逐次推定過程の情報理論的基礎づけ*

有　本　卓**

Abstract

This paper attempts to establish a theoretic basis for sequential estimation problems from the viewpoint of information theory. The problem which is dealt with consists in that one has to make a decision which point of a set of categories an observed sequence of random variables belongs to. It is supposed that the distributions of the random variables depend on a point of the finite set of categories. Much attentions are paid to constructing the quantities of equivocation with respect to decisions, which satisfy some axiomatically required conditions of goodness measure. A necessary and sufficient condition for an equivocation to be minimized by the Bayes decision rule is obtained when the number of categories is larger than two. In the simplest case when only two categories are possible the Bayes decision rule minimizes an infinite class of equivocation quantities.

1. はじめに

ある確率変数の系列があって，その実現値を観測することにより，それらがある有限個のカテゴリのうちどれに属するかを決定する問題を考える．この種の問題は，信号検出，パターン認識，学習や適応制御などにおいて基本的である．これらの問題には，従来，統計的仮設検定理論に基づく決定法が応用されてきた．しかし，ここでは全く立場を変えて，このような有限個のパラメータ推定の問題に情報理論的な基礎を与える．

この節では，推定過程を情報理論の立場から大まかにながめてみたいのだが，その前に問題を形式化しておく．

n 個のカテゴリ（仮設検定論では「仮設」，情報理論では「情報源」に相当するもの）

$$X=\{X_1, X_2, \cdots, X_n\} \tag{1.1}$$

が考えられ，この n 個のうちの一つから m 個の信号

$$Y=\{Y_1, Y_2, \cdots, Y_m\} \tag{1.2}$$

が発生するものと考える．時刻 $t=1, 2, \cdots$ において受けとる信号の系列をデータと呼び，それを記号で

$$s_N=Y_{i_N}Y_{i_{N-1}}\cdots Y_{i_1} \tag{1.3}$$

と表わす．ここに Y_{i_t} は時刻 t において受け取った信号で Y のうちのどれかである．以下では，簡略にするために，(1.3) を

$$s_N=i_Ni_{N-1}\cdots i_1 \tag{1.4}$$

と書くことにする．また，s_N の全体を

$$S_N=\{s_N=i_Ni_{N-1}\cdots i_1/Y_{i_t}\in Y, i_t=1, 2, \cdots, m\} \tag{1.5}$$

と表わす．ここでは，データ s_N が X のどれに属するかは未知であるが，それの生起する確率

$$P(s_N/X_j)=P(i_Ni_{N-1}\cdots i_1/X_j) \tag{1.6}$$

は既知であると仮定しよう．すなわち，データ s_N が情報源 X_j から発生するものなら，その発生する確率 (1.6) はわかっているものとする．

さて，われわれはすべての j に対して (1.6) が既知であることを利用して，受け取ったデータ s_N が X の中のどのカテゴリに属するかを決定したい．ただし，s_t の属する真のカテゴリは t のいかんにかかわらず不変だと仮定する．そして，ここではつぎのようなプロセスで定められる一種の学習的決定法を考える．

1)　$t=0$ において，各仮設 X_j の確からしさ $P(X_j)=\theta_j$ を任意に設定する．ただし，

$$\theta_j>0, \quad \sum_{j=1}^{n}\theta_j=1$$

こうして定められた $(\theta_1, \theta_2, \cdots, \theta_n)$ を事前確率ある

* The Basis of Sequential Estimation Process from the Viewpoint of Information Theory, by Suguru Arimoto (Faculty of Scientific Engineering, Osaka University)

** 大阪大学・基礎工学部

いは事前決定と呼ぶ．

2) $t=1$ において s_1 を得れば，$P(s_1/X_j)$ $(j=1, 2, \cdots, n)$ を考慮に入れて，事前確率 θ_j を $\theta_j{}^1$ に代える．

ここに，$\theta_j{}^1 \geq 0,\ \sum_{j=1}^{n} \theta_j{}^1 = 1$

3) $t=N$ までに s_N を得れば，$P(s_N/X_j)$ $(j=1, 2, \cdots, n)$ を考慮に入れて，いままでの決定 $\theta_j{}^{N-1}$ を改めて $\theta_j{}^N$ に代える．

