数理解析研究所講究録
第48巻 1968年 131-146

統計的学習理論の応用に関する研究

阪大 工　魚崎 勝司　今村 秀樹
田坂 誠男　杉山 博

§1 はじめに

近年における制御理論の発展の一つの側面は いわゆる適応制御，学習制御に関する理論の開発であろうと思われる。

すなわち 従前の理論では 'given'（'known'）としてきた制御対象の特性，構造，また入力，外乱の特性などを本来の姿である 'unknown' なものとしてとらえて，analysis や synthesis をおこなうという立場にたっての理論の開発と発展であり，これは記述函数的表現や非線型理論の発展とも関連している。

このとき重要なのは "学習" 過程というもののとらえ方である。"学習" は 明らかに過去の行動（こゝには環境，刺激，応答，結果などをも含む）と密接に結びついたものとして 行動の変化を示す現象であり，この観点から "学習" 過程を構成することが必要である。

我々はこゝ二，三年にわたって，心理学において学習行動の解析の手段として提案されていたR. BushとF. Mostellerとによる統計的学習理論を"学習"の手法として応用することを考え，理論的，実験的な考察をおしすゝめてきた。こゝではその一端を紹介したい。

§2 統計的学習理論の概要[1]

人間や動物の学習行動を説明するための試みは，従来から心理学的な立場から，種々行なわれてきている。以下で述べるR. BushとF. Mostellerによる統計的学習理論もその一つである。彼らはW. Estesによる刺激標本モデル（Stimulus sampling model）を整理，拡張して学習オペレータなるものを導入し，学習過程を一つの確率過程におきかえて解析をすゝめている。彼らの学習過程のモデルを図示するとFig. 1のようになる。

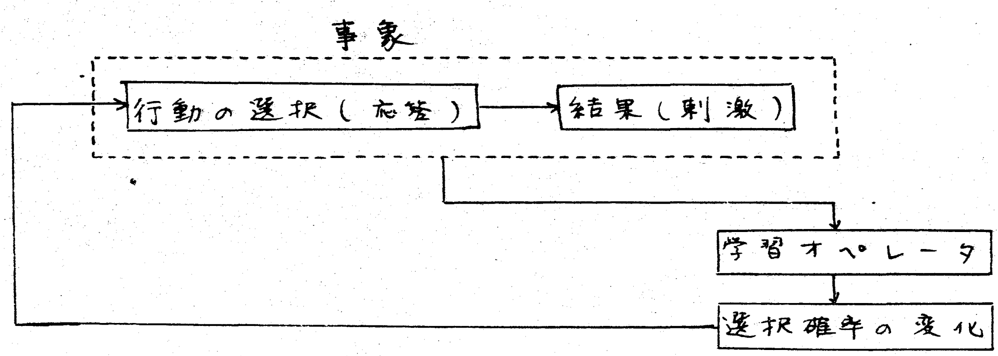

Fig. 1. Bush-Mosteller 学習モデル概念図

行動の選択Aと結果Oとを組み合わせた一つの事象Eに対応して選択確率を変化させる学習オペレータがあるのであるが，その形は一般には，r個の選択と，s個の結果の種類があるときには，

(2・1) $$Q_{ij}\,p=\alpha_{ij}\,p+a_{ij}\ , \qquad i=1,2,\cdots,r\ ,\ j=1,2,\cdots,s$$

または

(2.2) $$Q_{ij}\,p=\alpha_{ij}\,p+(1-\alpha_{ij})\,\lambda_{ij}\ , \qquad \begin{matrix} i=1,2,\cdots,r\ , \\ j=1,2,\cdots,s \end{matrix}$$

こゝに，学習オペレータの作用をうけたのちも確率としての性質を維持するために，

$$0<\alpha_{ij}\leq 1\ ,\quad 0\leq\lambda_{ij}\leq 1 \qquad \begin{matrix} i=1,2,\cdots,r \\ j=1,2,\cdots,s \end{matrix}$$

の制限が，パラメータに付加される。

(2・1)からわかるように，このオペレータは形式的には，一般的なオペレータ

$$Qp=a_0+a_1p+a_2p^2+\cdots$$

のある近似をなしている。

これらのオペレータは，選択，結果によって確率的に作用されることになり，その確率をp_{ij}と表わすとき，このp_{ij}のとり方により，学習に3つの型があることが示される。以下では$r=2$，$s=2$の場合について述べるが，その他の場合も同様である。

