연구논문집 제 9 권 제 2 호 p. 33~36 1978

2차원인 미지의 상태공간에서 손실정보의 비교

최 우 진
교 양 과 정 부

〈요 약〉

2차원인 미지의 상태공간에 대하여 얻은 손실정보량 사이의 부등식관계를 밝히고 m차원과 m-1차원의 상태공간에서 표준판정의 경우에 대한 정보교란을 비교 검토 하였다.

A note on the relations between the missing information and standard errors

Woo Jin Choi
Dept. of General Education

〈Abstract〉

The relations between the amount of missing imformation and the errors of standard decisions are studied in a two-dimensional space of nature.

I. 서 론

$(S, \&)$를 可測인 상태공간, ξ를 $(S, \&)$上에 정의된 事前확률밀도함수라 하고 $(S, \&)$上의 測度λ에 관하여 절대연속이고 $d\xi=\xi(s)d\lambda$라 두자. 미지의 하나의 상태 $s\in S$에 대한 불확실의 측도 $H(\xi)$를 다음과 같이 정의 한다. [5]

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

특히 $S^{(m)}=\{s_1, s_2, \cdots, s_m\}$, $\xi=\{\xi(s_1)\xi(s_2), \cdots, \xi(s_m)\}=\{\xi_1, \xi_2, \cdots, \xi_m\}$, $\xi_i>0$, $\sum_{i=1}^{m}\xi_i=1$이면

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

이다. 이 때 $H(\xi)$는 $\xi\in S^{m-1}$의 오목함수(concave function)이고 S^{m-1}은 $m-1$차원 simplex이다. X를 확율밀도함수 $f(x|s)$ $s\in S$를 갖는 가측공간 $(X, \mathscr{X})$上의 확율변수라 하고 측도 μ가 $(X, \mathscr{X})$위에 주어져 있다고 하자. 만약 결정자가 X의 출현치 x값을 알 수 있을 때 $s\in S$에 대한 정보 $e(X)$가 결정자에게 주어진다고 말하고 x를 $e(X)$의 通報라고도 한다. 만약 $S^m=\{s_1, s_2, \cdots s_m\}$, $X=\{X_1, X_2, \cdots, X_n\}$이면 정보 $e(X)$는 마르코브 행렬

$$\Lambda=\|\lambda_{ij}\|,\ i=1,2,\cdots,m,\ j=1,2,\cdots,n$$

로 표시되고 이 때

$$\lambda_{ij}=P_r(X=x_j|\tilde{S}=s_i)=f(x_j|s_i)$$

이다. 이 행렬을 $e(X)$의 정보행렬이라 부른다. 여기서 $s\in S$에 대한 정보 $e(X)$는 사전확율밀도함수 ξ와는 무관계 하며 $X=x$가 관찰된 후 事後확율밀도함수$\xi(s|x)$는 Bayesξ의 정리에 의하여

$$\xi(s|x)=\xi(s)f(x|s)/f(x),\ f(x)=\int\xi(s)f(x|s)d\lambda$$

이다. 이후부터 $\xi(s|x)=\xi(x)$, $\xi(s)=\xi$로 표기하기로 한다. 만약 $S^{(m)}=\{s_1, s_2, \cdots, s_m\}$ $\xi=(\xi_1, \xi_2, \cdots\xi_m)\in S^{m-1}$이면$\xi(x)=(\xi(s_1|x), \cdots\xi(s_m|x))=(\xi_1(x), \cdots\xi_m(x))\in S^{m-1}$임은 분명하다.

이제 $X=x$가 관찰된 후의 불확실의 측도를 $H(\xi(x))$로 표시하여 이것의 X에 관한 기대치를 다음과 같이 구한다.

$$m(X|\xi)=E[H(\xi(X))]=\int H(\xi(x))f(x)p\mu \quad (4)$$

정의 1. $m(X|\xi)$를 $e(X)$가 관측된 후의 손실된 정보량이라 한다.

