On Some Properties of Amount of Information

Jae Joo Kim*
Kil Won Suh*

0. Introduction

To gain knowledge about the unknown state Θ of nature, we perform experiments. Sufficiency of an experiment for another experiment has been introduced for the comparison of experiments [1]. The knowledge before and after an experiment makes it possible to discuss the amount of information provided by the experiment [3], [4].

The amount of information is defined as the expected difference between the entropy of the prior distribution over Θ and the entropy of the posterior distribtion [4]. The amount of information has some guiding post to the information value which is defined as the expected difference between the Bayes risk of the prior distribution over Θ and the Bayes risk of the posterior distribution [6]. The loss functions are at hand in the information value. The information cost or the experimental cost should be taken into consideration for decision problems [5], but it will be excluded in this paper.

The object of this paper is as follows:
In section 1, we shall analyze the binomial dichotomy experiments and show that the view of Lindley [4] is more powerful than that of Blackwell [1] for the comparision of experiments.
In section 2, we shall find a relation between the amount of information

* The authors are Assistant Professor of Statistics and Reserch Assistant, Seoul National University, respectively.

and the Bayes risk for the special but very frequent decision problem in which the prior distribution is continuous and normal.

In section 3, we shall show that the information value can be evaluated by means of the amount of information.

1. Comparision of the Binomial Dichotomy Experiments

Let (Θ,β) be a measurable state space, where β is a σ-field of the subsets of Θ. Let ξ be a probability measure of (Θ,β) and absolutely continuous with respect to a measure m on (Θ,β). We denote $d\xi=\xi(\theta)dm$. Then the amount of information against ξ is defined by

$$I_0=\int\xi(\theta)\log\xi(\theta)dm,$$

where the logarithm has base 2.

Let X be a random variable on a measurable space (X,χ) whose probability density function $f(x|\theta)$ given $\theta\in\Theta$ with respect to a measure η on (X,χ) is assumed to be known, where χ is a σ-field of the subsets of X. We also assume that a probability measure P_0, $\theta\in\Theta$, on (X,χ) which specifies the random variable X has the distribution function $f(x|\theta)$ with respect to the measure η on (X,χ) and let $P=\{P_\theta\}$. Then the ordered triple (X,Θ,P) characterizes an experiment $\mathscr{E}$.

After an experiment has been performed, we have an observed value x of X. Then the posterior distribution is given by

$$\xi(\theta|x)\equiv\xi(x)=\frac{f(x|\theta)\xi(\theta)}{f(x)},$$

where $f(x)=\int f(x|\theta)\xi(\theta)dm$.

The amount of information against the posterior distribution $\xi(x)$ is defined by

$$I_1(x)=\int\xi(x)\log\xi(x)dm.$$

Definition 1.1 The amount of information provided by an experiment $\mathscr{E}$, with a prior distribution $\xi(\theta)$, when the observed value of X is x, is

$$I(\mathscr{E},\xi(\theta),x)=E[I_1(X)]-I_0.$$

Consider two binomial dichotomy experiments, $\mathscr{E}_1=(X,\Theta,P_1)$ and $\mathscr{E}_2=(X,\Theta,P_2)$, where $X=\{0,1\}$, $\Theta=\{\theta_1,\theta_2\}$ and $Y=\{0,1\}$. P_1 is defined for

$$f_x(x=1|\theta_i)=p_i=1-f_x(x=0|\theta_i) \quad (i=1,2),$$

when $0\leqq p_1\leqq p_2\leqq 1$. P_2 is defined for

$$f_y(x=1|\theta_i)=q_i=1-f_y(x=0|\theta_i) \quad (i=1,2),$$

where $0\leqq q_1\leqq q_2\leqq 1$.

Theorem 1.2 $\mathscr{E}_2$ is sufficient for $\mathscr{E}_1$ if and only if,

$$(1.1) \quad \frac{1-q_2}{1-q_1}\leqq\frac{1-p_2}{1-p_1}\leqq\frac{p_2}{p_1}\leqq\frac{q_2}{q_1}$$