(i) Experimenter controlled events

ある事象の起る確率が，実験者によって定められていて，試行によっても変化しない場合で，p_{ij}はその時点の選択の確率pに依存しない。学習オペレータは

$$
(2\cdot 3)\quad
\begin{aligned}
Q_1 p &= \alpha_1 p + (1-\alpha_1)\lambda_1 && \text{with probability } \pi_1 \\
Q_2 p &= \alpha_2 p + (1-\alpha_2)\lambda_2 && \text{with probability } \pi_2 = 1-\pi_1
\end{aligned}
$$

となる。この場合の試行のくり返しによる選択確率の変化のようすをFig. 2に示す。

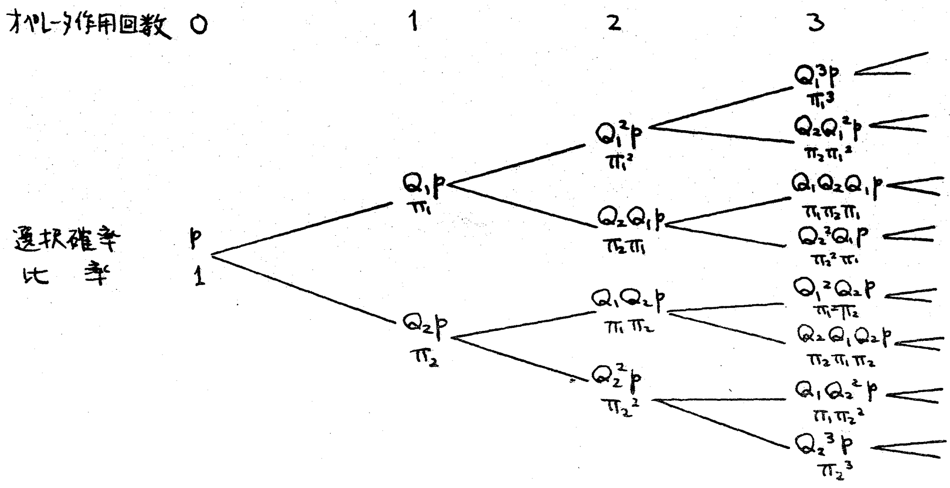

Fig. 2　確率の推移

(ii) Subject-controlled events

選択と事象とが一致する場合で，p_{ij}はその時点の選択の確率pそのものになる。学習オペレータは次で与えられる。

$$
(2\cdot 4)\quad
\begin{aligned}
Q_1 p &= \alpha_1 p + (1-\alpha_1)\lambda_1 && \text{with probability } p \\
Q_2 p &= \alpha_2 p + (1-\alpha_2)\lambda_2 && \text{with probability } q = 1-p
\end{aligned}
$$