정의 를 사전확율밀도함수라 할 때 $e(X)$가 제공하는 정보량 $I(X|s)$를 다음과 같이 정의한다.

$$I(X|\xi)=H(\xi)-m(X|)$$

여기서 $I(X|\xi)\geq 0$는 증명된 사실이다. 〔5〕

Ⅱ. 손실정보의 량과 $S^{(2)}=\{s_1, s_2\}$인 경우의 표준결정오차와의 관계

$\{X_n\}, (n=1,2,\cdots)$를 확율변수의 數列이라 하자. X_n의 분포는 s_1, s_2의 값에 따라 변한다고 가정하고 $\tilde{S}=s_1, \tilde{S}=s_2$일 때 X_n의 확율밀도함수를 각각 $f_1(x_n)$, $f_2(x_n)$라 둔다. 여기서 $\tilde{S}$는 사전확율밀도 ξ_1, ξ_2 $(\xi_1+\xi_2=1)$에 따라 s_1과 s_2를 취하는 확율변수로 간주한다. 또 $e(X(n))$를 확율변수x에 의하여 생성된 크기 n인 표본으로 부터 얻어지는 표본정보라 하고 $X(n)$은 n차원 확율벡타 $(X_1, X_2, \cdots, X_n)$이다.

I_n을 $\tilde{S}$에 관한 $X(n)$에 포함된 정보량이라 하면 다음과 같다.

$$(2\cdot 1)\quad I_n=I(X(n)|\xi)=H(\xi)-m(X(n)|\xi),$$

$$(2\cdot 2)\quad H(\xi)=\xi_1\log\frac{1}{\xi_1}+\xi_2\log\frac{1}{\xi_2}$$

지금 $X(n)$이 관찰된 후 $\xi_1(x(n))>\xi_2(x(n))$이면 S_1을 $\xi_2(x(n))>\xi_1(x(n))$이면 S_2을 $\xi_1(x(n))=\xi_2(x(n))$이면 s_1과 s_2를 각각 확율 ξ_1과 ξ_2로 선택하기로 하는 결정을 표준결정이라 부른다. 확율변수 ϕ_n를 다음과 같이 정의한다. 이것을 결정함수라 부른다.

$$(2\cdot 3)\qquad \phi_n=\begin{cases} s_1; & \text{표준결정이 } s_1\text{을 수용} \\ s_2; & \text{표준결정이 } s_2\text{을 수용}\end{cases}$$

이제 n개의 관찰치가 얻어진 후에 표준결정의 오차 $\varepsilon_n^{(2)}$는 표준결정이 틀릴 확율로 정의 한다. 즉

$$(2\cdot 4)\quad \varepsilon_n^{(2)}=P_r(\phi_n\neq\tilde{S})=\xi_1P_r(\phi_n=s_2|\tilde{S}=s_1)+\xi_2P_r(\phi_n=s_1|\tilde{S}=s_2)$$ 〔3〕.

$(A,\mathscr{A})$를 가측행동공간, $W(\cdot,\cdot)$을 $(S\times A, \&\times\mathscr{A})$ 위에 정의된 가측손실함수라 하자. 만약 $X(n)=x(n)$이면 결정함수 ϕ_n은 표본공간 X를 $X_{(1)}$과 $X_{(2)}$로 분할하여 $x(n)\in X(j), j=1,2$이면 $\phi_n=s_j$를 수용한다. 또 손실함수 $W(\phi_n, s_j)$는 결정이 정확하면 $W=0$, 그 외에는 $W=1$로 채택하기로 하자. 따라서 표준결정오차는 다음과 같다〔1〕.

$$(2\cdot 5)\quad \varepsilon_n^{(2)}=\xi_1\int X_{(2)}f_1(x(n))dx(n)+\xi_2\int X_{(1)}f_2(x(n))dx(n),$$

여기서

$$X_{(2)}\equiv\left\{x(n);\ \frac{\xi_2}{\xi_1}\cdot\frac{f_2(x(n))}{f_2(x(n))}\geq 1\right\}$$

$$(2\cdot 6)\quad X_{(1)}\equiv\left\{x(n):\ \frac{\xi_2}{\xi_1}\cdot\frac{f_2(x(n))}{f_1(x(n))}<1\right\}$$

이다.

정리 2·1 $\varepsilon_n^{(2)}\leq E\{H_2(\xi(x(n))\}\equiv m_2(X(n)|\xi$, (2·7)

여기서 H_2와 m_2의 값은 밑수 2인 대수값이다.