Proof We must show that there is a nonnegative function $h(x,y)$, where $x=0,1$ and $y=0,1$, such that the follwing equations are satisfied both for $x=0$ and for $x=1$:

$$f_x(x|\theta_1)=h(x,0)f_y(0|\theta_1)+h(x,1)f_y(1|\theta_1)$$

and $$f_x(x|\theta_2)=h(x,0)f_y(0|\theta_2)+h(x,1)f_y(1|\theta_2).$$

Putting $\Delta=q_2(1-q_1)-q_1(1-q_2)$, we have

$$(1.2) \quad h(0,0)=\{q_2(1-p_1)-q_1(1-p_2)\}/\Delta,$$

$$(1.3) \quad h(0,1)=\{(1-q_1)(1-p_2)-(1-p_1)(1-q_2)\}/\Delta,$$

$$(1.4) \quad h(1,0)=(p_1q_2-q_1p_2)/\Delta \text{ and}$$

$$(1.5) \quad h(1,1)=\{p_2(1-q_1)-p_1(1-q_2)\}/\Delta.$$

Since $q_2\geqq q_1$, we have $\Delta=q_2(1-q_1)-q_1(1-q_2)>0$ and hence all the numerators in (1.2), (1.3), (1.4) and (1.5) must be nonnegative. Then we have following inequalities:

$$q_2/q_1\geqq(1-p_2)/(1-p_1),(1-q_2)/(1-q_1)\leqq(1-p_2)/(1-p_1),$$

(1.6) $q_2/q_1 \geqq p_2/p_1$ and $(1-q_2)/(1-q_1) \leqq p_2/p_1$.

From (1.6) and the assumption, we have (1.1). Since $h(0,0)+h(1,0)=1$ and $h(0,1)+h(1,1)=1$, the proof is completed.

Remark 1.3 Since $\mathscr{E}_2$ is sufficient for $\mathscr{E}_1$,$\mathscr{E}_2$ is not less informative than $\mathscr{E}_1$ in view of Backwell [1] and Lindley [4].

Now we consider that Θ is uniformly distributed, *i.e.*,

$$\xi(\theta_1)=\xi(\theta_2)=\frac{1}{2}.$$

Let $I(p_1,p_2)$ be the amount of information after performing $\mathscr{E}_1$. Then

$$I(p_1,p_2)=S(\frac{1}{2}p_1+\frac{1}{2}p_2)-\frac{1}{2}S(p_1)-\frac{1}{2}S(p_2),$$

where $S(x)=-x\log x-(1-x)\log(1-x)$ $(0 \leqq x \leqq 1)$ and the logarithm has base 2.

Since $\frac{\partial I(p_1,p_2)}{\partial p_1} \geqq 0$ and $\frac{\partial I(p_1,p_2)}{\partial p_2} \leqq 0$,

$I(p_1,p_2)$ is increasing for p_1 and decreasing for p_2. These enable us to compare $\mathscr{E}_1$ and $\mathscr{E}_2$ when $p_i=q_i$, $(i=1,2)$.

Lemma 1.4 There exists a unique $p(0 \leqq p \leqq \frac{1}{2})$ satisfying $I(p_1,p_2)=I(p,1-p)$, where p_1 and p_2 are fixed. And $I(p,1-p)<I(q,1-q)$ for $p<q$.

Proof Since $S(p)=S(1-p)$, it suffices to show that there exists a unique $p(0 \leqq p \leqq \frac{1}{2})$ satisfying

(1.7) $S(p)=1-\left[S(\frac{1}{2}p_1+\frac{1}{2}p_2)-\frac{1}{2}S(p_1)-\frac{1}{2}S(p_2)\right]=1-I(p_1,p_2)$.

Since $S(x)$ is concave for x, we have $I(p_1,p_2) \leqq 0$. Since $0 \leqq S(x) \leqq 1$ for $x(0 \leqq x \leqq 1)$ and $0 \leqq p_i \leqq 1$, $(i=1,2)$, we have $I(p_1,p_2) \leqq 1$. Hence we have

(1.8) $0 \leqq 1-I(p_1,p_2) \leqq 1$.

Since $S(p)$ is continuous and one-to-one function and $0 \leqq S(p) \leqq 1$ in the interval [0, 1], we have the first part of this lemma from (1. 7) and (1. 8). The second part immediately follows from (1. 7) and concavity of $S(p)$.