この場合の選択確率の変化のようすをFig. 3に示してある。

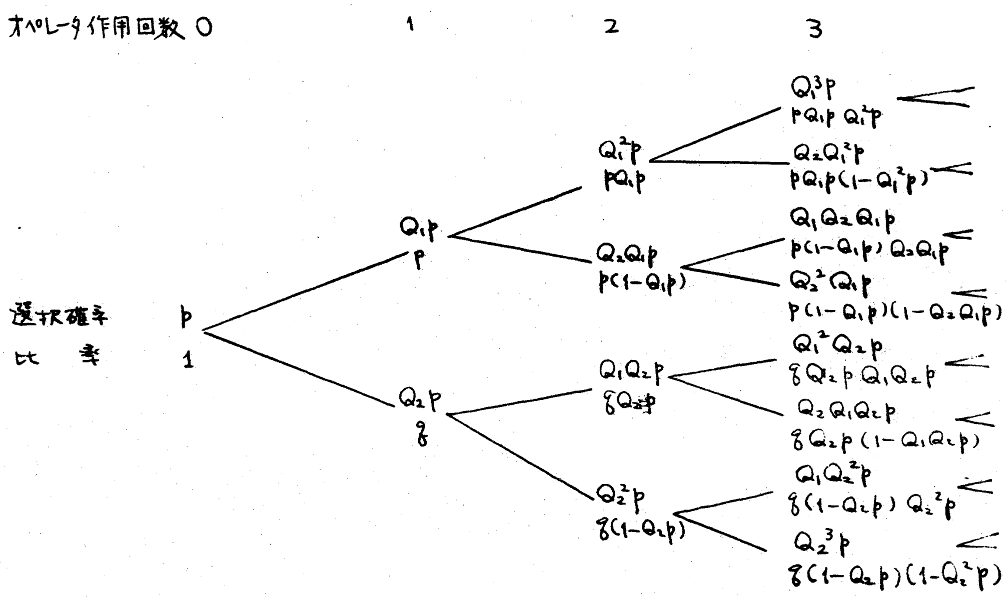

Fig. 3 確率の推移

(iii) Experimenter - Subject - controlled events

事象の起る確率が一定である(i)と，事象の起る確率が，選択の確率に等しい(ii)の二つの考え方を結びつけたもので，ある選択がなされたとき，その選択がなされたという条件のもとで，ある結果の起る条件付確率が一定であるとする。学習オペレータは

$$
\begin{aligned}
Q_{11}p &= \alpha_{11}p + (1-\alpha_{11})\lambda_{11} && \text{with pr. } p\pi_1 \\
Q_{12}p &= \alpha_{12}p + (1-\alpha_{12})\lambda_{12} && \text{with pr. } p(1-\pi_1) \\
Q_{21}p &= \alpha_{21}p + (1-\alpha_{21})\lambda_{21} && \text{with pr. } q\pi_2 = (1-p)\pi_2 \\
Q_{22}p &= \alpha_{22}p + (1-\alpha_{22})\lambda_{22} && \text{with pr. } q(1-\pi_2) = (1-p)(1-\pi_2)
\end{aligned}
\tag{2.5}
$$

によって与えられる。

Bush と Mosteller は，これらのモデルにおいて適当なパラメータを決めることにより，いくつかの学習実験のデータに対しよい一致を得ることができることを示し，データからパラメータを統計的に推定するいくつかの方法についても若干の考察をおこなっている。

§3 proportional parameter の推定に対する応用

二つの事象 A，B が比率 p，$q(=1-p)$ で起る系列が与えられたとき，A の生起の比率のパラメータ p を推定するための手段として，Experimenter controlled events の場合の Bush-Mosteller learning model を採用した場合について考察した結果を述べる。

(i) $\lambda_1=1$，$\lambda_2=0$ とし，ある先験的知識より定めた初期値 $\hat{p}_0$ $(0\leq\hat{p}_0\leq1)$ より出発し，A，B の生起にしたがって次のオペレータ $Q_{(1)}$，$Q_{(2)}$ を作用させる。

$$(3.1)\qquad \begin{aligned}\hat{p}_{n+1}&=Q_{(1)}\hat{p}_n=\alpha\hat{p}_n+1-\alpha, &&\text{if A occurs}\\ \hat{p}_{n+1}&=Q_{(2)}\hat{p}_n=\alpha\hat{p}_n &&\text{if B occurs}\end{aligned}$$

このとき得られる系列 $\{\hat{p}_n\}$ は

$$(3.2)\qquad \hat{p}_n=\sum_{i=1}^{n}(1-\alpha)\alpha^{i-1}y_i+\alpha^n\hat{p}_0,\quad n=1,2,\cdots$$

こゝに $\{y_i\}$ は

$$(3.3)\qquad \begin{aligned}y_i&=1 &&\text{if A occurs}\\ y_i&=0 &&\text{if B occurs}\end{aligned}$$

をとる確率変数列として表わされるので,

$$
(3.4)\quad \begin{aligned} E(\hat{p}_n) &= (1-\alpha^n)p + \alpha^n \hat{p}_0 \\ &\to p \qquad\qquad n \to \infty \end{aligned}
$$

が言える。しかし

$$
(3.5)\quad \begin{aligned} Var(\hat{p}_n) &= p(1-p)(1-\alpha^{2n})(1-\alpha)/(1+\alpha) \\ &\to p(1-p)(1-\alpha)/(1+\alpha) \qquad n \to \infty \end{aligned}
$$

で,一般に,分散は $n\to\infty$ としても0とはならないので,学習の一つの評価函数である真値からのずれ, $E(\hat{p}_n-p)^2$ も $n\to\infty$ で0とはならない。

(ii) (3.1)のオペレータをBush - Mostellerが考えたよりも,広い意味にとり, α が time-dependent ,すなわち n に依存するモデルとして採用すると,

$$
(3.6)\quad \begin{aligned} \hat{p}_{n+1} &= Q_{(1)n}\hat{p}_n = \alpha_n \hat{p}_n + (1-\alpha_n) && \text{if A occurs.} \\ \hat{p}_{n+1} &= Q_{(2)n}\hat{p}_n = \alpha_n \hat{p}_n && \text{if B occurs.} \end{aligned}
$$