증명.

$$(2\cdot 8)\qquad f_1\equiv f_1(x(n))=\prod_{r=1}^{n}f_1(x_r|x_1,\cdots,x_{r-1})$$

$$(2\cdot 8)'\qquad f_2\equiv f_2(x(n)=\prod_{r=1}^{n}f_2(x_r|x_1,\cdots,x_{r-1})$$ 로 부터

$$\begin{aligned}(2\cdot 9)\quad E\{H_2(H_2(\xi(x))\}&=\int_X H_2(\xi(x(n))\{\xi_1f_1+\xi_2f_2\}dx(n)\\
&=\xi_1\int_X H_2(\xi(x(n)f_1dx(n)+\xi_2\int_X H_2(\xi(x(n))f_2dx(n)\\
&=\xi_1\int_X\Big[\frac{\xi_1f_1}{\xi_1f_1+\xi_2f_2}\log_2\Big(1+\frac{\xi_2f_2}{\xi_1f_1}\Big)+\frac{\xi_2f_2}{\xi_1f_1+\xi_2f_2}\log_2\Big(1+\frac{\xi_1f_1}{\xi_2f_2}\Big)\Big]f_1dx(n)\\
&\quad+\xi_2\int_X\Big[\frac{\xi_1f_1}{\xi_1f_1+\xi_2f_2}\Big(\log_2\Big(1+\frac{\xi_2f_2}{\xi_1f_1}\Big)+\frac{\xi_2f_2}{\xi_1f_1+\xi_2f_2}\log_2\Big(1+\frac{\xi_1f_1}{\xi_2f_2}\Big]f_2dx(n)\\
&\geq\int_{X_{(2)}}\frac{\xi_{12}f_1^2+\xi_1\xi_2f_1f_2}{\xi_1f_1+\xi_2f_2}\log_2\Big(1+\frac{\xi_2f_2}{\xi_1f_1}\Big)dx(n)\\
&\quad+\int_{X_{(1)}}\frac{\xi_1\xi_2f_1f_2+\xi_2^2f_2^2}{\xi_1f_1+\xi_2f_2}\log_2\Big(1+\frac{\xi_1f_1}{\xi_2f_2}\Big)dx(n)\\
&=\xi_1\int_{X_{(2)}}f_1\log_2\Big(1+\frac{\xi_2f_2}{\xi_1f_1}\Big)dx(n)+\xi_2\int_{X_{(1)}}f_2\log_2\Big(1+\frac{\xi_1f_1}{\xi_2f_2}dx(n)\\
&=\varepsilon_n^{(2)},\end{aligned}$$

여기서 $X_{(2)}\equiv\left\{x(n):\frac{\xi_2f_2}{\xi_1f_1}\geq 1\right\}$

$X_{(1)}\equiv\left\{x(n):\frac{\xi_2f_2}{\xi_1f_1}<1\right\}$

정리 2·2 $\lambda_{12}=\int_X\sqrt{f_1(x(n))f_2(x(n))}dx(n)$이라 두면 다음부등식이 성립 한다.

$0\leq E\{H(\xi(x(n))\}\leq 2C\sqrt{\xi_1\cdot\xi_2}\ \lambda_{12}$,, 여기서

$$C=\max_{0\leq x\leq 1}\frac{h(x)}{\sqrt{x}}\Big(=\max_{'\leq x\leq 1}\frac{h(x)}{\sqrt{1-x}}\Big)$$

$$h(p)=p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p}$$

증명 $f(x(n))=\xi_1 f_1(x(n))+\xi_2 f_2(x(n))$이므로

(2·10) $E\{H(\xi(x(n)))\}=\xi_1\int_X H(\xi(x(n)))f_1 dx(n)$

$+\xi_2\int_X H(\xi(x(n)))f_2 dx(n)$. $C=\max_{0\le x\le 1}\frac{h(x)}{\sqrt{x}}$와 $h(x)\le C\sqrt{x}$는 필요충분조건관계이므로 다음을 얻는다.

(2·11) $H(\xi(x(n)))\equiv h((\xi_1(x(n)))\le C[\xi_1(x(n))]^{\frac{1}{2}}$,

(2·12) $H(\xi)(x(n)))\equiv h(\xi_2(x(n)))\le C[\xi_2(x(n))]^{\frac{1}{2}}$.

Bayes의 정리에 의하여

$$\xi_1(x(n))=\frac{\xi_1 f_1}{\xi_1 f_1+\xi_2 f_2}\le\frac{\xi_1}{\xi_2}\cdot\frac{f_1}{f_2}$$

$$\xi_2(x(n))=\frac{\xi_2 f_2}{\xi_1 f_1+\xi_2 f_2}\le\frac{\xi_2}{\xi_1}\cdot\frac{f_2}{f_1}$$