Theorem 1.5 Let $k(0<k<1)$ be a constant. If $p_2-p_1=q_2-q_1=k$ and $0\leq q_1 < p_1 < \frac{1-k}{2}$, then we have

$$I(p_1,p_2)<I(q_1,q_2).$$

Proof Consider

$$(1.9)\quad I(p_1,p_2)=S(p_1+\tfrac{1}{2}k)-\tfrac{1}{2}S(p_1)-\tfrac{1}{2}S(k+p_1)\equiv I(p_1).$$

From the assumption $p<(1-k)/2$, we can show

$$(1.10)\quad (2-2p_1-k)^2 2p_1(2k+2p_1)<(2p_1+k)^2(2-2p_1)(2-2k-2p_1).$$

Differentiating (1. 10) with respect to p_1, we have

$$(1.11)\quad \frac{d}{dp_1}I(p_1)=\frac{1}{2}\log\frac{\left(1-p_1-\frac{1}{2}k\right)^2 p_1(k+p_1)}{\left(p_1+\frac{1}{2}k\right)^2(1-p_1)(1-k-p_1)^2}$$

From (1. 10) and (1. 11), we have $\frac{dI(p_1)}{dp_1}<0$ and hence $I(p_1)$ is decreasing for p_1. Therefore, if $q_1<p_1<\frac{1-k}{2}$, we have $I(p_1)<I(q_1)$, *i.e.*,

$$I(p_1,p_2)<I(q_1,q_2).$$

Remark 1.6 If $\mathscr{E}_2$ is not less informative than $\mathscr{E}_1$ in view of Blackwell, then so is it in view of Lindley. But Theorem 1. 5 shows that the converse is not true.

2. Applications of the Amount of Information in Decision Theory

Consider experiments $\mathscr{E}_i=(X,\Theta,P)$, $(i=1,2,\cdots\cdots,n)$, $X_i\subset X$ and $\theta\subset\Theta$, where P is defined for $f(x|\theta)$ given by

$$(2.1)\quad f(x|\theta)=\left(\frac{r}{2\pi}\right)^{\frac{1}{2}}\exp[-r(x-\theta)^2/2].$$

We denote by x_i an observation after performing $\mathscr{E}_i$. We also assume that the experiments $\mathscr{E}_i$, $(i=1, 2, \cdots, n)$ are performed independently against the specified continuous prior distribution given by

$$(2.2)\quad \xi(\theta)=\left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}} \exp[-\tau(\theta-\mu)^2/2].$$

We define experiments $\mathscr{E}^{(i)}(i=1, 2, \cdots, n)$ as follows:

$$\mathscr{E}^{(1)}\equiv\mathscr{E}_1,\ \mathscr{E}^{(2)}\equiv(\mathscr{E}_2,\mathscr{E}^{(1)}),\cdots,\mathscr{E}^{(n)}\equiv(\mathscr{E}_n,\mathscr{E}^{(n-1)}),$$

where $(\mathscr{E}_i, \mathscr{E}^{(i)})$ is the sum of two experiments, $\mathscr{E}_i$ and $\mathscr{E}^{(i)}$ [4].

On the basis of $\mathscr{E}^{(n)}$, a statistician decides whether the mean of a normal distribution which has specified precision r is smaller or larger than the value θ_0, where the prior distribution is (2.1). We assume that the loss functions $L_i(\theta)$ resulting from the possible decision d_i, $(i=1, 2)$ are of the following forms:

$$L_1(\theta)=\begin{cases}0 & \text{for } \theta\leqq\theta_0\\ \theta-\theta_0 & \text{for } \theta>\theta_0,\end{cases}$$

$$L_2(\theta)=\begin{cases}\theta_0-\theta & \text{for } \theta\leqq\theta_0\\ 0 & \text{for } \theta>\theta_0,\end{cases}$$

Then Bayes risk $\rho_n{}^*(\xi)$ after performing $\mathscr{E}^{(n)}$ is given

$$\rho_n{}^*(\xi)=\frac{1}{\sqrt{r}}\psi[\sqrt{r}(\theta_0-\mu)]-\frac{1}{\sqrt{\tau_n}}\psi[\sqrt{\tau_n}(\theta_0-\mu)]$$

where $\psi(s)=\int_s^\infty (x-s)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$ and $\tau_n=\frac{\tau(\tau+nr)}{nr}$ [2].

Since $\frac{d}{dn}\rho_n{}^*(\xi)=-\frac{\tau^2\exp[\tau_n(\theta_0-\mu)^2]}{2\sqrt{2\pi}r\tau_n{}^{3/2}n^2}<0,$

$\rho_n{}^*(\xi)$ is a monotonically decreasing function of n.