の学習オペレータとなる。こゝで種々の criterion に従って,適切なパラメータ系列 $\{\alpha_n\}$ を選定することが重要である。

(3.6)の学習オペレータを用いると, $\{\hat{p}_n\}$ は

$$
(3.7)\quad \hat{p}_n = \sum_{i=1}^{n}(1-\alpha_i)\prod_{j=i+1}^{n}\alpha_j y_i + \prod_{i=1}^{n}\alpha_i \hat{p}_0 \qquad n=1,2,\cdots
$$

となる。よって

$$
(3.8)\quad E(\hat{p}_n) = (1-\prod_{i=1}^{n}\alpha_i)p + \prod_{i=1}^{n}\alpha_i \hat{p}_0
$$

$$
(3.9)\quad Var(\hat{p}_n) = p(1-p)\{(1-\alpha_n)^2 + \sum_{i=1}^{n-1}(1-\alpha_i)^2\prod_{j=i+1}^{n}\alpha_j^2\}.
$$

したがって n までに少くとも1個の整数 k が存在し，

$$\alpha_k = 0 \tag{3.10}$$

ならば，

$$E(\hat{p}_n) = p \tag{3.11}$$

として，不偏性が示される。

この不偏性の条件下で最小分散を与えるパラメータ列は，

$$\alpha_n = 1 - (1/n) \tag{3.12}$$

なることが導びける。このパラメータを用いた学習では，次式の成立することがわかる。

$$\begin{aligned} &E(\hat{p}_n) = p \\ &Var(\hat{p}_n) = E(\hat{p}_n - p)^2 = p(1-p)/n \end{aligned} \tag{3.13}$$

これは，p の最尤推定量のもつ性質と一致している。

(iii) 単に $E(\hat{p}_n - p)^2$ を最小にするパラメータ列は，

$$\alpha_n = 1 - (1/(n+\hat{a})), \tag{3.14}$$

ただし

$$\hat{a} = p(1-p)/(p-\hat{p}_0)^2$$

により与えられる。この学習では，

$$\begin{aligned} &E(\hat{p}_n) = p/(1+\hat{a}/n) + \hat{p}_0\hat{a}/(n+\hat{a}) \\ &Var(\hat{p}_n) = np(1-p)/(n+\hat{a})^2 \\ &E(\hat{p}_n - p)^2 = p(1-p)/(n+\hat{a}) \end{aligned} \tag{3.15}$$

である。

(iv) 真値 p は学習の前段階では未知だからパラメータ列として，上述の $\{\alpha_n\}$ をとることができないので，この近似として，パラメータ列 $\{\alpha_n\}$ を次のようにとることを考える。

(3.16) $$\alpha_n = 1 - (1/(n+a)) \qquad n = 1, 2, \cdots$$

この学習では，初期値 $\hat{p}_0$ の選択について

(3.17) $$p - \{2p(1-p)/a\}^{1/2} \leq \hat{p}_0 \leq p + \{2p(1-p)/a\}^{1/2}$$

の関係が成立していれば，$E(\hat{p}_n - p)^2$ は (ii) の学習より小となる。

(v) また学習理論における regression 関係 (3.8) を用いて，過去の $\{\hat{p}_i\}$ の系列から，p を最小2乗推定することも可能で，この学習による推定は，

(3.18) $$\tilde{p}_n = \sum_{i=1}^{n-1}\left(1 - \prod_{j=1}^{i}\alpha_j\right)\left(\hat{p}_{i+1} - \left(\prod_{j=1}^{i}\alpha_j\right)\hat{p}_0\right) \Big/ \sum_{i=1}^{n-1}\left(1 - \prod_{j=1}^{i}\alpha_j\right)^2$$

であり，次の性質をもっている。

$$E(\tilde{p}_n) = p$$

(3.19) $$\mathrm{Var}(\tilde{p}_n) = E(\tilde{p}_n - p)^2 = p(1-p)\sum_{i=1}^{n}(1-\alpha_i)^2\left\{\sum_{j=i}^{n}\left(1-\prod_{k=1}^{j}\alpha_k\right)\prod_{l=i+1}^{j}\alpha_l\right\}^2 \Big/ \left\{\sum\left(1-\prod_{j=1}^{i}\alpha_j\right)^2\right\}^2$$