ここに $\theta_j{}^N \geq 0,\ \sum_{j=1}^{N} \theta_j{}^N = 1$

4) こうして $t \to \infty$ としたとき，ある j に対して $\theta_j{}^t \to 1$ となれば，X_j が真の仮設であると決定する．

現在実用されている逐次決定の方法は，この種の学習的プロセスに基づいているものが多いようである．しかし，この学習的決定法には，決定法の良さを判定する規準が存在しないという最大の難点がある．そして，この論文の目的は，このような難点を克服するために，決定法の良さを判定する尺度を情報理論的に構成することにある．

上記の 1)～4) は一種の学習過程である．t が大きくなってデータの量が増すと，それのもたらす「情報量*」も増え，それに従って「あいまいさ」が減じるような良い決定が下せるようになるだろう．だが，不幸なことに，学習過程は一種の非平衡プロセスなので，非平衡の熱力学におけるエントロピーと同じように，各時刻 t においてもたらされるデータの情報量をうまく定義することはできない．したがって，ここでは少し観点を変えて，決定法の平均的な良さを表わす「あいまい度」が人為的に定義されるものかどうかを検討する．

第2節では決定法の数学的な定義を行ない，情報理論的あいまい度を先見的に定義する．第3節では，情報理論的あいまい度が持っているいくつかの性質を述べる．第4節では，逆に，一般にあいまい度と名付けてよい尺度はどのようなものであるかを論じ，第5節では，$n \geq 3$ のとき，Bayes 決定法が最適となる（あいまい度を最小にするという意味で）ようなあいまい度は，第2節で定義された情報理論的あいまい度でなければならないことを示す．$n=2$ のときは，Bayes 決定法はいろいろな意味で最適である．すなわち，無数に定義できるあいまい度を Bayes 決定法は最小にする．

このように，決定法の良さをある程度の根拠でもって定めることができたが，これらの拡張は信号 Y の元の数が無限の場合，特に信号が R^m の点をとるような場合には容易に実行できる．また，どのような条件のもとであいまい度をゼロに収束させることができるかという基本的な問題なども，ここで述べた考察を基にして議論することができる．これらの問題を検討した結果は，紙数の都合上，別の機会に発表する予定である[1]．

* ここでいう情報量は数学的にきちんと定義されたものではなく，単に philosophical な言葉として用いている．情報理論では不確定さを表わす尺度として情報量（エントロピー）が定義されたが，それは，本来，情報源に関するものであることに注意する必要がある．

2. 決定法，情報理論的あいまい度

決定法の意味を正確に述べるために，つぎの記法を導入しておこう．

$$\theta = (\theta_1, \theta_2, \cdots, \theta_n)$$

$$1 > \theta_i > 0,\quad \sum_{i=1}^{n} = 1 \tag{2.1}$$

(2.1) のような θ の全体を Θ^n と書き，その閉包を $\bar{\Theta}^n$ としるそう．

定義 1 任意の時刻 $t=N$ において，写像 φ^N: $S^N \to \bar{\Theta}^n$ が定められたとき，この写像の系列

$$\varphi = \{\varphi^0, \varphi^1, \varphi^2, \cdots, \varphi^N, \cdots\} \tag{2.2}$$

を決定法と呼ぶ．ただし，φ^0 については，$S_0 = \phi$（空集合の意味）と考えて形式的に

$$\varphi^0(\phi) = \theta \in \Theta^n \tag{2.3}$$

と定める．すなわち，φ^0 は Θ^n の中から n 次元ベクトル θ を選ぶこととみなされる．この選ばれた θ を**事前確率**あるいは**事前決定**と呼び，φ^N を時刻 N における**決定**ということにする．また，φ^N の全体を Φ^N，φ の全体を Φ などとしるすこともある．

このように定義したとき，時刻 $t=N$ の決定 φ^N は n 個の確率変数の集りと考えることができる．そして，決定法 φ はこのよう確率変数の n 次元ベクトルの無限列とみなされる．

例 1 Bayes 決定法はつぎのように定められている．

(1) $\varphi^0(\phi) = \theta \in \Theta^n$

(2) $\varphi^1(s_1)$ の第 j 成分を $\varphi_j{}^1(s_1)$ と表わすと

$$\varphi_j{}^1(s_1) = \frac{p(s_1/X_j)\theta_j}{\sum_{l=1}^{n} p(s_1/X_l)\theta_l} \qquad (j=1, \cdots, n)$$