을 얻는다. 식 (2·11)와 (2·12)로 부터

$$H(\xi(x(n)))\le C\left(\frac{\xi_1}{\xi_2}\right)^{\frac{1}{2}}\left(\frac{f_1}{f_2}\right)^{\frac{1}{2}}$$

$$H(\xi(x(n)))\le C\left(\frac{\xi_1}{\xi_2}\right)^{\frac{1}{2}}\left(\frac{f_2}{f_1}\right)^{\frac{1}{2}}$$

을 얻는다. 식 (2·)의 어느 項도

$$C\int_X(\xi_1\cdot\xi_2)^{\frac{1}{2}}(f_1\cdot f_2)^{\frac{1}{2}}dx(n)$$

보다 크지 않다. 따라서

$$E\{H(\xi(x(n)))\}\le 2C(\xi_1\xi_2)^{\frac{1}{2}}\lambda_{12}$$

이 성립되고 정리 2·1로 부터 $0\le E\{H(\xi(x(n)))\}$가 얻어진다.

Ⅲ. 정보교란과 완전정보의 비교

$I_n, n=1,2,\cdots$은 감소함수가 아니며 $I_n\le H(\xi)$임은 증명된 사실이다. [2]. 그러므로 $\lim_{n\to\infty}I_n=I^*$는 항상 존재한다. 만약 $I^*=H(\xi)$이면 관찰치 $\{X_n\}$, $(n=1,2,\cdots)$은 $\tilde{S}$에 관한 완전정보를 준다고 하고 $I^*<H(\xi)$이면 관찰치 $\{X_n\}$은 $\tilde{S}$에 관한 완전정보를 주지 않는다고 한다[3].

보조정리 3·1.

$$\prod_{r=1}^{n}\lambda^{(r)}{}_{ij}\le\sqrt{\varepsilon^{(m)}{}_n}\frac{\sqrt{\xi_i}+\sqrt{\xi_j}}{\sqrt{\xi_i\xi_j}},\quad r=1,2,\cdots,n$$

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

$$\varepsilon_n^{(m)}=\xi_1[\int_{X_{(2)}}f_1(x(n))dx(n)+\cdots+\int_{X_{(m)}}f_1(x(n))dx(n)]$$
$$+\xi_2[\int_{X_{(1)}}f_2(x(n))dx(n)+\int_{X_{(3)}}f_2(x(n))dx(n)+\cdots+\int_{X_{(m)}}f_2(x(n)dx(n)]$$
$$+\xi_m[\int_{X_{(1)}}f_m(x(n))dx(n)+\cdots+\int_{X_{m-1}}f_m(x)(n()dx(n)].$$

증명. $\prod_{r=1}^{n}\lambda^{(2)}{}_{ij}=\int_X\sqrt{f_i(x(n))f_jx(n))}dx(n)$, $i\neq j=1,2,\cdots,m$. $X_{(i)}=X-X_{(i)}$라 놓자 $f_i(x(n))$가 밀도함수임을 유의하여 Cauchy-Schwarz 부등식을 이용하면

$$\int_{X_{(i)}}\sqrt{f_i(x(n)f_j(x(n))}dx(n)\le(\int_{X_{(i)}}f_j(x(n))dx(n)^{\frac{1}{2}}$$

$$\int_{\overline{X}_{(i)}}\sqrt{f_i(x(n))f_j(x(n))}dx(n)\le(\int_{\overline{X}_{(i)}}f_i(x)n))dx(n))^{\frac{1}{2}}$$

이 성립한다.

$$\varepsilon_n^{(m)}=\sum_{i\neq j=1}^{m}\xi_j\int_{X_{(i)}}f_j(x(n))dx(n)$$

이므로

$$\frac{\varepsilon_n^{(m)}}{\xi_j}\ge\int_{X_{(i)}}f^j(x(n))dx(n),$$

$$\frac{\varepsilon_n^{(m)}}{\xi_i}\ge\int_{\overline{X}_{(i)}}f_i(x(n))dx(n)$$

이 되므로

$$\int_X\sqrt{f_i(x(n))f_j(x(n))}\,dx(n)\le\left(\frac{\varepsilon_n^{(m)}}{\xi_i}\right)^{\frac{1}{2}}+\left(\frac{\varepsilon_n^{(m)}}{\xi_i}\right)^{\frac{1}{2}}=\frac{\sqrt{\varepsilon_n^{(m)}}(\sqrt{\xi_i}+\sqrt{\xi_j}}{\sqrt{\xi_i\cdot\xi_j}}$$

이 성립한다.