Lemma 2.1 Let I_n be the amount of information of $\mathscr{E}^{(n)}$. Then we have

$$I_n=\frac{1}{2}\log\left(1+n\frac{r}{\tau}\right).$$

Proof Against the prior distribution (2.2), the amount of information I_0 is as follows:

$$I_0=\int_{-\infty}^{\infty}\left(\frac{r}{2\pi}\right)^{1/2}\exp[-r(\theta-\mu)^2/2]\log\left[\left(\frac{r}{2\pi}\right)^{1/2}\exp\{-r(\theta-\mu)^2/2\}\right]d\theta$$

$$=\frac{1}{2}\log\frac{r}{2\pi}-\frac{1}{2}/\log_e 2.$$

Using Bayes theorem, the posterior distribution $\xi(x)$ can be found as

$$\xi(x)=\left(\frac{\tau'}{2\pi}\right)^{1/2}\exp[-\tau'(\theta-\mu')^2/2],$$

where $\tau'=n+nr$, $\mu'=\dfrac{\tau\mu+nrx}{\tau+nr}$ and $x=\dfrac{1}{n}\sum_{i=1}^{n}x_i$ [2].

Hence, we have

$$E[I_1(X)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\tau'}{2\pi}\right)^{1/2}f(x)\exp\{-\tau'(\theta-\mu')^2/2\}$$

$$\log\left[\left(\frac{\tau'}{2\pi}\right)^{1/2}\exp\{-\tau'(\theta-\mu')^2/2\}\right]d\theta dx=\frac{1}{2}\log\frac{\tau'}{2\pi}-\frac{1}{2}\log 2.$$

Therefore, we have $I_n=I_0-E[I_1(X)]=\log\left(1+n\dfrac{r}{\tau}\right)$.

Theorem 2.2 If we exclude the infomation cost and if we have, for some n, $\rho_n^*(\xi)=\varepsilon$ and $I_n=I$, then for any integer $N\geqq n$, we have $\rho_N^*(\xi)\leqq\varepsilon$ and $I_N\geqq I$.

Proof The derivative of I_n in Lemma 2.1 with respect to n is positive and hence I_n is monotonically increasing for n. Since $\rho_n^*(\xi)$ is monotonically decreasing for n, this theorem immediately follows.

Remark 2.3 The criterion for the Bayes risk is equivalent to the criterion for the amount of information in this special case.

3. Relation between the Amount of Information and the Information Value

Let (Θ,β) be a measurable parameter space, where $\Theta=\{\theta_1,\theta_2,\cdots,\ \theta_n\}$ is the finite space of state of nature and where β is a σ-field of the subsets of Θ. Let $\xi=(\xi_1,\ \xi_2,\cdots,\ \xi_n)$ be the prior probability distribution over Θ. Let $(\boldsymbol{A},\alpha)$ be a measurable action space, where $\boldsymbol{A}=\{a_1,\ a_2,\cdots,\ a_n\}$ and where α is a σ-field of the subset of $\boldsymbol{A}$. Then we can, without loss of generality, define the loss function w as shown in diagram 1, where $w_{ij}=0$ and $w_{ij}\geqq 0$, $(i=1, 2, \cdots, n$ and $j=1, 2, \cdots, n)$.

	a_1	a_2	$\cdots$	a_n
θ_1	w_{11}	w_{12}	$\cdots$	w_{1n}
θ_2	w_{21}	w_{22}	$\cdots$	w_{2n}
	$\cdots$	$\cdots$	$\cdots$	$\cdots$
θ_n	w_{n1}	w_{n2}	$\cdots$	w_{nn}

Diagram 1.

For these $\Theta,\xi,\ \boldsymbol{A}$ and w, we define the basic decision problem $D_0\equiv\{\Theta,\xi,\boldsymbol{A},w\}$. Now we assume that the distribution law $f(x\,|\,\theta=\theta_i)\equiv f_i(x)$ is known for the given $\theta=\theta_i\in\Theta$, where x is observed value of a random variable X. Then the posterior probabilty distribution is given by $\xi(x)=(\xi_1(x),\ \xi_2(x),\cdots,\ \xi_n(x))$.

The information value $V(X\,|\,\xi,w)$ for the observed value x of the random variable X is defined as follows [5], [6]:

$$V(X\,|\,\xi,w)=R(D_0)-R(D),$$

where $R(D_0)\equiv R_0(\xi\,|\,w)\equiv\inf\limits_{a\in A}\int w(\theta,a)\,d\xi(\theta)$

and $R(D)\equiv R(X\,|\,\xi,w)\equiv E[R_0(\xi(X))]$.