(3) 一般に

$$\varphi_j{}^N(s_N) = \frac{p(s_N/X_j)\theta_j}{\sum_{l=1}^{N} p(s_N/X_l)\theta_l} \qquad (j=1, \cdots, n)$$

つぎに，情報理論からの類推で先見的にあいまい度を定義しよう．仮りにデータ s_N は n 個のカテゴリ X のうちで k 番目に属しているとしよう．X を仮設の集合とみなすと，X_k のみが真であるから，このことは

$$p(X_j)=\delta_{kj}=\begin{cases}1 & j=k\\ 0 & j\neq k\end{cases} \tag{2.4}$$

を意味する．そこで，決定法 φ^N に関してつぎの量を定義しよう．

定義 2

$$\begin{aligned}J^*(\varphi^N:\delta_{kj})&=-\sum_{s_N\in S_N}\sum_{j=1}^{n}p(s_N/X_j)p(X_j)\\&\quad\times\log\varphi_j{}^N(s_N)\\&=-\sum_{s_N\in S_N}p(s_N/X_k)\log\varphi_k{}^N(s_N)\end{aligned} \tag{2.5}$$

そして，この量を時刻 N における決定 φ^N に関する**真のあいまい度**と呼ぶことにする．ここで，log の底は何でもよいのだが，便宜上 e（自然対数）とする．

ところで，上に定義した真のあいまい度は未知の値 δ_{kj} を用いている．すなわち，上記の J^* は実際には計算できないものであり，したがって非現実的である．そこで代わりにつぎのような量を考えよう．

定義 3 事前確率 $\theta\in\Theta^n$ を一つ選んで固定する．そのとき，つぎの量

$$J(\varphi^N:\theta)=-\sum_{s_N\in S_N}\sum_{j=1}^{n}p(s_N/X_j)\theta_j\log\varphi_j{}^N(s_N) \tag{2.6}$$

を時刻 $t=N$ における決定 φ^N に関する**情報理論的あいまい度**と呼ぶことにする．

こうして定義された情報理論的あいまい度は，事前確率 θ を選んできてそして決定 φ^N を与えると，すでに既知の値 (1.6) を用いて計算できる．この量が「あいまい度」と呼ぶにふさわしいような性格を持っているかどうかは，次節で示される．

なお，定義3において $N=0$ のときは，便宜上つぎのように定める．

定義 4

$$J(\varphi^0:\theta)=-\sum_{j=1}^{n}\theta_j\log\theta_j=H(\theta) \tag{2.7}$$

すなわち，事前決定 θ のあいまい度はそのエントロピーであると定める．

3. 情報理論的あいまい度の性質

定理 1 事前確率 $\theta\in\Theta^n$ を選んで固定する．そのときつぎのことがらが成立する．

1) 任意の $\varphi^N\in\Phi^N$ に対して

$$J(\varphi^N:\theta)\geq 0 \tag{3.1}$$

2) 任意の $s_N\in S_N$ に対して

$$p(s_N/X_1)=p(s_N/X_2)=\cdots=p(s_N/X_n)$$

が成立するならば

$$\min_{\varphi^N\in\Phi^N}J(\varphi^N:\theta)=H(\theta) \tag{3.2}$$

3) 任意の組 (j,l)（ただし $j\neq l$，$j,l=1,2,\cdots,n$）に対して

$$\sum_{s_N\in S_N}\sqrt{p(s_N/X_j)\cdot p(s_N/X_l)}=0$$

が成立するならば

$$\min_{\varphi^N\in\Phi^N}J(\varphi^N:\theta)=0 \tag{3.3}$$

4) つぎの不等式を満足するようなある決定法 φ が存在する．

$$J(\varphi^N:\theta)\geq J(\varphi^{N+1}:\theta)\quad(N=0,1,2,\cdots) \tag{3.4}$$

5) 例1の Bayes 決定法は $J(\varphi^N:\theta)$ を最小にする．

6) ある決定法 $\varphi=\{\varphi^0,\varphi^1,\varphi^2,\cdots\}$ に対して

$$J(\varphi^N:\theta)\to 0\quad\text{as } N\to\infty \tag{3.5}$$

が成立するならば

(a) $E_k(\varphi_j{}^N)\to\delta_{kj}$ as $N\to\infty$ (3.6)

(b) $E_k\{\varphi_j{}^N-E_k(\varphi_j{}^N)\}^2\to 0$ as $N\to\infty$ (3.7)

