1})=f(x(n+1)). \end{aligned} \tag{4.12'}$$

Then from (4.11) and the definition of $H(\cdot)$, we have

$$\begin{aligned} I_d=&\int_{X(n)}\int_X\Big[\frac{\xi_1 f_1(x(n))f_1(x_{n+1})}{f(x(n+1))}\log\frac{\xi_1 f_1(x(n))f_1(x_{n+1})}{f(x(n+1))} \\ &+\frac{\xi_2 f_2(x(n))f_2(x_{n+1})}{f(x(n+1))}\log\frac{\xi_1 f_1(x(n))f_1(x_{n+1})}{f(x(n+1))} \\ &+\cdots+\frac{\xi_m f_m(x(n))f_m(x_{n+1})}{f(x(n+1))}\log\frac{\xi_m f_m(x(n))f_m(x_{n+1})}{f(x(n+1))}\Big]f(x(n+1))\,d\mu^{n+1} \\ &-\int_{X(n)}\Big[\frac{\xi_1\xi_1(x(n))}{f(x(n))}\log\frac{\xi_1 f_1(x(n))}{f(x(n))}+\frac{\xi_2 f_2(x(n))}{f(x(n))}\log\frac{\xi_2 f_2(x(n))}{f(x(n))} \\ &+\cdots+\frac{\xi_m f_m(x(n))}{f(x(n))}\log\frac{\xi_m f_m(x(n))}{f(x(n))}\Big]f(x(n))\,d\mu^n \\ \equiv& A_1-A_2, \end{aligned} \tag{4.13}$$

where A_1 and A_2 are defined by the context. Since

$$\int_X f_1(x_{n+1})\,d\mu=\int_X f_2(x_{n+1})\,d\mu=\cdots=\int_X f_m(x_{n+1})\,d\mu=1, \tag{4.14}$$

we can rewrite A_2 as follows:

$$\begin{aligned} A_2=&\int_{X(n)}\int_X\Big[\frac{\xi_1 f_1(x(n))f_1(x_{n+1})}{f(x(n))}\log\frac{\xi_1 f_1(x(n))}{f(x(n))} \\ &+\frac{\xi_2 f_2(x(n))f_2(x_{n+1})}{f(x(n))}\log\frac{\xi_2 f_2(x(n))}{f(x(n))} \\ &+\cdots \end{aligned} \tag{4.15}$$

$$+\frac{\xi_m f_m(x(n))f_m(x_{n+1})}{f(x(n))}\log\frac{\xi_m f_m(x(n))}{f(x(n))}\Big]f(x(n))d\mu^{n+1}.$$

Then from (4.13) and (4.15), we have

$$I_d=A_1-A_2$$

$$=\int_{X(n)}\int_X\Big[\xi_1 f_1(x(n))f_1(x_{n+1})\log\frac{f(x(n))f_1(x_{n+1})}{f(x(n+1))}$$

(4.16) $$+\xi_2 f_2(x(n))f_2(x_{n+1})\log\frac{f(x(n))f_2(x_{n+1})}{f(x(n+1))}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+\xi_m f_m(x(n))f_m(x_{n+1})\log\frac{f(x(n))f_m(x_{n+1})}{f(x(n+1))}\Big]d\mu^{n+1}.$$

That is, from (4.12′) and (4.16),

$$I_d=-\int_{X(n)}\int_X\xi_1 f_1(x(n))f_1(x_{n+1})\log\frac{\sum_{i=1}^m\xi_i f_i(x(n))f_i(x_{n+1})}{f(x(n))f_1(x_{n+1})}d\mu^{n+1}$$

$$-\int_{X(n)}\int_X\xi_2 f_2 x(n))f_2(x_{n+1})\ \log\frac{\sum_{i=1}^m\xi_i f_i(x(n))f_i(x_{n+1})}{f(x(n))f_m(x_{n+1})}d\mu^{n+1}$$