We denote, for simplicity of notation, as follows:

$$R_0 \equiv R(D_0).\ V \equiv V(X|\xi, w),$$

$$W \equiv \max(\sum_{i=1}^{n} w_{ij},\ j=1, 2, \cdots, n),$$

$$w(0) \equiv \min_{a \in A}((\sum_{i=1}^{n} \xi_i w_{ij})/W,\ j=1, 2, \cdots, n)$$

and $w(x) \equiv \min_{a \in A}((\sum_{i=1}^{n} \xi_i(x) w_{ij})/W, j=1, 2, \cdots, n)$.

Then we can evaluate the information value $V(X|\xi, w)$ by means of the amount of information $I(X|\xi)$ as follows:

Theorem 3.1 If there exists at least one n-tuple solution $(y_1, y_2, \cdots, y_n)$ such that $0<y_i<1$, $(i=1, 2, \cdots, n)$ satisfying

$$\text{(3.1)} \quad \sum_{i=1}^{n} \xi_i y_i = w(0) \text{ and } \sum_{i=1}^{n} \xi_i(x) y_i = w(x),$$

then we have

$$\text{(3.2)} \quad WI(X|\xi) \geqq (R_0 - V)\log\frac{R_0 - V}{R_0} + (W - R_0 + V)\log\frac{W - R_0 + V}{W - R_0}.$$

Proof Since

$$\text{(3.3)} \quad R_0 \equiv R(D_0) = Ww(0) \text{ and } R(D) = Ww(x)f(x)dx,$$

where $f(x) = \sum_{i=1}^{n} \xi_i f_i(x)$, we have

$$\text{(3.4)} \quad V \equiv V(X|\xi, w) = W\{w(0) - w(x)f(x)dx\}.$$

From the assumption (3.1), there exist $a_i(x) = y_i$ such that $0<a_i(x)<1$, $(i=1, 2, \cdots, n)$ satisfying

$$\text{(3.5)} \quad \sum_{i=1}^{n} \xi_i a_i(x) = w(0) \text{ and } \sum_{i=1}^{n} \xi_i(x) a_i(x) = w(x),\ x \in X.$$

Since $\int f(x)dx = 1$, we have from (3.4) and (3.5)

$$\text{(3.6)} \quad V/W = \sum_{i=1}^{n} \xi_i \int a_i(x) f(x)dx - \sum_{i=1}^{n} \xi_i(x) a_i(x) dx$$

$$=\sum_{i=1}^{n}\int \xi_i a_i(x)f(x)[1-\frac{\xi_i(x)}{\xi_i}]dx$$

$$\equiv\sum_{i=1}^{n}\int g_i(x)[1-r_i(x)]dx,$$

where $\xi_i(x)/\xi_i\equiv r_i(x)$ and $\xi_i a_i(x)f(x)\equiv g_i(x)$.

From the definition of the amount of information, we have

$$I=\sum_{i=1}^{n}\int f(x)\xi_i(x)\log\xi_i(x)dx-\sum_{i=1}^{n}\xi_i\log\xi_i.$$

Since $E[\xi_i(X)]=\int\xi_i(x)f(x)dx=\xi_i$, $(i=1,2,\cdots,n)$, we have

$$I=\sum_{i=1}^{n}\int f(x)\xi_i(x)\log\xi_i(x)dx-\sum_{i=1}^{n}\int f(x)\xi_i(x)\log\xi_i\ dx$$

$$=\sum_{i=1}^{n}\int \xi_i a_i(x)f(x)\xi_i(x)\log_i(x)dx\ +\sum_{i=1}^{n}\int\xi_i[1-a_i(x)]f(x)r_i(x)\log r_i(x)dx$$

$$\equiv\sum_{i=1}^{n}\int g_i(x)r_i(x)\log r_i(x)dx+\sum_{i=1}^{n}\int k_i(x)r_i(x)\log r_i(x)dx,$$

where $k_i(x)\equiv\xi_i[1-a_i(x)]f(x)\equiv\xi_i f(x)-g_i(x)$.

Let $G(x)=\sum_{i=1}^{n}g_i(x)r_i(x)\log r_i(x)$ and $K(x)=\sum_{i=1}^{n}k_i(x)r_i(x)\log r_i(x)$.