(c) 一般に l を正整数とすれば

$$E_k(\varphi_j{}^N)^l\to\delta_{kj}\quad\text{as } N\to\infty \tag{3.8}$$

ここに，$E_k(\varphi_j{}^N)=\sum_{s_N\in S_N}p(s_N/X_k)\varphi_j{}^N(s_N)$ の意味である．

この定理の証明を述べる前に，定理の意味について言及しておこう．

定理の 1) はあいまい度が負にならないことをいっている．2) は，n 個の情報源がデータをすべて等確率で発生するときには n 個のカテゴリ間に差はあり得ないことになり，したがって，どのように良い決定を選んでも，その決定のあいまい度は事前決定のあいまい度（すなわち，事前決定のあいまい性を表わすエントロピー）より小さくすることはできない，といっている．つぎに，3) の前提が満足されれば，どの一つのデータ s_N を受け取っても完全に deterministic に正しい決定が下せるのであるから，当然あいまい度をゼロにするような決定が存在しなければなるまい．3) はこのことを示している．4) は，見かけ上データがふえれば，それに従って決してあいまい度はふえないような良い決定法が存在しなければならないことをいっている．要するに，定理の 1)～4) は，あいまい度が当然持たねばならない性質を前節で定義した情報

理論的あいまい度は持っていることを示している．6)は，あいまい度がゼロに近づけば，決定 φ^N が真の値に収束してくれることを示している．しかも，平均収束の意味でも，確率1の意味でも，φ^N は真の仮設分布 (2.4) に収束してくれるのである．

定理1の証明 1) は定義3の (2.6) 式から明らかである．2) は，一般に関数

$$J(x)=-\sum_{j=1}^{n}\theta_j\log x_j,\quad x_j\geq 0,\quad \sum_{j=1}^{n}x_j=1 \tag{3.9}$$

の最小値は $x_j=\theta_j$ のとき実現されることから明らかである．3) はつぎのようにして示される．いま，任意に $s_N\in S_N$ を選んでみよう．そのとき，3) の前提条件から，$p(s_N/X_j)>0$ となる X_j は X の中でたかだか一つしか存在しない．したがって，もし，このような X_j が存在するとき $\varphi_j{}^N(s_N)=1$, $\varphi_l{}^N(s_N)=0$ $(l\neq j)$ とする．また，$p(s_N/X_j)>0$ となるような X_j が存在しないときは，$\varphi^N(s_N)$ は Θ^n の任意の元としてよい．このように φ^N を定めたとき，$J(\varphi^N\colon\theta)=0$ となることは明らかであろう．4) は，(2.6) で定義された情報理論的あいまい度ばかりでなく，もっと一般のあいまい度についても成立する．したがって，このことは，もっと一般な定理2で証明されるだろう．5) は Lagrange の乗数法を用いて簡単に示されるので省略する．6) はつぎのようにして示される．(3.5) が成立するとき，任意の $\varepsilon>0$ に対してある $N(\varepsilon)$ が存在し，$t>N(\varepsilon)$ ならば

$$\begin{aligned}\varepsilon &> -\sum_{s_t\in S_t}\sum_{j=1}^{n}p(s_t/X_j)\theta_j\log\varphi_j{}^t(s_t)\\ &> -\theta_k\Big[\sum_{s_t\in S_t}p(s_t/X_k)\log\varphi_k{}^t(s_t)\Big]\end{aligned} \tag{3.10}$$

となる．ところで，一般に $0<x\leq 1$ のとき

$$\log x\leq-\Big[(1-x)+\frac{1}{2}(1-x)^2+\cdots\cdots+\frac{1}{l}(1-x)^l\Big] \tag{3.11}$$

が成立するが，いま $l=1$ の場合の不等式を (3.10) に適用すると

$$\begin{aligned}\varepsilon/\theta_k &> \sum_{s_t\in S_t}p(s_t/X_k)\{1-\varphi_k{}^N(s_t)\}\\ &=1-E_k(\varphi_k{}^N)\end{aligned}$$

となる．ε は任意だから，結局

$$E_k(\varphi_k{}^N)\to 1\quad\text{as } N\to\infty \tag{3.12}$$

が導かれた．特に，$\varphi_j{}^N\geq 0$, $\sum_{j=1}^{n}\varphi_j{}^N=1$ であることから，(3.12) は (3.6) を意味する．こうして，6) の (a) が示されたが，(b) と (c) も，(3.10) に不等式 (3.11) と数学的帰納法を適用すれば，逐次証明できる．