(4.17) $$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$-\int_{X(n)}\int_X\xi_m f_m(x(n))f_2(x_{n+1})\log\frac{\sum_{i=1}^m\xi_i f_i(x(n))f_i(x_{n+1})}{f(x(n))f_m(x_{n+1})}d\mu^{n+1}$$

$$\equiv B_1+B_2+\cdots+B_m,$$

where $B_1, B_2, \cdots, B_m$ are defined by the context. Since by Bayes's theorem

$$\xi_i f_i(x(n))/f(x(n))=\xi_i(x(n)),\quad i=1,2,\cdots,m$$

we can rewrite $B_1, B_2, \cdots$ and B_m as follows:

(4.18) $$B_1=-\int_{X(n)}\int_X\xi_1(x(n))f_1(x_{n+1})\log\Big[\xi_1(x(n))+\xi_2(x(n))\frac{f_2(x_{n+1})}{f_1(x_{n+1})}$$

$$+\cdots+\xi_m(x(n))\frac{f_m(x_{n+1})}{f_1(x_{n+1})}\Big]f(x(n))d\mu^{n+1},$$

(4.19) $$B_2=-\int_{X(n)}\int_X\xi_2(x(n))f_2(x_{n+1})\log\Big[\xi_1((x(n))\frac{f_1(x_{n+1})}{f_2(x_{n+1})}$$

$$+\xi_2(x(n))+\cdots+\xi_m(x(n))\frac{f_m(x_{n+1})}{f_2(x_{n+1})}\Big]f(x(n))d\mu^{n+1},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

(4.20) $$B_m=-\int_{X(n)}\int_X\xi_m(x(n))f_m(x_{n+1})\log\Big[\xi_1(x(n))\frac{f_1(x_{n+1})}{f_m(x_{n+1})}$$

$$+\xi_2(x(n))\frac{f_2(x_{n+1})}{f_m(x_{n+1})}+\cdots+\xi_m(x(n))\Big]f(x(n))d\mu^{n+1}.$$

Now from the definitions of $X_{(1)}$, $X_{(2)},\cdots,X_{(m)}$ given by (4.10), we have followings:

$$X_{(1)}=\left\{x_{n+1}:\frac{f_2(x_{n+1})}{f_1(x_{n+1})}\leqslant 1,\ \frac{f_3(x_{n+1})}{f_1(x_{n+1})}\leqslant 1,\cdots\ \textit{and}\ \frac{f_m(x_{n+1})}{f_1(x_{n+1})}\leqslant 1\right\},$$

$$(4.21)\quad X_{(2)}=\left\{x_{n+1}:\frac{f_1(x_{n+1})}{f_2(x_{n+1})}<1,\ \frac{f_3(x_{n+1})}{f_2(x_{n+1})}\leqslant 1,\cdots\ \textit{and}\ \frac{f_m(x_{n+1})}{f_2(x_{n+1})}\leqslant 1\right\},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$X_{(m)}=\left\{x_{n+1}:\frac{f_1(x_{n+1})}{f_m(x_{n+1})}<1,\ \frac{f_2(x_{n+1})}{f_m(x_{n+1})}<1,\ \cdots\ \textit{and}\ \frac{f_{m-1}(x_{n+1})}{f_m(x_{n+1})}<1\right\}.$$

Then from (4.21) and the fact that each term under the logarithmic operation in B_1, $B_2,\cdots$ and B_m given by (4.18), (4.19), and (4.20) respectively is non-negative, we have the following inequalities : For $x_{n+1}\in X_{(1)}$, we have

$$(4.22)\quad \begin{aligned}&-\log\left[\xi_1(x(n))+\xi_2(x(n))\frac{f_2(x_{n+1})}{f_1(x_{m+1})}+\cdots+\xi_m(x(n))\frac{f_m(x_{n+1})}{f_1(x_{n+1})}\right]\\&\equiv-\log G_1\leqslant-\log\xi_1(x(n)),\end{aligned}$$

where G_1 is defined by the context. For $x_{n+1}\in X_{(2)}$, from (4.21) we have

$$(4.23)\quad -\log G_1\leqslant-\log(\xi_1(x(n))+\xi_2(x(n))).$$

If we proceed this process succesibly, then we have

$$(4.24)\quad -\log G_1\leqslant-\log(\xi_1(x(n))+\xi_m(x(n)))\ \text{for}\ x_{n+1}\in X_{(m)}.$$