なお最初の先験的分布がベータ分布 $B(\bar{a}, \bar{b})$ としたときの Bayesian type の学習過程を，この問題に適用すると，推定値は，

$$\bar{p} = \int_0^1 p^{m+1}(1-p)^{n-m}dH(p) \Big/ \int_0^1 p^m(1-p)^{n-m}dH(p)$$

$$= B(m+\bar{a}+1, n-m+\bar{b}) / B(m+\bar{a}, n-m+\bar{b})$$

となり，(iv) において

$$a = \bar{a} + \bar{b}$$

とした場合に相当することに注意する。

以上の各種の学習を比較すると

不偏性のもとで分散小　　(ii) $\succ$ (v)

$E(\hat{p}_n-p)^2$ 小　　(iii) $\succ$ (iv) $\succ$ (ii) $\succ$ (v)

この議論は，三個以上の事象の生起確率の推定にも拡張できる。

§ 4　Markovian sequence の予測問題への応用

r 個の状態をもつ単純 Markov chain の遷移確率行列は

$$
\mathrm{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{bmatrix} \tag{4.1}
$$

で与えられるが，これを学習オペレータを用いて推定するには，時点 n で状態 E_i にあって，時点 $n+1$ で状態 E_j に遷移したなら，遷移確率行列の第 i 行の成分 p_{ik} の推定値 $\hat{p}_{ik}$ を，学習オペレータにより，次のように変化せしめる。

$$
\begin{aligned} \hat{p}_{ij} &\to \alpha\hat{p}_{ij} + 1 - \alpha \\ \hat{p}_{ik} &\to \alpha\hat{p}_{ik} \qquad k \neq j, \quad k=1,\cdots,r \end{aligned} \tag{4.2}
$$

たゞしこゝで α は前節で述べたような意味にとる。このとき，前節の議論を若干変更するだけで，成分が row-wise に推定されることが示される。

この手法を，s 個の事象 $A_1,\cdots,A_s$ を含む，周期 r の系列 (

rZS）の予測問題に適用する。この周期性には多少の乱れがあってもよく，単に系全体として，r重 Markov chain として表現可能でありさえすればよい。このような場合でも，r個のあいつぐ事象列をあらためて，1つの事象と見なすことにより，S^r個の状態をもつ単純 Markov chain におきかえうるから上の手法が適用できるのである。ただ注意するのは，推定すべき成分は，各 row で S個のみであることである。このようにして推定が行なわれれば，pure strategy（次に生起する確率の最大なる事象を予測），もしくは mixed strategy（次に生起する確率に応じた比率で各事象を予測する）により，予測を行なう。

いま一例として，事象数が2，各々を0，1と表わす，で周期が4の場合の計算機実験の例を示す。とりあげたのは

…00110011001…

なる系列で，この系列の遷移確率行列は次で与えた。

$$
P = \begin{array}{c}
0000\\0001\\0010\\0011\\0100\\0101\\0110\\0111\\1000\\1001\\1010\\1011\\1100\\1101\\1110\\1111
\end{array}
\left[\begin{array}{cccccccccccccccc}
\varepsilon & 1-\varepsilon & & & & & & & & & & & & & & \\
 & & \varepsilon & 1-\varepsilon & & & & & & & & & & & & \\
 & & & & \frac{1}{2} & \frac{1}{2} & & & & & & & & & & \\
 & & & & & & 1-\varepsilon & \varepsilon & & & & & & & & \\
 & & & & & & & & \varepsilon & 1-\varepsilon & & & & & & \\
 & & & & & & & & & & \frac{1}{2} & \frac{1}{2} & & & & \\
 & & & & & & & & & & & & 1-\varepsilon & \varepsilon & & \\
 & & & & & & & & & & & & & & 1-\varepsilon & \varepsilon \\
\varepsilon & 1-\varepsilon & & & & & & & & & & & & & & \\
 & & \varepsilon & 1-\varepsilon & & & & & & & & & & & & \\
 & & & & \frac{1}{2} & \frac{1}{2} & & & & & & & & & & \\
 & & & & & & 1-\varepsilon & \varepsilon & & & & & & & & \\
 & & & & & & & & \varepsilon & 1-\varepsilon & & & & & & \\
 & & & & & & & & & & \frac{1}{2} & \frac{1}{2} & & & & \\
 & & & & & & & & & & & & 1-\varepsilon & \varepsilon & & \\
 & & & & & & & & & & & & & & 1-\varepsilon & \varepsilon
\end{array}\right]
$$