なお，この証明からほとんど自明なことであるが，つぎのことが示される．

系 1 (3.5) が成立すれば，確率1で $\varphi_k{}^t\to 1$ as $t\to\infty$ となる．

4. あいまい度の一般的な構成

第2節で情報理論的あいまい度を先見的に定義し，前節ではそれがあいまい度と呼ぶにふさわしいいくつかの性質を持っていることを導いてきた．ここでは立場を逆にして，まず，あいまい度が持つべき最小限の条件を抽出し，このような条件を満足するあいまい度を一般的に構成する．

第2節で定義したように，決定 φ^N は確率変数の n 次ベクトルと見ることができた．したがって，決定 φ^N に関するあいまい度はある種の平均量とみるべきであろう．しかし，真のカテゴリがいずれであるかがわかっていない限り，確率変数の真の期待値は計算することができない．したがって，事前確率を θ と一たん定めた上では，データ s_N と X_j の同時確率を $p(s_N/X_j)\theta_j$ と仮定する以外にどうしようもなく，期待値もこの確率でもってとることにせざるを得まい．ここでは，このような立場から，つぎのような形をしているあいまい度だけを考えていく．

$$F(\varphi_N\colon\theta)=\sum_{s_N\in S_N}\sum_{j=1}^{n}p(s_N/X_j)\theta_j f\{\varphi_j{}^N(s_N)\} \tag{4.1}$$

ここに，$f(x)$ は実数値をとる一価関数であって，かつ，$0<x\leq 1$ で連続とする．

さて，あいまい度と呼ばれる量が当然満足すべき必要最小限の条件は，定理1の注釈にも述べたように，つぎのようなものであるべきであろう．

あいまい度が満足すべき性質

1) 任意の $\varphi^N\in\Phi^N$ に対して $F(\varphi^N\colon\theta)\geq 0$．また，少なくともある一つの φ^N に対して $F(\varphi^N\colon\theta)>0$．

2) 任意の組 (j,l) (ただし $j\neq l$, $j,l=1,2,\cdots,n$) に対して

$$\sum_{s_N\in N_N}\sqrt{p(s_N/X_j)\cdot p(s_N/X_l)}=0$$

が成立するとき

$$\inf_{\varphi_N\in\Phi_N}F(\varphi^N\colon\theta)=0 \tag{4.2}$$

3) 任意の $s_N\in S_N$ に対して

$$p(s_N/X_1)=p(s_N/X_2)=\cdots\cdots=p(s_N/X_n)$$

が成立するとき

$$\inf_{\varphi^N\in\Phi^N} F(\varphi^N:\theta)=F(\theta) \tag{4.3}$$

ここに $F(\theta)$ は N に無関係で各 $0\leq\theta_j\leq1$ に関して連続な関数である.

4) ある決定法 $\varphi=\}\varphi^1,\varphi^2,\cdots\}$ があって

$$\inf_{\varphi^1\in\Phi^1} F(\varphi^1:\theta)\leq F(\theta)$$

$$\inf_{\varphi^{N+1}\in\Phi^{N+1}} F(\varphi^{N+1}:\theta)\leq \inf_{\varphi^N\in\Phi^N} F(\varphi^N:\theta) \tag{4.4}$$

(4.1) の形のあいまい度が上記の 1)～4) を満足するための必要十分条件をつぎに示そう.

定理 2 (4.1) の $F(\varphi^N:\theta)$ があいまい度の性質 1)～4) を満足するための必要十分条件は, 関数 $f(x)$ がつぎの条件を満足することである.

(1) $\inf_{0<x\leq1} f(x)=0$

(2) 少なくともある $0<x\leq1$ に対して $f(x)>0$

定理 2 の証明 まず必要であることを示そう.