From (4.18), (4.22), (4.23) and (4.24), we have

$$(4.25)\quad \begin{aligned}B_1&\leqslant-\int_{X(n)}\xi_1(x(n))\log\xi_1(x(n))\left[\int_{X(1)}f(x_{n+1})d\mu\right]f(x(n))d\mu^n\\&=-\alpha_1\int_{X(n)}[\xi_1(x(n))\log\xi_1(x(n))]f(x(n))d\mu^n,\end{aligned}$$

where α_1 is defined by (4.19). By the similar method, we have

$$(4.26)\quad \begin{aligned}B_2&\leqslant-\int_{X(n)}\xi_2(x(n))\log\xi_2(x(n))\left[\int_{X(2)}f_2(x_{n+1})d\mu\right]f(x(n))d\mu^n\\&=-\alpha_2\int_{X(n)}[\xi_2(x(n))\log\xi_2(x(n))]f(x(n))d\mu^n.\end{aligned}$$

where α_2 is defined by (4.10), $\cdots$ and

$$(4.27)\quad \begin{aligned}B_m&\leqslant-\int_{X(n)}\xi_m(x(n))\log\xi_m(x(n))\left[\int_{X(m)}f_m(x_{n+1})d\mu\right]f(x(n))d\mu^n\\&=-\alpha_m\int_{X(n)}[\xi_m(x(n))\log\xi_m(x(n))]f(x(n))d\mu^n,\end{aligned}$$

where α_m is defined by (4.9).

Therefore, from (4.17), (4.25), (4.26), (4.27), and the definition of α by (4.8), we have

$$I_d = B_1 + B_2 + \cdots + B_m$$
$$\leqq \alpha \int_{X(n)} \left[-\sum_{i=1}^{m} \xi_i(x(n)) \log \xi_i(x(n)) \right] f(x(n)) d\mu^n,$$

that is

$$I_d \leqq \alpha E[H(\xi(X(n)))]. \tag{4.28}$$

Now, in our case, since

$$\int_{X(n)} [f_i(x(n)) f_j(x(n))]^{\frac{1}{2}} d\mu^n = \prod_{r=1}^{n} \int_X [f_i(x_r) f_j(x_r)]^{\frac{1}{2}} d\mu = (\beta_{ij})^n \tag{4.29}$$

by remark 1.4, (4.28) and (4.29) we have

$$I_d \leqq \alpha C \sum_{i \neq j=1}^{m} \sqrt{\xi_i \cdot \xi_j} (\beta_{ij})^n. \tag{4.30}$$

Then, from (4.4), (4.11) and (4.30), we have

$$0 \leqq R(D_n) - R(D_{n+1}) = V(X(n+1) \mid \xi, W) - V(X(n) \mid \xi, W)$$
$$\leqq [2CKR(D_n)]^{\frac{1}{2}} \alpha^{\frac{1}{2}} \left[\sum_{i \neq j=1}^{m} \sqrt{\xi_i \cdot \xi_j} (\beta_{ij})^n \right]^{\frac{1}{2}}$$

as should be proved.

5. The case in which S is infinite.

Let us suppose that $S = \{s; -\infty < s < \infty\}$ or an interval. It is now necessary to remark that as Shannon's measure of information is not invariant under a change of description of the state space.