ただし $\varepsilon = 0.03$

遷移確率行列の推定のようすをFig. 4に，10回毎の誤り個数の推移をFig. 5に示す。この結果は10回のくり返し実験の平均である。ただ問題となるのは周期が長く事象の種類の多いときで、このときには、状態数が非常に大きくなることによる難点が生じてくる。

これから pure strategy の場合には，αによる影響はそれほど大きくないように見える．しかし別の解析から，αを小とすれば，分散が大きくなり，誤る危険率が高くなることが示されるので，αは1に十分近くとることが望ましいと結論される。これに反し，mixed strategy では全体的に pure strategy より誤る率が高いことが明らかで，これは予期されることである。また学習の初期では，αの小なるほど誤る割合は小さいが，これはαのもつ意味からもうなづける結果である。

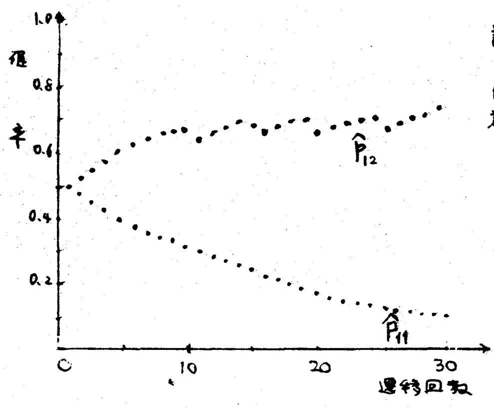

Fig.4. Pの第1行成分の推定のようす　α=0.95

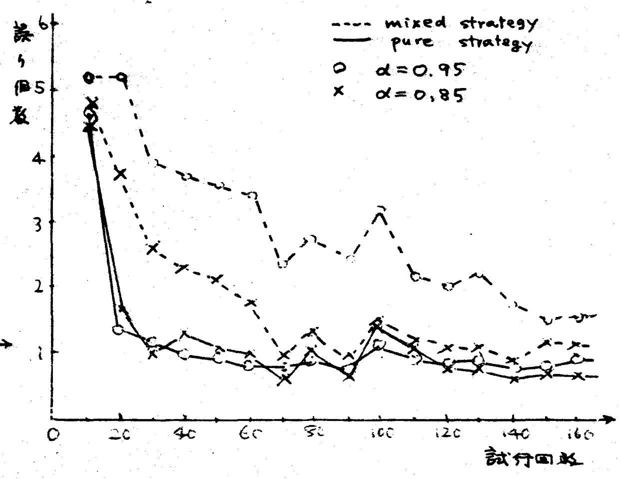

Fig. 5 誤り個数の変化

§5 Two-armed bandit problem への応用

Two-armed bandit problem とは，二種類の選択の可能性（以

下では2個のcoin，AとBを想定する、があって，その各々の選択において，1 unit のgainを得る（表が出る）確率π_1，π_2が定まっているが，実験者には未知であるとするとき，N回の試行で，gainをできるだけ大きくするには，どのように選択していけば良いかということをとり扱う問題で，主にBayesianの立場から考察がなされている。ここでは，min-max approach，Bayesian approach が採用されて考察がなされてきているが，これは状況が不明のときの制御問題のとり扱いと同様である。ここでは，自然とのgameと考えて，min-max approachを採用するが，特に，実際的な立場から，最初の何回かは，二つの選択の可能性を許すが，ある時点以降は，一方にのみきめて試行をおこなう場合の考察を行なう。

このような場合について，Vogel が一つの argorithm を提案している[2]。それによると 最初はcoin A と coin B を対にしてふり，次の確率変数を導入する．

(5.1) $$U_k = \sum_{i=1}^{k} (X_i - Y_i)$$

ここにX_i, Y_iはそれぞれ，coin A，coin B について，表が出たとき1，裏が出たとき0をとる確率変数である．さらにある正の整定数νを導入し，

(5.2) $$-\nu < U_k < +\nu$$

である限り，coin A と coin B の対の実験を行なうが，もし

(5.3) $$U_k \geq +\nu \quad (\leq -\nu)$$

なら，残りの（$N-2K$）回は coin A（coin B）のみをとる。このとき Loss function，

(5.4) $$L(\sigma,\tau,\nu)=\sigma-S_N/N \qquad \sigma=\max(\pi_1,\pi_2),\ \tau=\min(\pi_1,\pi_2)$$