2) の前提条件が満足されているとすると, 任意の s_N に対して $p(s_N/X_j)>0$ となるような X_j はたかだか一つしか存在しない. そこで, このような j を $j(s_N)$ と書こう. こうして

$$\inf_{\varphi^N\in\Phi^N} F(\varphi^N:\theta)=\sum_{s_N\in S_N} p[s_N/X_{j(s_N)}]\theta_{j(s_N)} \inf_{0<x\leq1} f(x)$$

となる. ところが, $f(x)$ が $0<x\leq1$ で連続であることと, 2) が成立することから

$$f(x)\geq0, \qquad \inf_{0<x\leq1} f(x)=0$$

でなければならない. こうして (1) が示されたが, (2) が成立しなければならないことは 1) より明らかであろう. つぎに十分であることを示そう. 1) と 2) はほとんど明らかであろう. そこで, 3) を示すために

$$F(\theta_1,\cdots,\theta_n)=\inf_{x\in\bar{\Theta}^n}\sum_{j=1}^n \theta_j f(x_j) \tag{4.5}$$

とおこう. この論文の主旨には本質的でないので理由は述べないが, $f(x)$ が $0<x\leq1$ で連続であることと (1) とより, $F(\theta)=F(\theta_1,\cdots,\theta_n)$ は $\bar{\Theta}^n$ で連続であることがわかる. さて, 3) の前提条件が満足されていると仮定すれば

$$\inf_{\varphi^N\in\Phi^N} F(\varphi^N:\theta)=\sum_{s_N\in S_N} p(s_N/X_1)\inf_{x\in\bar{\Theta}^n}\left[\sum_{k=1}^n\theta_k f(x_k)\right]=F(\theta_1,\cdots,\theta_n)$$

となるが, このことは 3) が証明されたことに等しい. つぎに 4) を示そう. まず, $F(\theta)$ は concave な関数であることを注意しよう. すなわち, $\theta^1,\theta^2\in\Theta^n$ を任意にとり, α を $0\leq\alpha\leq1$ の任意の数とすると, つぎの関係が成立する.

$$\begin{aligned}\alpha F(\theta^1)+(1-\alpha)F(\theta^2)&=\inf_{x\in\Theta^n}\left[\sum_{k=1}^n\alpha\theta_k{}^1 f(x_k)\right]+\inf_{x\in\Theta^n}\left[\sum_{k=1}^n(1-\alpha)\theta_k{}^2 f(x_k)\right]\\&\leq\inf_{x\in\Theta^n}\sum_{k=1}^n\{\alpha\theta_k{}^1+(1-\alpha)\theta_k{}^2\}f(x_k)\\&=F\{\alpha\theta^1+(1-\alpha)\theta^2\}\end{aligned} \tag{4.6}$$

この性質を利用して 4) の第 2 の不等式を示そう. 第 1 の不等式はほとんど同様に示される. 一般に, データ s_{N+1} は時刻 N までに s_N を受け取った後, 時刻 $N+1$ において信号 Y_l を受け取ったものと考えて

$$s_{N+1}=Y_l s_N$$

と書ける. そして, s_N を受け取ったときの Y_l の条件つき確率を

$$p(Y_l/s_N,X_j)=\frac{p(s_{N+1}/X_j)}{p(s_N/X_j)}$$

と書こう. そのとき

$$F(\varphi^{N+1}:\theta)=\sum_{s_N\in S_N}\left[\sum_{l=1}^m\sum_{j=1}^n p(Y_l/s_N,X_j)\times p(s_N/X_j)\theta_j f\{\varphi_j{}^{N+1}(s_{N+1})\}\right]$$

と表わされる. そこで簡単のために, すべての $p(s_N/X_j)$ はゼロでないと仮定しよう. そして

$$\theta_j{}^*=\frac{p(s_N/X_j)\theta_j}{\sum_{j=1}^n p(s_N/X_j)\theta_j}, \qquad p_{lj}=p(Y_l/s_N,X_j)$$

とおき, $\theta^*\in\Theta^n$, $\sum_{l=1}^m p_{lj}=1$ であるとを考慮すると, つぎの式が得られる.

$$\begin{aligned}\inf_{\varphi^{N+1}\in\Phi^{N+1}} F(\varphi^{N+1}:\theta)&=\inf_{\varphi^{N+1}\in\Phi^{N+1}}\sum_{s_N\in S_N}\left[\left\{\sum_{j=1}^n p(s_N/X_j)\theta_j\right\}\times\sum_{l=1}^m\sum_{k=1}^n p_{lk}\theta_k{}^* f(\varphi_k{}^{N+1})\right]\\&=\sum_{s_N\in S_N}\left\{\sum_{j=1}^n p(s_N/X_j)\theta_j\right\}\times\left[\sum_{l=1}^m q_l F\left(\frac{q_{l1}}{q_l},\cdots,\frac{q_{ln}}{q_l}\right)\right]\end{aligned} \tag{4.7}$$