Then we have

$$(3.7)\quad I=\int G(x)dx+\int K(x)dx,\ \text{where}\ J_1\equiv G(x)dx.$$

From (3.2) and the definitions of $g_i(x)$ and $k_i(x)$, we have

$$(3.8)\quad \sum_{i=1}^{n}\int g_i(x)dx=w(0)=R_0/W\ \text{and}$$

$$(3.9)\quad \sum_{i=1}^{n}\int k_i(x)dx=1-w(0)-R_0/W.$$

Using the extended Jensen's inequality [7] for the convex function $x\log x$ $(x>0)$, we have

$$\frac{\int G(x)dx}{\sum_{i=1}^{n}\int g_i(x)dx} \geqq \frac{\sum_{i=1}^{n}\int g_i(x)r_i(x)dx}{\sum_{i=1}^{n}\int g_i(x)dx}\log\left(\frac{\sum_{i=1}^{n}\int g_i(x)r_i(x)dx}{\sum_{i=1}^{n}\int g_i(x)dx}\right),$$

(3.10)

$$\frac{\int K(x)dx}{\sum_{i=1}^{n}\int k_i(x)dx} \geqq \frac{\sum_{i=1}^{n}\int k_i(x)r_i(x)dx}{\sum_{i=1}^{n}\int k_i(x)dx}\log\left(\frac{\sum_{i=1}^{n}\int k_i(x)r_i(x)dx}{\sum_{i=1}^{n}\int k_i(x)dx}\right).$$

Hence, from (3.7) to (3.10), we have

$$J_1 \geqq [\sum_{i=1}^{n}\int g_i(x)r_i(x)dx]\log[\{\sum_{i=1}^{n}\int g_i(x)r_i(x)dx\}/w(0)],$$

$$J_2 \geqq [\sum_{i=1}^{n}\int k_i(x)r_i(x)dx]\log[\{\sum_{i=1}^{n}\int k_i(x)r_i(x)dx\}/\{1-w(0)\}].$$

From (3.6) and (3.8), we have

$$(3.11)\quad \sum_{i=1}^{n}\int g_i(x)r_i(x)dx = R_0/W - V/W$$

and from the definition of $k_i(x)$ and (3.11), we have

$$(3.12)\quad \sum_{i=1}^{n}\int k_i(x)r_i(x)dx = \sum_{i=1}^{n}\int \xi_i r_i(x)f(x)dx - \sum_{i=1}^{n}\int g_i(x)r_i(x)dx$$

$$= \sum_{i=1}^{n}\xi_i - (R_0/W - V/W)$$

$$= 1 - R_0/W + V/W.$$

From (3.10), (3.11) and (3.12), we have

$$I = J_1 + J_2 \geqq (R_0/W - V/W)\log[W/R_0(R_0/W - V/W)]$$

$$+(1-R_0/W+V/W)\log\frac{1}{1-R_0/W}(1-R_0/W+V/W)],\ i.e.,$$

$$WI(X|\xi) \geqq (R_0-V)\log\frac{R_0-V}{R_0} + (W-R_0+V)\ \log\frac{W-R_0+W}{W-R_0}.$$

Remark 3.2 Miyasawa [6] has shown the inequality (3.2) without the assumption (3. 1), when $n=2$.

REFERENCES

〔1〕 Blackwell, David, "Comparision," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, 1951.

〔2〕 DeGroot, M. H., *Optimal Statistical Decisions,* New York, McGraw-Hill, 1970.

〔3〕 DeGroot, M. H., "Uncertainty, Information and Sequential Experiments," *Annal of Mathematical Statistics,* Vol. 33, 1962.

〔4〕 Lindley, D. V., "On a Measure of the Information Provided by an Experiment," *Annal of Mathematical Statistics.* Vol. 27, 1956.

〔5〕 Marshack, Jacob, "Economics of Information Systems," *Journal of the American Statistical Association,* Vol. 66, Number 333 invited paper section, 1971.

〔6〕 Miyasawa, K., *Theory of Information and Statistical Decision,* Tokyo, Ewanami, 1971.

〔7〕 Renyi, Alfred, "On Some Basic Problems of Statistics from the Point of View of Information Theory," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability,* 1967.

요 약

제 목 : 정보량의 성질에 관한 연구
Bayes 통계학 입장에서

제 1 절에서는 실험비교의 척도로써 정보량이 우월성을 가졌음을 보이고,
제 2 절에서는 사전확률이 연속정규분포를 할 때 정보량과 Bayes risk 와의 관계를 보였으며,
제 3 절에서는 정보량과 정보가치의 관계를 보였다.