정리 3·1 $S^{(m-1)}$, $S^{(m)}$上의 사전밀도함수를 각각 $\xi=(\xi_1,\xi_2,\cdots\xi_{m-1})$, $\xi'=(\xi_1,\xi_2,\cdots,\xi_m)$라 두자. $\varepsilon^{(m-1)}{}_n$, $\varepsilon_n^{(m)}$을 $X(n)$을 관찰한 후의 표준 결정의 오차라 하자. 이제 $S^{(m-1)}$과 $S^{(m)}$에 관한 정보교란의 上界値를 각각 $E\{H^{(m-1)}(\xi(x(n)))\}$과 $E\{H^{(m)}(\xi(x(n))\}$라 둔다. 그러면

$$0\le E\{H^{(m)}(\xi(x(n)))\}-E\{H^{(m-1)}(\xi(x(n)))\}$$
$$=2C\sqrt{m-1}\{\sqrt{m(m-1)}\max_{\xi'}(\sqrt{\varepsilon_n^{(m)}})-(m-2)\max_{\xi}(\sqrt{\varepsilon_n^{(m-1)}})\},$$

여기서 C는 다음과 같이 주어진다.

$C=\max\limits_{0\le x\le 1}\frac{h(x)}{\sqrt{x}}\left(=\max\limits_{0\le x\le 1}\frac{h(x)}{\sqrt{1-x}}\right)$이고

$h(x)=x\log\frac{1}{x}+(1-x)\log\frac{1}{1-x}$이다.

증명. $0\le E\{H(\xi(x(n)))\}\le C\sum\limits_{i\ne j=1}^{m}\sqrt{\xi_i\xi_j}\lambda_{ij}$와

$\prod\limits_{r=1}^{n}\lambda^{(r)}{}_{ij}\le\sqrt{\varepsilon^{(m)}_n}\,\frac{\sqrt{\xi_i}+\sqrt{\xi_j}}{\sqrt{\xi_i\xi_j}}, r=1,2,\cdots n$로 부터

$$
\begin{aligned}
&E\{H^{(m-1)}(\xi(x(n)))\}\\
&=\max_{\xi}\left[C\sum_{i\ne j=1}^{m-1}\sqrt{\xi_i\xi_j}\,\sqrt{\varepsilon^{(m)}_n}\,\frac{\sqrt{\xi_i}+\sqrt{\xi_i}}{\sqrt{\xi_i\xi_j}}\right]\\
&=\max_{\xi}\left[C\sum_{i\ne j=1}^{m-1}\sqrt{\varepsilon^{(m)}{}_n}(\sqrt{\xi_i}+\sqrt{\xi_j})\right]\\
&=C\max_{\xi}\left[\sqrt{\varepsilon^{(m-1)}_n}\right]\max_{\xi}\left[\sum_{i\ne j=1}^{m-1}(\sqrt{\xi_i}+\sqrt{\xi_j})\right]\\
&=2C\max_{\xi}[\sqrt{\varepsilon^{(m-1)}_n}]\max_{\xi}[(m-2)(\sqrt{\xi_1}\\
&\quad+\sqrt{\xi_2}+\cdots+\sqrt{\xi_{m-1}})]\\
&=2C\max_{\xi}[\sqrt{\varepsilon^{(m-1)}_n}]\,[(m-2)\sqrt{m-1}]
\end{aligned}
$$

마찬가지로

$E\{H^{(m)}(\xi(x(n)))\}=2C\max\limits_{\xi'}[\sqrt{\varepsilon^{(m)}_n}]\ \ [(m-1)\sqrt{m}]$

이 성립한다. 따라서 $\max\limits_{\xi'}[\sqrt{\varepsilon^{(m)}_n}]\ \ge\max\limits_{\xi}[\sqrt{\varepsilon^{(m-1)}_n}]$

로 부터 주어진 부등식이 성립된다.

참 고 문 헌

1. Koichi Miyaiiawa. "情報・決定理論序說", 岩波書店 1971.
2. Koichi, Miyasawa. "Information value and the entropy formulas" (To be presented at the second world Congress of the Econometric Society in Canbrridge England; september, 1970)
3. Rényi, A. "On same basic proclem of statistics from the point of vew of imformation theory." proc. of Fith Berkeley Symposium on Mathemiatical Statistics and Proqability, University of California press, 1967.
4. De Groot, M.H., "Uncertainty, Information and Sequential Experiments," Ann. Math stat. Vol. 33, 1962.
5. Lindley, D.V, "On the measure of the information Provided by an expeɹimcnt," Ann. Math Stat Vol27. 1956.