ここに

$$q_{lj}=p_{lj}\theta_j{}^*, \qquad q_l=\sum_{j=1}^n q_{lj}$$

とおいた．ところで，(4.7) は $F(\theta)$ の concavity より

$$\begin{aligned}&\inf_{\varphi^{N+1}\in\Phi^{N+1}} F(\varphi^{N+1}:\theta)\\ &\leq \sum_{s_N\in S_N}\{\sum_{j=1}^{n} p(s_N/X_j)\theta_j\}F(\theta_1^*,\theta_2^*,\cdots,\theta_n^*)\\ &= \inf_{\varphi^N\in\Phi^N} F(\varphi^N:\theta) \qquad (4.8)\end{aligned}$$

と変形される．こうして 4) の第2式が示された．なお，ある s_N, j に対して $p(s_N/X_j)=0$ となる場合も 4) の等2式が成立することは，関数 $F(\theta)$ が $\bar{\Theta}^n$ で連続であることから容易にわかろう．

つぎに，一般的なあいまい度が定理1の 6) に対応する性質を持つための十分条件を示しておく．

定理 3　$f(x)$ が定理2の (1) と (2) の他につぎの条件を満足するとする．

(3)　$f(x)\geq c_1(1-x)+c_2(1-x)^2+\cdots+c_l(1-x)^l$

ここに，$c_1, c_2, \cdots, c_l$ はすべて正の定数である．このとき，任意に固定された $\theta\in\Theta^n$ に対して

$$\lim_{t\to\infty} F(\varphi^t:\theta)=0$$

ならば，$1\leq r\leq l$ なる整数 r に対して

$$E_k(\varphi_j^t)^r \to \delta_{kj} \quad \text{as } t\to\infty$$

$$E_k\{\varphi_j^t-E_k(\varphi_j^t)\}^r \to 0 \quad \text{as } t\to\infty$$

が成立する．

証明は定理1の場合と全く同様なので省略する．

例 2　$f(x)=-\log x$ はまさに不等式 (3) を満足している．その他，たとえば，$f(x)=\sqrt{(1-x)/x}$ についても不等式 (3) が成立するような定数 $c_1, c_2, \cdots, c_l$ を適当に選ぶことができる．ここでは詳しく述べないが，$f(x)=\sqrt{(1-x)/x}$ はパラメータの数が $n=2$ の場合，重要な役割を果たし，また情報理論においては binary channel の場合にしばしば重用される．これは．一つには，$-\log x\leq\sqrt{(1-x)/x}$ という不等式が成立することからきており，もう一つには，例3で述べることからきているものと考えられる．

5.　Bayes 決定法とあいまい度との関係

第3節において，Bayes 決定法は情報理論的あいまい度を最小にする決定法であることを述べた．ここでは，逆に，Bayes 決定法が最適となるような（すなわち，あいまい度を最小にするという意味で）あいまい度は，情報理論的あいまいでなければならないことを示そう（ただし，$n\geq3$ のとき）．

定理2の証明から，特に (4.8) 式の後半の等式から，Bayes 決定法が最適になるためには

$$\inf_{x\in\Theta^n}\sum_{j=1}^{n}\theta_j f(x_j)=\sum_{j=1}^{n}\theta_j f(\theta_j) \qquad (5.1)$$

とならねばならないことがわかる．こうして，つぎの定理が導かれる．

定理 4　$n\geq3$ とする．$f(x)$ がつぎの 1) と 2) の条件を満足するならば，$f(x)=c\log x$ である．ここに c は負のある定数である．すなわち，Bayes の決定法を最適にするあいまい度は情報理論的あいまい度でなければならない．

1)　$f(x)\geq0$，ある $0<x\leq1$ で $f(x)>0$,

$$\min_{0<x\leq1} f(x)=0$$

2)　$f(x)$ は $0<x\leq1$ で微分可能であって，その導関数 $f'(x)$ も $0<x\leq1$ で連続，かつ

$$\inf_{x\in\Theta^n}\sum_{j=1}^{n}\theta_j f(x_j)=\sum_{j=1}^{n}\theta_j f(\theta_j)$$