First, we shall investigate the uncertainty measure of $s \in S$. For example, let $\xi(s)$ be uniformly distributed between 0 and b. Then by (1.1) we have $H(\xi) = \log b$, which takes on all real values as b ranges over $(0, \infty)$. In fact $H(\xi)$ may be $+\infty$ or $-\infty$. As another example, let $\xi(s)$ be normally distributed with mean μ and variance σ^2 such that $-\infty < \mu < \infty$ and $\sigma > 0$. Then we have

$$H(\xi) = \frac{1}{2} \log(2\pi e \sigma^2) \tag{5.1}$$

whose values are positive or negative according to σ^2. By the above two examples, we know that, $H(\xi) \geqq 0$ is not always held. It is well known that $H(\xi) \geqq 0$ in the case where S is finite.

Next, let us investigate the quantity, $M(X(n) \mid \xi)$ which is defined by

$$M(X(n) \mid \xi) = E[H(\xi(X(n)))] = -\iint f(x(n)) \xi(x(n)) \log(\xi(x(n))) d\lambda dx(n). \tag{5.2}$$

For example, let us suppose that $X_1, X_2, \cdots, X_n$ is a random sample from a normal distribution with an unknown value of the mean S and a specified value of the variance r^2.

Suppose also that the prior distribution of S, $\xi(s)$, is a normal distribution with mean μ and variance σ^2. Then the posterior distribution of S when $X_i=x_i$ $(i=1,\cdots,n)$ is a normal distribution with mean μ' and precision $\frac{1}{\sigma^2}+\frac{n}{r^2}$, where $\mu'=\frac{\mu\sigma^2+n\bar{x}(n)r^2}{r^2+n\sigma^2}$ [5]. By (5.2), we can easily have

$$M(X(n)\,|\,\xi)=\frac{1}{2}\log\left\{2\pi e\frac{\sigma^2r^2}{n\sigma^2+r^2}\right\} \tag{5.3}$$

By (5.3), we know that $M(X(n)\,|\,\xi)\geqq 0$ is not always held. But by theorem 1.3 we know that $M(X(n)\,|\,\xi)\geqslant 0$ in the case when S is finite. By the above two results, we know that our results in section 1 cannot be generalized to the case where the set of possible values of S is infinite.

REMARK 5.1. Information amount which is defined by (1.5) can give some guiding rules for the Bayes decision in the case where the set of possible values of S is infinite [5], [7].

Acknowledgement.

The author would like to express his deep gratitude to professors K. Miyasawa and Y. Suzuki at Tokyo University whose help and guidance were immeasureable in the preparation of this paper.

References

[1] Albert, Perez., *Information, ε-sufficiency and Data Reduction Problems*, Kybernetika, (1965), 297–323.

[2] Albert, Perez., *Information Theory Methods in Reducing Complex Decision Problems.* Transactions of the Fourth prague Conferenceon Information Theory, Statistical Decision Functions, Random Processes. (1965), 33–55.

[3] Ash, R. B., *Information Theory*, John Wiley & Sons, London. Sydney, (1967).

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

[5] DeGroot, M. H., *Optimal Statistical Decision*, McGraw-Hill (1970).

[6] Halmos, P. R., *Measure Theory*, Van Nostrand, 1962.

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

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

[9] Matusita, K., *On the notion of affinity of serveral distributions and some of its application*, Ann. Inst. Stat. Math., Vol. **19**, No. 2, (1967) 137–155.

[10] Matusita, K., *Some properties of affinity and applications*, Ann. Inst. Math., Vol. **23**, No. 2, (1971) 137–155.

[11] 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.

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

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

[14] Rényi, *A.*, *On some basic problems of statistics from the point of view of information theory*, Proc. 5th Berkeley Symp. Math. Statistics and Probability 1, University of California Press, Berkeley, Calif. (1967), 531-543.
[15] Maritz, J. S., *Empirical Bayes Methods*, Methuen and Co. Ltd. (1970), 1-151.

Seoul National University