を考え，$\min_{\nu}\max_{\sigma,\tau} L(\sigma,\tau,\nu)$ を与える ν として

(5.5) $$\nu = 0.292\,N^{\frac{1}{2}} \qquad N \geqq 100$$

をとればよいことを示した。

我々は，これに対し，equal α でかつ $\lambda_{11}=\lambda_{22}=1$，$\lambda_{12}=\lambda_{21}=0$ の Experimenter–Subject controlled events の学習により，coin A を選択する確率を変化させることを考えた。すなわち

(5.6) $$\begin{aligned} Q_{11}p &= \alpha p+1-\alpha && \text{with pr.} \quad p\pi_1 \\ Q_{12}p &= \alpha p && \text{with pr.} \quad p(1-\pi_1) \\ Q_{21}p &= \alpha p && \text{with pr.} \quad (1-p)\pi_2 \\ Q_{22}p &= \alpha p+1-\alpha && \text{with pr.} \quad (1-p)(1-\pi_2) \end{aligned}$$

のオペレータを用いる。ただし c，1−c に吸収壁を設け，

(5.7) $$p \geq 1-c \qquad (\leq c)$$

ならばそれ以降は coin A（coin B）のみを用いることにした。この方法(S_L)を用いたときの，$N=100$ に対する Loss function の値を爆発法により計算し，図示したものが Fig. 6 である。なお，coin A の選択確率を

(5.8) $$\begin{aligned} Qp &= p+\Delta p && \text{with pr.} \quad p\pi_1+(1-p)(1-\pi_2) \\ &= p-\Delta p && \text{with pr.} \quad p(1-\pi_1)+(1-p)\pi_2 \end{aligned}$$

で変化させた場合（S_A）についても，Vogel の argorithm（S_V）についても Fig. 7，Fig. 8 に示した。

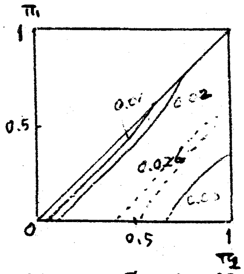

Fig. 6 S_L $\alpha=0.99$, $c=0.47$

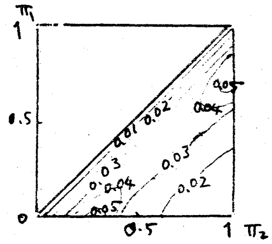

Fig.7 S_A $\Delta p=0.1$

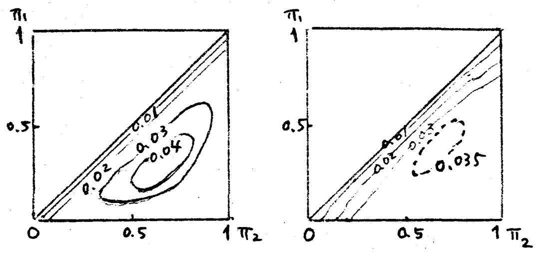

Fig. 8 S_V $\nu=2$ $\nu=3$

これから３つのstrategyは π_1, π_2 の組み合わせに対し，お互いに相補う性質をもつことがわかると共に，学習オペレータを用いた我々の方法は，Vogelの方法より，minmaxの意味でより良い性質を有していることが示された。

§6 結び

いくつかの例で示したように，統計的学習理論を用いた我々の方法は，望ましい学習に対する一つの指針を与えていることがわかる。こゝでとりあげた例では，制御理論との結びつきが，それほど鮮明でないが，将来の応用には十分期待されるものがあると考える。たとえば，こゝでとりあげた予測の問題は適応制御への一つのapproachとして，またTwo-armed bandit problemへの応用はidentificationの問題への一つのapproachを示唆するものといえないであろうか。

今後の課題としては，こゝでとりあげたいくつかの実験の理論的背景を追求するとともに，大局的な視野からみた制御理論への応用を考えていきたい。

参考文献

1) R. R. Bush, F. Mosteller, *Stochastic Models for Learning.* John Wiley (1955)

2) W. Vogel, "A sequential design for the two armed bandit," Ann. Math. Statist., 31, 430 (1960)