を満足する．

定理 4 の証明　Lagrange の乗数の定理から 2) よりある実数 λ が存在してすべての j に対して

$$\theta_j f'(\theta_j)-\lambda=0 \qquad (j=1, 2, \cdots, n)$$

とならねばならない．これより，$j\neq l$ に対して

$$\theta_j f'(\theta_j)=\theta_l f'(\theta_l)$$

となるが，これは $n\geq3$ のとき

$$\theta f'(\theta)=\text{const.} \quad \text{for } 0<\theta<1$$

を意味する．したがって，c, c_1 をある定数として

$$f(\theta)=c\log\theta+c_1$$

とならねばならないが，1) より $c_1=0$, $c<0$ が得られる．こうして定理は証明された．

例 3　$n=2$ のときは，定理4の条件 2) が加わっても，情報理論的あいまい度は一意的には定まらない．たとえば

$$f(x)=c\sqrt{\frac{1-x}{x}} \qquad (c>0)$$

という関数は定理2の 1) と 2)，定理4の 1) と 2) をすべて満足し，また，この f で定義されたあいまい度を Bayes 決定法は最小にする．一般に，$n=2$ のとき

$$f'(x)=\frac{1}{x}g\{x(1-x)\}$$

のような関数であれば，Bayes 決定法は最適になる．

6.　おわりに

ここでは，決定の良さに関する尺度として自明に要求される条件を満足するような「あいまい度」を構成した．また，Bayes 決定が最適になるようなあいまい

度は情報理論的あいまい度でなければならないことを導いた．このような考察は，著者の知る限り，まだなされていないように思われる．しかし，決定に関する「あいまい度」という概念を通さずに，データのもたらす情報量を直接定義しようという試みは早くからあったようである．その最初の試みはおそらく D.V. Lindley の論文[2]であると思われるが，A. Rényi[3] はさらに「missing information」という概念を導入し，それが統計的決定の基本問題において有効であることを指摘している．しかし，そこでは初めから Bayes 決定法が仮定されているのであって，Bayes 決定法を用いたことと彼の定義した missing information との関係は明らかにされてはいなかった．これらについての詳しい議論と，Rényi が最初に扱った統計的決定の基本問題が「あいまい度」という概念を用いることによってどのように明瞭にされるかを検討した結果は，別の機会に発表される予定である[1]．

最後に，データのもたらす情報の量をどのように定義したらよいか一言しておきたい．事前決定 θ のあいまい度を $F(\theta)$ とすると，時刻 $t=N$ の決定 φ^N に関するあいまい度が $F(\varphi^N:\theta)$ に減るので，時刻 $t=N$ までのデータのもたらす情報量は

$$I_N(\varphi^N:\theta)=F(\theta)-F(\varphi^N:\theta)$$

と定めるのが素直であろう．しかし，この量は決定 φ^N のとり方に依存するので，それを避けるために，データのもたらす情報の量は最善の決定を行なったときのあいまい度の差し引き

$$I_N(\theta)=F(\theta)-\inf_{\varphi^N\in\Phi^N}F(\varphi^N:\theta)$$

と定めるのがよかろう．しかし，これも事前決定 θ に依存するので，結局，データのもたらす情報量は

$$I_N=\sup_{\theta\in\Theta^n}\{F(\theta)-\inf_{\varphi^N\in\Phi^N}F(\varphi^N:\theta)\} \quad (6.1)$$

と定めるのが最も素直であるように思える．特に，情報理論的あいまい度を用いると，(6.1) 式は

$$I_N=\sup_{\theta\cdot\Theta^n}\{H(\theta)-\inf_{\varphi^N\cdot\Phi^N}J(\varphi^N:\theta)\} \quad (6.2)$$

となるが，これは情報理論で現われる channel capacity と同じ形をしていることに気づく．この量がデータのもたらす情報量として採用されるべき必然性は，まだわかっていないようである．

(昭和43年11月5日受付)

参 考 文 献

1) 有本卓："ベイズ決定法とあいまい度"，電子通信学会論文誌に投稿予定．
2) D.V. Lindley: "On a Measure of the Information Provided by an Experiment" Ann. Math. Statist., Vol. 27, 1956, pp. 986 1005.
3) A. Rényi: "On Some Basic Problems of Statistics from the Point of View of Information Theory" Proceedings of the Fifth Berkeley Symposium, Vol. 1, University of California Press, 1967, pp. 531～543.